



Published in final edited form as:

J Mol Graph Model. 2019 July ; 90: 59–76. doi:10.1016/j.jmglm.2019.04.003.

Deciphering the Three-Domain Architecture in Schlafens and the Structures and Roles of Human Schlafen12 and SerpinB12 in Transcriptional Regulation

Jiaxing Chen^{a,1} Leslie A. Kuhn^{a,b}

^aProtein Structural Analysis and Design Lab, Department of Biochemistry and Molecular Biology, Michigan State University, 603 Wilson Road, East Lansing, MI 48824-1319 USA

Abstract

Schlafen proteins are important in cell differentiation and defense against viruses, and yet this family of vertebrate proteins is just beginning to be understood at the molecular level. Here, the three-dimensional architecture and molecular interfaces of human schlafen12 (hSLFN12), which promotes intestinal stem cell differentiation, are analyzed by sequence conservation and structural modeling in light of the functions of its homologs and binding partners. Our analysis shows that the schlafen or divergent AAA ATPase domain described in the N-terminal region of schlafens in databases and the literature is a misannotation. This N-terminal region is conclusively an Alba_2 DNA/RNA binding domain, forming the conserved core of schlafens and their sequence homologs from bacteria through mammals. Group III schlafens additionally contain a AAA NTPase domain in their C-terminal helicase region. In hSLFN12, we have uncovered a domain matching rho GTPases, which directly follows the Alba_2 domain in all group II-III schlafens. Potential roles for the GTPase-like domain include antiviral activity and cytoskeletal interactions that contribute to nucleocytoplasmic shuttling and cell polarization during differentiation. Based on features conserved with rSlfn13, the Alba_2 region in hSLFN12 is likely to bind RNA, possibly as a ribonuclease. We hypothesize that RNA binding by hSLFN12 contributes to an RNA-induced transcriptional silencing/E3 ligase complex, given the functions of hSLFN12's partners, SUV39H1, JMJD6, and PDLIM7. hSLFN12's partner hSerpinB12 may contribute to heterochromatin formation, based on its homology to MENT, or directly regulate transcription via its binding to RNA polymerase II. The analysis presented here provides clear architectural and transcriptional regulation hypotheses to guide experimental design for hSLFN12 and the thousands of schlafens that share its motifs.

^bCorresponding author; KuhnL@msu.edu.

¹Current address: Bioinformatics and Genomics Graduate Program, Pennsylvania State University, 201 Huck Life Sciences Building, University Park, PA 16802 USA

Competing Interests

The authors have no competing interests to declare.

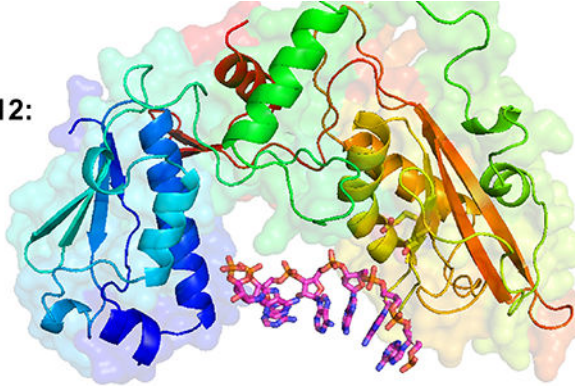
Appendix A. Supplementary data

The following supplementary data associated with this article can be found in the online version:

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Graphical Abstract

Human schlafen12: RNA binding in transcriptional regulation



Keywords

schlafen AAA domain; schlafen ATPase; AlBA_2 domain; GTPase-like proteins; SUV39H1; JMJD6; PDLIM7; MENT; DDB1; cullin4; CLRC (Clr4 cullin ring ligase); RNA-induced transcriptional silencing (RITS); functional annotation; enterocyte differentiation

Introduction

Proteins in the schlafen family (Slfns) are present widely in vertebrates, including mammals, fish, and amphibia [1]. Slfns 1–4 were first reported 20 years ago and found to be differentially regulated during T cell development and thymocyte maturation [2]. The name schlafen means “to sleep” in German, based on the observation that Slfn1 arrests the cell cycle at G0/G1 [2]. Since then, Slfns have become associated with roles including anti-proliferation and cell differentiation [3,4], inhibition of viral replication [5,6], prevention of cancer cell migration and invasion [7], and sensitization of cancer cells to DNA-damaging drugs [8]. However, the molecular mechanisms underlying these functions of Slfns remain poorly understood.

Research has focused on Slfn proteins from mouse (mSlfn1, mSlfn2, mSlfn3, mSlfn4, mSlfn5, mSlfn8, mSlfn9, mSlfn10, and mSlfn14) and humans (abbreviated in capitals as hSLFN5, hSLFN11, hSLFN12, hSLFN13, and hSLFN14) [3]. These schlafens fall into three groups based on protein length: group I (37–42 kDa), group II (58–68 kDa), and group III (100–104 kDa) [4]. All contain a Slfn box motif, which bears the sequence ILPQYVSAFANTDGGYLFGLNE in hSLFN12. Not found in other proteins, the Slfn box is referred to as a “conserved domain signature (COG2865) found in a variety of putative proteins assigned as transcriptional regulators or helicases” [4]. The Slfn box is followed by what has been called a schlafen (or divergent) AAA domain, which lacks the Walker A motif associated with NTPase activity [3,4]. The Walker A or P-loop (phosphate loop) motif is a nucleotide binding structural motif found in the AAA family of ATPases [9] that binds phosphate groups in the nucleotide. This loop has the conserved amino acid sequence G-XXXX-GK-[T or S], linking a β strand to an α helix in the active site [10]; X represents nonconserved positions in the sequence. Alteration or nonconservation of the P-loop

sequence in an otherwise AAA-like domain is associated with loss of ATPase function. Group II-III Slfns contain the Slfn-specific SWADL sequence motif, with unknown function [11]. Group III Slfns additionally contain a AAA domain in their C-terminal region matching superfamily I DNA/RNA helicases [4]. Of the 15 929 sequences identified as Slfns in the Conserved Domains Database [12], molecular functions have been partially characterized for hSLFN5, mSlfn8, hSLFN11, rat Slfn13 (rSlfn13), hSLFN13, and hSLFN14. hSLFN5 is a transcriptional co-repressor of signal transducer and activator of transcription 1 (STAT1)-stimulated interferon responses to infection and glioblastoma [13]. hSLFN11 destabilizes the replication factor A (RPA)-ssDNA complex, inhibiting checkpoint maintenance and DNA repair. hSLFN11 also cleaves tRNA associated with codons required by HIV-1 proteins or kinases that cause deleterious activity in cells with damaged DNA [14,15]. mSlfn8, rSlfn13, and hSLFN13 cleave ribosomal and/or transfer RNA, with the biological implications not yet known [16]. The N-terminal region in hSLFN14 also acts as an endoribonuclease of ribosomal RNA and ribosome-bound mRNA [17].

Our focus on hSLFN12 is due to its crucial role in the differentiation of prostate epithelial cells and nutrient-absorbing intestinal enterocytes and the desire to understand the underlying molecular mechanism [18,19]. At the same time, given the diverse descriptions of Slfn cellular functions and domain structures, we would like to characterize the common structural and functional features of Slfns in general. hSLFN12 is likely to have important roles in cell types beyond enterocytes, since it is expressed in 133 different healthy tissues at various stages of development (<https://bgee.org>) [20]. Understanding and being able to regulate enterocyte differentiation at a molecular level may present useful opportunities for clinical intervention. For instance, downregulating the replenishment of enterocytes could help control obesity, while stimulating the renewal of enterocytes would aid in reversing the intestinal mucosal atrophy that occurs from prolonged fasting, which makes patients vulnerable to bacterial infections [21,22].

In general, Slfns are very distantly related to other protein families, creating a fascinating bioinformatics challenge. Searching for protein sequence homologs of hSLFN12 with multiple rounds of PSI-BLAST [23] only identifies Slfns or uncharacterized proteins. Understanding the structure and function of Slfns has been complicated by inconsistent and unsupported annotations that have propagated through databases and the scientific literature. InterPro uses the Pfam conserved domain sequence signature for the Alba_2 nucleic acid binding domain (PF04326; http://pfam.xfam.org/family/Alba_2) as the basis to detect schlafen AAA domains (<https://www.ebi.ac.uk/interpro/entry/IPR007421>). On the other hand, InterPro detects members of the Slfn superfamily (IPR029684) by using the PANTHER conserved sequence signature derived from a selected set of sequences annotated as Slfns (<http://www.pantherdb.org/panther/family.do?clsAccession=PTHR12155>). The term schlafen AAA domain is assigned by InterPro to the sequence-conserved N-terminal region between residues 200–300 in all the human and rodent schlafens (<http://www.ebi.ac.uk/interpro/entry/IPR029684/proteins-matched>). The intentionality of using an Alba_2 domain signature to detect Slfn AAA domains was confirmed by correspondence with InterPro (interhelp@ebi.ac.uk; February 1, 2018), although Alba_2 and AAA domains are incompatible in structure and function (see section 3.1). References or data are not provided in InterPro to support the designation of this region in Slfns as a AAA domain or the

accompanying description of AAA function in Slfns: “AAA ATPases form a large, functionally diverse protein family belonging to the AAA+ superfamily of ring-shaped P-loop NTPases”. InterPro description of this region as a AAA domain has carried over to the Pfam database, which is the basis for domain annotation by many other bioinformatics servers. For instance, the Pfam entry for the Alba_2 family (http://pfam.xfam.org/family/Alba_2), which is known to be active in nucleic acid binding [24,25], provides descriptive text from the InterPro entry for Slfns (IPR007421), incorrectly describing Alba_2 as a family of AAA ATPases. For individual proteins (e.g., <https://www.uniprot.org/uniprot/Q8IYM2>), the UniProt database also refers to the incorrect InterPro annotation as the basis for stating that Slfns have “ATP binding” function.

Similarly, a review of the Slfn literature identifies a number of papers that mention the presence of AAA domains without providing supportive evidence or that cite another paper(s) mentioning AAA domains or ATPase activity in the N-terminal region of Slfns (often described as “adjacent to the schlafen box sequence”) without supportive evidence [1,3,6,26–31]. In other publications, authors note the well-documented presence of Walker motif-containing ATPase domains that occur only in the C-terminal domain of the helicase-like group III Slfns [4] as the basis for inferring the presence of AAA domains adjacent to the Slfn box sequence in the N-terminal region of Slfns, despite this being a distal and nonhomologous part of the protein [11,32–34]. Three papers include evidence for the presence of a divergent AAA_4 domain in the N-terminal region in Slfns, citing the Pfam04326 motif as the basis for identifying the AAA_4 domain [16,35,36]. The expectation value provided in one reference, 2.75×10^{-08} , indicates a very low probability of this sequence match occurring at random [36]. However, the Pfam04326 motif detects Alba_2 nucleic acid binding domains, not AAA ATPase domains. This misannotation derives from the incorrect AAA domain description accompanying the Alba_2 sequence signature in InterPro, which is repeated in Pfam and UniProt. Geserick et al. identify a confident match to a AAA domain in the C-terminal region of group III Slfns, with a conserved Walker ATPase motif, and infer that these group III Slfns belong to a new subfamily of RNA-dependent ATPases in helicase superfamily I [4]. Their identification of a C-terminal AAA domain in group III Slfns is well-supported and uncontroversial; the confusion occurs when AAA domains are postulated to occur in the N-terminal and central regions of Slfns without supporting evidence.

Here, we use structural bioinformatics, supported by recent high-resolution structural and activity data on a close relative, rSlfn13 [16], augmented by experimental data on four binding partners, to predict hSLFN12 architecture, conserved motif roles, and molecular function in partnership with proteins and nucleic acids. The identity of an Alba_2 versus AAA domain in the N-terminal region of all Slfns is addressed in detail. This analysis substantially clarifies the domain structure of the mammalian Slfns. We then discuss the implications for cell differentiation [18,19] and gene regulation, functions often ascribed to Slfns.

2. Materials and methods

2.1. Objective definition of domains and motifs in hSLFN12

To identify regions of high conservation in hSLFN12 and their relationship to binding sites, the degree of evolutionary conservation of each residue was calculated by ConSurf (<http://consurf.tau.ac.il>). ConSurf automates the identification and alignment of homologs, while sampling homologous sequences evenly across evolutionary time [37–39]. Cellular localization and the presence of localization motifs in the hSLFN12 sequence were assessed by LocTree3 [40], a module of PredictProtein (<http://open.predictprotein.org>) [41,42]. Remote homologs were detected by HHblits (<https://toolkit.tuebingen.mpg.de/#/tools/hhblits>) [43]. Protein BLAST (BLASTP) and its PSI-BLAST (remote homolog detection) and DELTA-BLAST (domain identification) variants were used to identify and align the protein sequences of homologs with hSLFN12 (<https://blast.ncbi.nlm.nih.gov/>) [23,44,45]. Sequence-independent structural superposition was performed via Dali (<http://ekhidna2.biocenter.helsinki.fi/dali/>) [46]. PDBsum (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>) [47] facilitated comparing the topology of helices and sheets in Alba and AAA domains to the crystal structure of rSlfn13, Protein Data Bank (PDB) entry 5yd0 [16]. The term Alba is used to describe the 3-dimensional fold shared by all members of the Alba family (Alba_2, Alba, Rpp20, and SpoVS; <https://pfam.xfam.org/clan/Alba>) [48], whereas the term Alba_2 is used to refer to sequence matches that are specific to the Alba_2 family. BioGrid (<https://thebiogrid.org>) [49], accessed via UniProt (<https://www.uniprot.org>) [50], was used to identify literature on experimentally-defined binding partners of hSLFN12, and to explore the partners' additional partners. The PANNZER2 server was used to retrieve additional functional annotations of the hSLFN12 sequence (<http://ekhidna2.biocenter.helsinki.fi/sanspanz/>) [51].

2.2. Three-dimensional structural modeling of hSLFN12 and hSerpinB12, to drive functional hypotheses

For modeling the structure of hSLFN12, to help define its functional domains and accessible residues, the hSLFN12 protein sequence (UniProt entry Q8IYM2) was first submitted to BLASTP to identify any structurally characterized homologs in the PDB (<http://www.rcsb.org>) [52,53]. PSI-BLAST [23] and HHblits [43] were then used to identify any remote homologs of hSLFN12 with known structure or function. To model the structure of the C-terminal region of hSLFN12, which exhibited no detectable homology to sequences of known function or PDB structures, the MUSTER structural modeling algorithm [54] was used via the Pcons structure prediction metasever (<http://pcons.net>) [55,56]. Pcons submits the user's protein sequence (target) to more than a dozen different structure prediction servers. Together, their results identify the most favorable three-dimensional (3D) structural models based on sequence similarity, favorability of the 3D environment of each residue, and compatibility of the predicted secondary structure with experimentally resolved protein structure templates. Pcons then ranks the top-scoring models according to the degree of structural consensus with the other high-scoring models.

The 3D structure of the N-terminal region of hSLFN12 was modeled by homology using SWISS-MODEL (<https://swissmodel.expasy.org>) [57]. For homology modeling, SWISS-

MODEL constructs an alignment between the target sequence and the most similar sequence in the PDB (template structure) based upon HHblits alignment. The residues conserved between the two proteins are replicated in structure between the template structure and the model of the target protein. Residue insertions and deletions in the target protein are then modeled using a structural fragment library representing similar sequences in other structures, and side chain positions are rebuilt to be overlap-free. Finally, bond lengths and angles and interatomic contacts (structural stereochemistry features) in the resulting model are regularized to fit known favorable values by using a force field within SWISS-MODEL. It was also possible to model the 3D structure of hSerpB12, a binding partner of hSLFN12, by homology. In this case, the FFAS modeling program [58] was selected, based on generating the homology model with the greatest degree of consensus with other high-scoring models, according to Pcons. Stereochemical quality checking of all models was performed by PROCHECK [59] through the PDBsum interface (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>) [47] and by using the QMEAN metric [60] in the SWISS-MODEL Structure Assessment tool (<https://swissmodel.expasy.org/assess>). All molecular structure figures were rendered by using the PyMOL molecular graphics system v1.8.2.2 (<http://pymol.org>; Schrödinger, LLC). Other figures were drawn using the macOS version of PowerPoint (v14.6.8).

3. Results and discussion

3.1. Clear definition of Alba_2 and AAA_22/P-loop NTPase domains in Slfns

Clarifying whether the N-terminal region in Slfns contains an Alba_2 or AAA domain is crucial for understanding Slfn function, because Alba_2 domains bind DNA or RNA, whereas AAA P-loop NTPases perform GTP or ATP hydrolysis associated with the energy-dependent remodeling or translocation of other molecules [61]. Both the structures and functions of Alba_2 and AAA domains differ, making the distinction important.

An Alba_2 domain match was identified for residues 200–334 in the hSLFN12 protein sequence and the corresponding N-terminal region in all mammalian Slfns with the Conserved Domains Database sequence search utility (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>). CDD detects Alba_2 domains using the Pfam Alba_2 signature (<https://pfam.xfam.org/family/PF04326>), yielding an expectation value of 9.2×10^{-15} for the Alba_2 sequence match in hSLFN12 and a domain-match expectation value of 2.9×10^{-10} . These values indicate an extremely low probability of occurring at random and very high confidence that this region in hSLFN12 is an Alba_2 domain. The sequence match between hSLFN12 and the representative sequence for Alba_2 domains (Figure 1AB) underscores their close correspondence and is compelling because the highly conserved residues in the Alba_2 domain match the highly-conserved Slfn box residues. The Slfn box structural correspondence to an Alba_2 nucleic acid binding domain has not been recognized in the literature. Confusingly, InterPro describes the Alba_2 residue range in the hSLFN12 sequence and its murine and human homologs as a schlafen AAA domain (IPR007421) and a schlafen AAA domain superfamily member (IPR038461), even though InterPro detects this domain by using the Pfam signature for the Alba_2 nucleic acid binding domain (PF04326).

Because N-terminal AAA domain misannotation and confusion persists, we analyzed the N-terminal regions of Slfns in detail to address: could AlBA_2 and AAA domains be compatible in sequence and structure? The answer is no, based on the following analysis. Firstly, only AlBA_2 and no AAA or ATPase domain sequence matches were found in the N-terminal half of the 14 murine and human Slfns analyzed (Figure 2), after scanning for more than 40 types of AAA domains in the Pfam motif database (<http://pfam.xfam.org>). Secondly, the AlBA_2 and AAA domain structures have different topologies of α helix and β sheet connectivity and differ in having antiparallel versus parallel β sheets and two versus six helices, resulting in different tertiary structures (Figure 1CD). Thus, a sequence region cannot simultaneously be an AlBA_2 domain and a AAA domain. Most compellingly, the N-terminal domain present in the recent crystal structure of rSlfn13 (PDB entry 5yd0) [16] is one of the three structural instances of AlBA_2 domains in the PDB (<https://pfam.xfam.org/family/PF04326>), the other two being Slfn structural homologs of unknown function (PDB entries 3lmm and 2kyy). The AlBA_2 domain structure matches the ribonuclease (RNase) site in the N-terminal half of rSlfn13 (PDB entry 5yd0), which is also detected by the Pfam AlBA_2 domain signature. This AlBA_2 domain also occurs in the corresponding residues of hSLFN12, based on its homology with rSlfn13:

AlBA_2 β strand 1: hSLFN12 residues 203–207 (rSlfn13: 208–212)

AlBA_2 α helix 1: hSLFN12 residues 214–231 (rSlfn13: 218–237)

AlBA_2 β strand 2: hSLFN12 residues 235–239 (rSlfn13: 240–244)

AlBA_2 α helix 2: hSLFN12 residues 257–270 (rSlfn13: 262–276)

AlBA_2 β strand 3: hSLFN12 residues 284–293 (rSlfn13: 291–299)

AlBA_2 β strand 4: hSLFN12 residues 299–308 (rSlfn13: 304–313)

There is no structural match to a AAA domain in the rSlfn13 crystal structure, which covers the N-terminal 327 residues of the protein.

There is borderline sequence similarity over a short region between some pairs of known AlBA_2 domains and known AAA domains. Thus, predicting domain type based on low sequence identity over a short region can lead to misidentification of AAA domains in sequences that actually form AlBA_2 domains. For instance, one-third (47 residues) of the mSlfn10 AlBA_2 domain sequence is 28% identical with a subset of the AAA_22 domain residues in mSlfn8, and this AlBA_2 region is incorrectly annotated by InterPro as a schlafen AAA domain. Thus, loose domain detection signatures and overly permissive sequence matching, plus the absence of a Slfn crystal structure for comparison until recently, may have led to the “schlafen AAA domain” misconception of AlBA_2 regions. False matches can be avoided by using Pfam motif searching (<http://pfam.xfam.org>), which uses stringent patterns for both AlBA_2 and AAA domains. We also found that alignment between pairs of true AlBA_2 domain sequences in Slfns always yielded 38% identity across 75% or more of the sequence, while alignment between true AAA_22 domains (occurring in the C-terminal region of group III Slfns) always resulted in 62% identity across 95% or more of

the sequence. Thus, setting a suitably high threshold for identity, and requiring coverage over at least three-quarters of the domain, also rules out false positive domain matches. Residue ranges are summarized in Figure 2 for the AlbA_2 domains in group I-III murine and human Slfns, and for the AAA_22/P-loop NTPase domains occurring in group III Slfns only. The nucleic acid binding significance of the AlbA_2 domain is consistent with experimental confirmation that mSlfn8, hSLFN11, rSlfn13, hSLFN13, and hSLFN14 are ribonucleases [16,17]. Two glutamic acid residues in the AlbA_2 domain of rSlfn13 have been shown to be catalytically essential for its 3'-endonuclease function [16].

3.2. 3D structural modeling and domain interpretation of hSLFN12 and hSerpinB12

The N-terminal region of hSLFN12 (residues 13–346), showed 38% sequence identity (with a highly significant BLASTP expectation value of 7×10^{-69}) with residues 28–366 of the crystal structure of rSlfn13 (PDB entry 5yd0) [16]. This allowed the N-terminal structure of hSLFN12 to be modeled confidently by homology with the SWISS-MODEL software, using the rSlfn13 N-terminal crystal structure as the template. In cases like this, where a protein sequence to be modeled has at least 25% sequence identity over at least 80 residues with an experimentally defined protein structure such as rSlfn13, homology modeling is a quantifiably reliable technique, yielding a structural model accurate to within 2.5 Å RMSD for main-chain atom positions [64]. The structure and function of hSLFN12's C-terminal domain (residues 347–578) and the corresponding region in its homolog rSlfn13 have not been defined experimentally. For this region in both hSLFN12 and rSlfn13, there were no significant HHblits or BLASTP matches in either the BLAST non-redundant protein sequence database or the PDB with an expectation value < 1 (required for confidence that a match scores much better than a random alignment). Homology modeling of the C-terminal region in hSLFN12 was therefore not an option. However, structural modeling with the MUSTER (multi-sources threader) fold recognition software [54] through the Pcons server [55,56] identified the entire catalytic domain of the GTPase, dynamin (PDB entry 1jx2) [65], as the 3D structure with the greatest sequence, residue contact, and secondary structure compatibility with the C-terminal domain of hSLFN12. The dynamin-based C-terminal model for hSLFN12 generated by MUSTER also showed the highest degree of consensus with structural predictions from other fold recognition methods invoked by Pcons. The rho/ras family of small GTPases, including the corresponding domain in dynamin, differs in structure and function from AAA domains [66,67]. A related rho family small GTPase from an amoebaflagellate (UniProt entry D2VNU5_NAEGR), bears significant sequence identity (33%, expectation value of 1×10^{-05}) with residues 378–446 in the C-terminal domain of hSLFN12, which surround the rho GTPase nucleotide binding site. The *Naegleria gruberi* amoebaflagellate is an early eukaryote containing the ~4 300 genes of the last eukaryotic common ancestor [68].

The 33% sequence identity between the *Naegleria* GTPase and the hSLFN12 C-terminal domain is 5% above the threshold required for main chain structural homology between two sequences of this length [64]. The amoebaflagellate GTPase also matches the Pfam sequence signature for ras GTPases (PF0071) and is listed as an ortholog of hSLFN12 in the KEGG Orthology database (<https://www.genome.jp/kegg/ko.html>) [69,70]. The PDB homolog closest in sequence to the amoebaflagellate GTPase, chain D in PDB entry 1i4d (with 59%

identity over the entire sequence; expectation value of 4×10^{-78}), is a Rac rho GTPase that closely overlays with the PDB 1jx2 structure used to model the C-terminal domain of hSLFN12. Thus, a rho GTPase-like structure is likely for this domain in hSLFN12. DELTA-BLAST [45], which is optimized to detect homologous domains, indicates the amoebaflagellate organism also contains three proteins matching the hSLFN12 Alba_2 domain, the third of which is annotated as maintaining chromosome structure (sequence identifiers XP_002670686.1, XP_002680060.1, and XP_002671375.1) [71].

Overlaying the hSLFN12 C-terminal domain model and the dynamin structure bound to GDP (PDB entry 5d3q) with Dali [46] indicated that only a subset of the dynamin residues in contact with GDP [72] are conserved in hSLFN12. However, both the dynamin and hSLFN12 sites are lined with polar residues, and the binding site shape is conserved, as described below. Thus, molecules binding in this pocket in hSLFN12 are likely to be similar in size and polarity to GTP.

Human serpinB12 (hSerpB12; also known as yukopin) became of interest to the hSLFN12 story when it was discovered that hSLFN12 co-immunoprecipitates with hSerpB12 and also stimulates the differentiation of nutrient-absorbing enterocytes in the human gut [19]. At the molecular level, overexpression of either hSLFN12 or hSerpB12 increases the expression of sucrase-isomaltase and villin, two hallmarks of enterocyte differentiation [19]. hSerpB12 is a member of the ovalbumin-like serine protease inhibitor family expressed in many tissues, including brain, bone marrow, lymph node, heart, lung, liver, pancreas, testis, ovary, and intestine [73]. A complex involving both hSerpB12 and hSLFN12, which shows no protease-like features, was a mystery, as hSerpB12 has been characterized as an intracellular inhibitor of trypsin, plasmin, and cathepsin G [73]. Thus, we modeled and analyzed the structural features and functional relevance of hSerpB12 for contributing to cell differentiation in complex with hSLFN12 and its other partners. The structure of hSerpB12 was modeled based on its significant homology to myeloid and erythroid nuclear termination stage-specific protein (MENT), a serine protease active in chromatin remodeling. hSerpB12 and MENT were found to be 43% identical over the full-length (405-residue) protein, with a highly significant BLASTP expectation value of 7×10^{-111} . The FFAS software using the MENT crystal structure as a template (PDB entry 2h4r) [74] generated the highest-quality structural model, according to Pcons.

Evaluation of the quality of the modeled structures was performed with PROCHECK [59] and the SWISS-MODEL Structure Assessment tool [57,75]. SWISS-MODEL's QMEAN score evaluated the favorability of all-atom contacts, C_{β} contacts, solvation (favorability of the exposure of each residue to its hydrophobic or hydrophilic environment), and bond torsions (assessing whether bond configurations are strained). Table 1 summarizes the modeling statistics. The overall stereochemistry of all three structural models was good, according to the PROCHECK G-factor (0) and the percentage of residues with favorable main-chain dihedral angles (85%). According to SWISS-MODEL's QMEAN score, the N-terminal domain of hSLFN12 and the hSerpB12 model have similar characteristics to crystal structures, while the C-terminal domain model of hSLFN12 scores at the extreme edge of the crystal structure distribution (less favorable). The 3D structural models of all

domains of hSLFN12 and hSerpinB12 are shown in Figure 3, and their PDB-formatted atomic coordinate files are provided in Supplementary Data 1–3.

Structural modeling provides a valuable spatial context for interpreting residue conservation data and functionally important residues, as well as guiding the design of experiments to test interactions with molecular partners. A residue can be highly conserved due to being buried in the core of a domain, occurring at a domain or subunit interface, or contributing to a molecular binding or catalytic site. The 3D model and contacts made in homologous complexes suggest which of these roles is most likely. To identify conserved motifs and residues important for structure and function in hSLFN12, we employed ConSurf, which selected and aligned the protein sequences of 150 Slfn homologs from diverse organisms. For each sequence position, ConSurf assigned a value from most-conserved to most variable across the aligned Slfn sequences. The most conserved regions (Figure 4), including the Slfn box and SWADL motifs, were analyzed in the context of the hSLFN12 structural model (Supplementary Data 1 and 2), yielding the following observations.

The hSLFN12 Alba_2 region and a subset of its Slfn box residues are likely to bind RNA—A highly conserved surface patch is formed by residues R129, T132, S133, F198, E200, S201, E205, K207, Y225, A228, F229, N231, and G234 (labeled as conserved RNA binding region in Figure 3A). These residues are entirely conserved with rSlfn13 and include several schlafen box residues in the Alba_2 domain ((221)ILPQYVSAFANTDGGYLFIGN). Based on high conservation with rSlfn13, including two glutamic acid residues shown to be critical for RNase activity (Figure 4), this region in hSLFN12 is likely to form an RNA binding site and may function as a ribonuclease [16], similar to mSlfn8, hSLFN11, rSlfn13, and hSLFN14. In the recently sequenced genome of a vertebrate precursor, the Mediterranean amphioxus or lancelet (*Branchiostoma lanceolatum*) [77], a single Slfn protein is present (SLFN1; accession code BL17146_evm9 in <http://amphiencode.github.io>). The region of amphioxus Slfn match with mammalian Slfns is the Alba_2 domain, including conservation of the Slfn box and the RNase catalytic glutamic acid residues. Amphioxus SLFN1 shows 29% identity with residues 196–264 of hSLFN12 (expectation value of 0.096). This suggests that the Alba_2 domain and its RNA binding/cleaving residues comprise the most ancient and characteristic region within vertebrate Slfns.

Is hSLFN12 a monomer or dimer?—Proteins in the Alba superfamily are typically dimeric [78] and have highly conserved dimer interface residues [79]. The RNaseE structure (PDB entry 5f6c) that matches the RNase site in rSlfn13 and hSLFN12 [16,80] forms a dimer via its Alba domain [81]. hSLFN12 may dimerize similarly, despite its homolog, rSlfn13, being described as a monomer based on small-angle light scattering [16]. The Slfn box motif, centered on the sequence SAFANTDGGYLFIGN in the hSLFN12 Alba_2 domain, is one of the most highly conserved Slfn regions (Figures 1 and 4). Residues from the Slfn box contribute to an unusually hydrophobic surface patch in the 3D structure that would be energetically unfavorable if solvent-exposed, suggesting they are conserved to form a protein interface such as the dimer interface in RNase E; these residues are marked “potential dimer interface” in Figure 3A. The residues conserved in the Slfn box that lack

surface exposure are likely to be important for structural integrity rather than binding other molecules.

hSLFN12 can be modeled as an RNaseE-like (rSlfn13 homolog) dimer, by performing main-chain superposition of the hSLFN12 monomer on each chain in PDB entry 5f6c with Dali [46], as shown in Figure 3B. Dimerization requires repositioning the (155)RLYLRPPELLAKRPCVDIQEENNMKAL loop in hSLFN12 (Figure 3A), which is not homologous with rSlfn13 and was modeled by SWISS-MODEL in an arbitrary conformation. The following highly conserved and hydrophobic hSLFN12 residues are central to the predicted dimer interface formed between two β sheets, one from each monomer: Y236, F238, I247, V302 and A304. Mutation of Y236 to Phe reduces hSLFN12 activity by 40% [19]. CCHC zinc fingers also frequently act as dimerization motifs [82], making the CCHC zinc finger region in hSLFN12 (Figures 3A and 4) an additional candidate for a dimer interface in hSLFN12. When modeled as a dimer, the conserved RNA binding region forms an arc-shaped site large enough to bind single- or double-stranded nucleic acid (Figure 3B). Whether hSLFN12 proves to be monomeric or dimeric in solution, the conserved polar residues listed above are expected to be important for RNA binding. Specific residues contributing to the positive patch that interacts with RNA in rSlfn13 are marked + in Figure 4, while the RNase catalytic glutamic acids are marked #.

Conserved polar pocket—A second conserved surface patch in schlafens forms a polar pocket spatially proximal to the C-terminus of the N-terminal domain in hSLFN12, labeled as a conserved polar region in Figure 3A. This site could bind a nucleotide or polar segment of a larger molecule and is formed by the N53, S54, G55, G57, F106, V107, K108, S109, and W110 residues from the (53)NSGGG and (106)FVKSW epitopes. The (49)CALLNSGGGVI sequence in this region is the second most highly conserved region in Slfns, forming a helix-turn-strand motif in the 3D structure in which five of the residues, CALL and I, are buried.

Conservation of the CCHC/gag knuckle zinc finger—The histidine and cysteine residues in the (275)HFC and 310(FCCAVFA) motifs are highly conserved in Slfns, including hSLFN12. These His and Cys side chains are the four Zn ligands in a surface-exposed zinc finger observed in the rSlfn13 structure. This CCHC zinc finger, also known as a gag knuckle [83], binds ssRNA in some proteins [84]. Alternative possibilities for the Zn finger are forming a protein-protein interface or interacting with DNA [82,83].

A buried conserved motif—The conserved sequence (300)GYVCALRV forms a buried β strand.

Hydrophobic strip and loop regions are potential protein-protein interaction sites—The N-terminal and C-terminal domains of hSLFN12 lack any large hydrophobic surface patches as clear candidates for protein-protein interfaces. However, a narrow hydrophobic spiral around the N-terminal domain consists of A12, L14, V15, L16, A47, V58, I75, L87, L88, F89, V90, P91, L94, F96, M97, L104, P106, W322, V329, and M330, which could bind to a hydrophobic strip or flexible tether on another molecule. The disordered region (155)RLYLRPPELLAKRPCVDIQEENNMKAL is also unusually

hydrophobic for a surface loop, with 50% hydrophobic residues. The first three-quarters of these residues are absent from the rSlfn13 structure due to mobility. Thus, these residues are expected to be flexible and may interact with another molecule.

Role of the SWADL motif—(427)SWSVDL, the SWADL motif in hSLFN12, is buried rather than surface accessible, under the GDP binding site in the dynamin structure on which the C-terminal domain was modeled. The terminal VDL residues are identical between hSLFN12 and dynamin. Others have noted the SWADL motif appears to be Slfn-specific and is found in group II and III Slfns [11]. Across the group II-III Slfns shown in Figure 2, the SWADL motif in the middle of the GTPase-like domain matches the sequence pattern SW(A,S)(V,G,L)D(L,I,V) in all fourteen.

A second conserved motif in the C-terminal domain—(440)VLCDAL is the most conserved Slfn motif in the C-terminal domain; however, it is dissimilar in sequence to the corresponding region in dynamin. This region is predicted to be a mostly surface-exposed helix that clamps over the GDP/GTP site in dynamin.

Two other C-terminal regions highly conserved among Slfns—

(479)QKLAKIGGYTKKVCV is part of a helical turn-omega loop- β strand motif that is surface-exposed except for the terminal VCV residues. (530)LLKALFKAL is mostly exposed and part of a helix, as well as being unusually hydrophobic. This suggests it may interact with another protein.

hSLFN12 functions in the nucleus and cytoplasm—hSLFN12 binds two nuclear proteins, SUV39H1 and JMJD6 [85], and therefore is likely to function in the nucleus as well as the cytoplasm [19]. hSLFN12 is predicted to be nuclear by LocTree [40], with a nuclear localization sequence present near the amino terminus: (28)RKKMKDCKLRKK. All proteins of known cellular localization matching this [RK] \times [RK] \times [KR] \times {4,6}RKK motif are nuclear.

Conserved regions in hSerpB12—To identify functionally important regions, analysis of conserved residues was also performed for 150 homologs of hSLFN12's partner, hSerpB12, with ConSurf (Figure 5). When mapped onto the structural model of hSerpB12 (Supplementary Data 3), the most conserved surface patch among these serpins included residues N27, K169, I170, Y192, F193, K194, W197, F201, T206, V221, Y247, F302, T303, E305, K347, E351, V352, D353, E354, and P405. This corresponds to the breach and shutter regions in the main β sheet (Figure 3E), which facilitate structural transitions between the free and inhibitory conformations of serpins [86]. There were no hydrophobic surface patches of any significance on hSerpB12, and thus no obvious candidates for direct protein-protein interfaces aside from the reactive center loop, which inhibits proteases. The M-loop and reactive center loops are determinants of individual serpins' selectivity for their molecular partners, which differs between serpins, and thus these loops show less sequence conservation.

3.3. Schlafen12 is predicted to bind and possibly cleave RNA

Like its recently characterized homolog, rSlfn13 [16], hSLFN12 is likely to either bind RNA or act as a 3' RNA endonuclease. This prediction is based on conservation of the RNA binding residues and overall homology with rSlfn13 (Figure 4). The glutamic acid residues essential for rSlfn13 endonuclease activity are conserved in hSLFN12 (E205 and E210 in rSlfn13, corresponding to E200 and E205 in hSLFN12). Four of the six positive patch residues interacting with RNA in rSlfn13 (K38, R39, K42, R217, K224, and K276) [16] are conserved as positive residues in hSLFN12 (R28, K29, K219, and K270), with other positively charged residues present nearby. Yang et al. noted that RNA-cleaving schlafens mSlfn8, hSLFN13, and rSlfn14 all have higher concentrations of positive residues than the RNase rSlfn13 [16,17], whereas mSlfn1 and hSLFN5 have fewer positive residues and do not cleave tRNA (Suppl. Fig. 6 in [16]). hSLFN12 has a higher concentration of positive residues than the RNaseE region in rSlfn13 (residues 126–317 in PDB entry 5yd0), with 19 Arg and 33 Lys residues in hSLFN12 compared to 16 Arg and 31 Lys in rSlfn13. This suggests that hSLFN12 has the positive surface and catalytic Glu residues necessary to support RNA binding and cleavage.

A panel of related functions were provided by the PANNZER2 server [51], given only the sequence of hSLFN12. The following Gene Ontology [87,88] molecular functions had a confidence value of 0.40 or higher (where 1.0 indicates maximum confidence): adenylyl ribonucleotide binding (0.55), drug binding (0.54), purine ribonucleoside triphosphate binding (0.54), ribosome binding (0.42), endoribonuclease activity (0.41), and tRNA binding (0.40). The top 100 matches to hSLFN12 identified by remote homology detector HHblits [43] included a number of predicted transcriptional regulators with up to 42% sequence identity to the hSLFN12 AlbA_2 domain (e.g., UniProt entry F8EDI5).

3.4. Could schlafen12 participate in a gene-silencing or RNA-splicing complex?

Aside from co-immunoprecipitating hSerp1B12, hSLFN12 binds two nuclear RNA-binding proteins, SUV39H1 and JMJD6 [85]. SUV39H1 promotes the formation of heterochromatin in repetitive, non-coding DNA sequences through its histone methyltransferase activity [89] and is required for silencing S-phase genes during the terminal differentiation of cells [90], consistent with hSLFN12's role in enterocyte differentiation. The relevance of RNA binding by SUV39H1 was recently discovered. Low-level transcription can occur for constitutive heterochromatin regions in DNA that must be silenced for viability. The RNA transcripts from repetitive DNA are found to bind SUV39H1 in human cells, which tethers SUV39H1 to the repetitive DNA to silence it [91].

The other known nuclear partner of hSLFN12, JMJD6 (previously called Ptdsr), contains a jumonjiC enzymatic domain that hydroxylates histone lysine residues [92], inhibiting their subsequent methylation or acetylation. JMJD6 possibly also demethylates arginine residues [93–95]. Furthermore, JMJD6 binds an E3 ubiquitin ligase, UHRF1, that recognizes histone modifications, ubiquitinates histone 3, and maintains DNA methylation to block gene expression [92]. Proposed other roles for JMJD6 are promoter-proximal pause release of polymerase II (pol II), and pre-mRNA splicing [96]. JMJD6 binds single-stranded RNA but not DNA. Like SUV39H1, JMJD6 associates with nascent RNA transcripts in the nucleus

[97–99]. Its function is necessary for normal differentiation of intestinal and other tissues in mice [100]. Promoter-proximal pause release of pol II, one of the functions associated with JMJD6, is important for coordinating the transcription of genes during differentiation [101].

Transcriptional gene silencing coupled to pre-mRNA splicing is attributed to SUV39H1 in complex with argonaute (Ago) proteins [102], as follows. Chromatin modifying proteins are recruited by Ago complexes bound to double-stranded or microRNA, leading to repressive chromatin modifications such as histone 3 lysine 9/lysine 27 (H3K9/K27) methylation, which inhibit transcription at the promoter targeted by the RNA. The process is similar for small RNA-mediated alternative splicing, in which the Ago may be guided by single-stranded small RNA (possibly antisense transcripts) to chromatin at the 3' ends of alternative exons. Binding to additional chromatin components, such as the chromodomain protein HP1, then leads to repressive trimethylation of H3K9 by SUV39H1 and a second lysine methyl transferase. This slows pol II elongation, facilitating spliceosome recruitment to incorporate the alternative exon. Together, the roles of SUV39H1 and JMJD6 support their participation in an RNA binding complex that regulates gene transcription and possibly also contributes to pre-mRNA splicing.

How are Slfns and argonautes connected? The immediate precursors of eukaryotes, the Asgard group of archaeobacteria [103], contain a single protein chain in which an Alba_2 domain with a Slfn box in the N-terminal region is fused to a C-terminal, 400-residue Ago sequence [104]. The evolutionary biologists studying these archaeobacteria annotated this sequence as a Slfn-Ago fusion, and this sequence clusters together with archaeal Agos participating in DNA or RNA-guided DNA silencing to protect against DNA incorporation from other organisms [105]. This is reminiscent of the roles some mammalian Slfns are known to play in preventing viral DNA from being integrated into host cells, as discussed in the Introduction. A protein BLAST alignment of the *Heimdallarchaeota archaeon* Slfn-Ago fusion protein (GenBank OLS20375.1) shows it is significantly similar in the Slfn box and Alba_2 region with hSLFN12 (residues 191–252 are 29% identical; expectation value of 9×10^{-5}). For comparison, mammalian hSLFN12 and rSlfn13 show 38% identity in their N-terminal regions. Slfn and Ago proteins both form horseshoe-like structures that wrap around RNA (e.g., PDB entry 5yd0 for rSlfn13 and PDB entry 4krf for human Ago1). Our BLAST analysis also identified homologous proteins with 39–42% identity to the Asgard *Heimdallarchaeota archaeon* Slfn Alba_2-Ago sequence (GenBank OLS20375.1) in a number of other bacterial families from diverse marine, freshwater, soil, and animal environments: cyanobacteria, actinobacteria, cytophagaceae, proteobacteria, and planctomycetes/verrucomicrobia/chlamydiae (representative sequence identifiers from GenBank: WP_075598738.1, AHY48436.1, WP_081144923.1, WP_112072897.1, and RKU28595.1). In some archaea, the Slfn-homologous Alba_2 domain is alternatively fused to sirtuin2 (SirT2) [104], a nuclear H4K16 deacetylase that promotes H4K20 methylation and chromatin compaction [106]. The prevalence of these sequences across bacterial and archaeobacterial families, and their similarity to mammalian schlafens in the most conserved Slfn box-Alba_2 region, suggest that Slfn or Slfn-like proteins occur in archaeobacteria and bacteria as well as vertebrates.

Given Slfn RNA binding, its partnership with chromatin remodeling proteins, and the ancestral covalent connection between Slfns and Agos, could they together form a transcriptional silencing complex? Three of hSLFN12's binding partners are homologs of components in the RNA-induced transcriptional silencing-Clr4 cullin ring E3 ligase (RITS-CLRC) complex in fission yeast, suggesting they partner with an Ago to form an analogous complex. In fission yeast, as in mammals, siRNA associates with Ago1, targeting the complex to nascent RNA transcripts, particularly those of long non-coding DNA. Human SUV39H1 is homologous to the yeast H3K9 methyltransferase, Clr4, while human JMJD6 is homologous to the yeast JmjC-domain-containing demethylase, Lid2, which hypomethylates H3K4 to stabilize heterochromatin [107–110]. An additional hSLFN12 binding partner is human PDLIM7 (also known as ENIGMA), identified by affinity-capture mass spectrometry [111]. PDLIM7 is a positive regulator of osteoblast differentiation (GO: 0045669) and cardiac development, and is proposed to function in protein-protein interactions in gene transcription and the cytoskeleton via its LIM and PDZ domains [87,111–113]. The correspondences between human SUV39H1, JMJD6, and PDLIM7 with the fission yeast CLRC components Clr4, Lid2, and Stc1, respectively, are shown schematically in Figure 6. The CCHC zinc finger in hSLFN12 could form an interface with PDLIM7, since CCHC zinc fingers are common interfaces with LIM proteins, or with JMJD6, which binds other zinc finger-containing proteins [85,111].

The LIM-domain homolog of PDLIM7, Stc1, is key for recruiting the Clr4 CLRC complex to the Ago1 RITS complex in fission yeast, resulting in Clr4 methylation of H3K9 in cognate chromatin to promote heterochromatin formation [114]. Homologs of the yeast Ago1:Lid2:Clr4 RITS-CLRC complex have been characterized in higher organisms, including *Drosophila* [107], and the corresponding mechanisms are being dissected in mammals [115]. DDB1 is the human homolog of the yeast CLRC linker protein, Rik1. In complex with cullin4 and SUV39H1 [116], DDB1 forms a cullin4 ringbox1 E3 ubiquitin ligase similar to the yeast CLRC, and is vital for transcriptional control and antiviral activity [117]. The DDB1 CLRC ubiquitination targets are regulators of chromatin formation, development, the cell cycle, and cell proliferation [118]. The DDB1 CLRC and the human UHRF1 E3 ligase bound by JMJD6 have similar roles, since both ubiquitinate histones and recruit chromatin remodeling enzymes [117]. E3 ligases containing cullin4, including the one formed by DDB1, are known targets for hijacking by viruses to evade recognition and support viral propagation [118], connecting them with Slfns' viral defense function. The helicase region in group III Slfns has been shown to act as a virus sensor that protects cells against viral replication [1,3,5,6].

3.5. The GTPase-like C-terminal region of hSLFN12 is structurally related to MxB, which has Slfn-like antiviral activity

The region in Slfns corresponding to the hSLFN12 C-terminal domain, which follows the Alba_2 domain in all group II-III Slfns (Figure 2), has been the most mysterious region, not yet characterized structurally or functionally by others. The entire catalytic domain of the dynamin GTPase was found to be the structure showing the most primary, secondary, and tertiary structural compatibility with the C-terminal domain of hSLFN12 (section 3.2). What are the functional implications of a dynamin-like domain in Slfns? A candidate function is

preventing the integration of viral DNA into chromatin. Myxovirus resistance factor MxB (also known as Mx2) is a GTPase related to dynamin whose structure matches the C-terminal domain in hSLFN12. Similar to Slfns, MxB has potent antiviral activity in mammalian cells against HIV-1 and other RNA viruses by interacting with the capsid of HIV-1 and preventing chromosomal integration of viral DNA [6,119–121]. MxB protein is found in both the cytoplasm and nucleus, where it forms a granular pattern in the heterochromatin underlying the nuclear envelope [122,123]. Like the roles of MENT and archaeal Alba domains in heterochromatin formation [25,74,124], MxB is observed to form higher-order oligomeric arrays [125], as observed in its cryo-electron microscopy structure (PDB entry 5uot). The GTPase domain in MxB has 27 surface-exposed Lys and Arg residues, and the GTPase-like domain in hSLFN12 is similarly positively charged, with 31 surface-exposed Lys and Arg residues concentrated in the strip 1 and 2 regions (Figure 3D). Dynamin is also observed to bind RNA, with unknown biological implications [126]. A nuclear as well as cytoplasmic role for the C-terminal domain of hSLFN12 is consistent with the nuclear activities of the MxB GTPase and the observation from cofractionation experiments that JMJD6 binds dynamin-like proteins [127]. By moving between the nucleus and cytoplasm, the dynamin-like domain in hSLFN12 could shift between nuclear RITS and cytoplasmic cell-shaping roles (as described in Supplementary Data 4), both of which are essential for differentiation. The GTPase-like domain in hSLFN12 is found in all group II-III Slfns, based on their exhibiting 39% or greater identity to at least 142 residues in the hSLFN12 GTPase-like domain (Figure 2).

Interestingly, a significant viral protein match to Slfns has been found in camelpox virus. This protein is called 176R or v-slfn (UniProt entry Q775M44), and homologs have been found in a number of orthopoxviruses including cowpox, ectromelia, monkeypox, and taterapox [26,128]. The 176R v-slfn is 36% identical to residues 35–338 of hSLFN12 (expectation value of 9×10^{-54}), corresponding to the Slfn box-Alba_2 N-terminal region. The authors found that v-slfn does not inhibit mammalian cell growth when stably expressed. They propose that v-slfn is a virulence factor, since vaccinia virus administered intranasally in mice shows attenuated virulence when v-slfn expression is heterologously expressed in the virus, relative to the wild-type virus. However, reduction in virulence is not observed in intradermal administration of the same v-slfn expressing virus. The observed attenuation in intranasally administered vaccinia virus could result from heterologous v-slfn expression destabilizing the virus structure or its interactions with the host cell. Because v-slfn does not contain the MxB-like GTPase-like domain found in group II Slfns, it is not an obvious candidate for mimicking or competing with the hypothesized role of the GTPase-like domain in blocking the integration of viral DNA. However, given that v-slfn contains the Alba_2 region, which is proposed to act as a transcriptional regulator in hSLFN12 and a translational regulator in rSlfn13, one possibility is that v-Slfn competes with mammalian Slfn Alba_2 domains' inhibition of transcription or translation. This would free the cellular machinery to transcribe or translate viral proteins. Alternatively, the v-slfn could transcriptionally or translationally upregulate or downregulate other regions in the DNA or RNA. Given that Slfns are not associated with cell proliferation/growth, but rather with cell cycle pausing and differentiation, v-slfn expression could also influence the differentiation of mammalian cells.

3.6. hSerpinB12 participation in hSLFN12 function

Several observations suggest that direct or indirect interaction with hSerpinB12 is important for hSLFN12 function. hSLFN12 is observed to co-immunoprecipitate hSerpinB12, and hSerpinB12 and hSLFN12 both stimulate the expression of markers of enterocyte differentiation [19]. Two mutations (D233A and Y236F) in the Slfn box region of the AlbA_2 domain (ILPQYVSAFANTDGGYLFGLNE) not only inactivate hSLFN12 but also prevent hSLFN12:hSerpinB12 co-immunoprecipitation. hSLFN12 effects are also inhibited by reducing hSerpinB12 expression via synthetic siRNA [19]. hSerpinB12 is otherwise known to inhibit serine proteases trypsin, plasmin, and cathepsin G [73], and the serpin family is known to be involved in blood, immune, and cartilage cell differentiation [73,129,130]. Could hSerpinB12 regulate cell differentiation in partnership with hSLFN12? The homologous serpin MENT concentrates in heterochromatin and is the major non-histone chromatin protein in nuclei [131]. MENT is the closest homolog of hSerpinB12 in the PDB (43% identity over the entire protein) and the basis for modeling the hSerpinB12 structure. Using the AT-hook positively charged motif (PXRGRP) in its M-loop (also known as the CD loop, linking helices C and D in serpins), MENT binds two linker strands of DNA as they exit mononucleosomes and oligomerizes to form a homopolymer that compacts chromatin. Its oligomerization mechanism involves adding the MENT reactive center loop as a β strand to the edge (strand6A) of the major β sheet of an adjacent copy of MENT [74,124]. This causes MENT to condense chromatin into a higher order structure by forming a series of protein bridges connecting each MENT protein to an adjacent MENT, pulling the nucleosomes together into arrays [132].

The nucleic binding features of MENT are moderately conserved in hSerpinB12. In the structure of hSerpinB12 (Figure 3EF; also Supplementary Data 3) modeled on MENT, the M-loop sequence (SKEPDPCLKSNKQKAG) contains four positively charged Lys residues in the region corresponding to the Arg-rich M-loop in MENT (Figure 5). hSerpinB12 bears a second area of concentrated positive charge above the positively charged N-terminus, where the C-terminus of helix D meets the loops of sheet B (Figure 3F). This forms an 11-residue (K102, R105, K107, K194, K196, R231, R246, K249, K251, H390, and K392) positive surface patch adjacent to the M-loop, which together form a cleft that could bind nucleic acids. Seven of the nine residues in strand 6A (Figure 3E) are conserved between hSerpinB12 and MENT, making β sheet oligomerization a possibility for hSerpinB12. However, the binding of hSerpinB12 to RNA pol II [113] suggests an alternative role in RNA rather than DNA binding. Another binding partner of hSerpinB12 is the superfamily 2 DNA helicase CHD4, which contributes to the histone deacetylase activity of NuRD [133]. As presented above for RITS-CLRC complexes, hSerpinB12 could couple RNA pol II interaction with transcriptional regulation via its histone-modifying partner, CHD4.

3.7. Are there proteins of known function with the same domain architecture as hSLFN12?

Aside from schlafens, two other protein families contain both RNaseE-like (superfamily cl26557) and GTPase-like (superfamily cl21455) domains, according to the Conserved Domains Database [12]. The first, ADP-ribosylation factor (ARF)-like proteins, are involved in actin remodeling and vesicular trafficking [134]. Some members of this family have

enzymatically inactive GTPase domains, as predicted for hSLFN12. The second family includes the DEAD-box-like helicases involved in unwinding double-stranded nucleic acids and metabolizing RNA [135]. These helicases, like the group III schlafens containing helicase superfamily I motifs [4], also contain a true P-loop NTPase AAA domain. The NTPase AAA domain is N-terminal in DEAD-like helicases, whereas it is C-terminal in group III schlafens. One of these DEAD-like helicases, DDX3, co-precipitates with Ago2 and is hypothesized to modulate miRNA demethylation [136]. Additionally, a non-covalent association between RNaseE and an ATP-dependent helicase has been identified in the RNA degradosome [81]. Thus, proteins with similar domain architectures to Slfns are found to have RNA binding and metabolic functions similar to those suggested for hSLFN12.

Conclusions

In this work, we have modeled the domain architecture of hSLFN12, a protein that stimulates intestinal stem cell differentiation, and found that it contains an N-terminal domain related to and likely sharing the RNA binding function of rSlfn13, and a C-terminal domain that remotely resembles dynamin and may interact with viruses or actin, but likely does not hydrolyze GTP. We clarified by sequence conservation and structural analysis that the N-terminal region in Slfns is not a divergent or schlafen AAA domain; instead this RNase region in Slfn sequences and the rSlfn13 crystal structure closely matches the nucleic binding Alba_2 domain consensus sequence and 3D structure. An analysis of Alba_2, GTPase-like and AAA domains in 14 different Slfns clearly defines a hierarchical, three-domain architecture for mammalian Slfns. Group I Slfns only contain an N-terminal Alba_2 domain, while group II Slfns also include a C-terminal GTPase-like domain, and group III Slfns additionally contain a C-terminal AAA_22/helicase domain.

Analysis of conserved residues between hSLFN12 and other Slfns supports an RNA binding and possible 3' endoribonuclease function for the N-terminal domain. The C-terminal GTPase-like domain in hSLFN12, by analogy with dynamin and MxB, may contribute to chromatin and actin cytoskeleton interactions, both of which contribute to cell differentiation. In the nucleus, hSLFN12 is hypothesized to bind RNA in a RITS-like process contributing to silencing non-coding DNA or proliferation-associated gene programs via heterochromatin formation, while possibly also contributing to pre-mRNA splicing. hSLFN12 binding partners SUV39H1, JMJD6, and PDLIM7 are homologs of proteins in the yeast RITS-CLRC complex. hSLFN12 partner hSerp1B12 was found to be homologous to the heterochromatin-promoting serpin, MENT, and hSerp1B12 binds RNA pol II and the CHD4 DNA helicase of a histone deacetylase. Both these roles are relevant to the RITS-CLRC mechanism of transcriptional regulation shown in Figure 6.

Together, the results of this work help elucidate the architecture of shared domains in the ~16 000 proteins annotated as schlafens and provide specific hypotheses for the roles of the two domains of hSLFN12 and its partners in cell differentiation. Structural integrity versus functional roles for conserved sequence motifs in Slfns are also clarified, based on their structural context in rSlfn13 and the homologous hSLFN12 model. In combination with experimental data on binding partners, now readily accessible in BioGrid, the domain architecture described here is expected to provide useful insights into how hSLFN12 and

other members of the schlafen family function in stem cell differentiation and defense against viruses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Mary Walsh for stimulating our interest in decoding the structure and function of schlafens and Dr. Marc Basson for our early collaboration on rSlfn3 and hSLFN12. We appreciate feedback on the manuscript from Dr. Bill Henry (MSU), Dr. David Arnosti (MSU), and Dr. Sebastian Raschka (Univ. of Wisconsin, Madison). This manuscript is dedicated to Dr. John Schotland (undergraduate research supervisor), Dr. Jack Leigh (Ph.D. research advisor), and Dr. Les Dutton (graduate guidance committee chair), for their strong support of LK's early computational biology research at the University of Pennsylvania.

Funding: This work was partially supported by National Institutes of Health grant R01DK096137.

References

- [1]. Bustos O, Naik S, Ayers G, Casola C, Perez-Lamigueiro MA, Chippindale PT, Pritham EJ, de Casa-Esperón E, Evolution of the schlafen genes, a gene family associated with embryonic lethality, meiotic drive, immune processes and orthopoxvirus virulence, *Gene*. 447 (2009) 1–11. doi:10.1016/j.gene.2009.07.006. [PubMed: 19619625]
- [2]. Schwarz DA, Katayama CD, Hedrick SM, Schlafen, a new family of growth regulatory genes that affect thymocyte development, *9* (1998) 657–668.
- [3]. Liu F, Zhou P, Wang Q, Zhang M, Li D, The schlafen family: complex roles in different cell types and virus replication, *Cell Biol Int*. 41 (2018) 2–8. doi:10.1002/cbin.10778.
- [4]. Geserick P, Kaiser F, Klemm U, Kaufmann SHE, Zerrahn J, Modulation of T cell development and activation by novel members of the schlafen (slfn) gene family harbouring an RNA helicase-like motif, *Int. Immunol* 16 (2004) 1535–1548. doi:10.1093/intimm/dxh155. [PubMed: 15351786]
- [5]. Seong R, Seo S, Kim J, Fletcher SJ, Morgan NV, Schlafen 14 (SLFN14) is a novel antiviral factor involved in the control of viral replication, *Immunobiology*. 222 (2017) 979–988. doi:10.1016/j.imbio.2017.07.002. [PubMed: 28734654]
- [6]. Li M, Kao E, Gao X, Sandig H, Limmer K, Pavon-Eternod M, Jones TE, Landry S, Pan T, Weitzman MD, David M, Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11, *Nature*. 490 (2012) 125–128. doi:10.1038/nature11433.
- [7]. Sassano A, Mavrommatis E, Arslan D, Kroczyńska B, Beauchamp EM, Human schlafen 5 (SLFN5) is a regulator of motility and invasiveness of renal cell carcinoma cells, *35* (2015) 2684–2698. doi:10.1128/MCB.00019-15.
- [8]. Zoppoli G, Regairaz M, Leo E, Reinhold WC, Varma S, Ballestrero A, Doroshov JH, Pommier Y, Putative DNA/RNA helicase schlafen-11 (SLFN11) sensitizes cancer cells to DNA-damaging agents, *Proc Natl Acad Sci U S A*. 109 (2012) 15030–15035. doi:10.1073/pnas.1205943109. [PubMed: 22927417]
- [9]. Chène P, ATPases as drug targets: Learning from their structure, *Nat. Rev. Drug Discov* 1 (2002) 665–673. doi:10.1038/nrd894. [PubMed: 12209147]
- [10]. Walker JE, Saraste M, Runswick MJ, Gay NJ, Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold, *EMBO J*. 1 (1982) 945–951. [PubMed: 6329717]
- [11]. Neumann B, Zhao L, Murphy K, Gonda TJ, Subcellular localization of the schlafen protein family, *Biochem. Biophys. Res. Commun* 370 (2008) 62–66. doi:10.1016/j.bbrc.2008.03.032. [PubMed: 18355440]
- [12]. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z,

- Yamashita RA, Zhang D, Zheng C, Bryant SH, CDD: NCBI's Conserved Domain Database, *Nucleic Acids Res.* 43 (2015) D222–D226. doi:10.1093/nar/gku1221. [PubMed: 25414356]
- [13]. Arslan AD, Sassano A, Saleiro D, Lisowski P, Kosciuzuk EM, Fischietti M, Eckerdt F, Fish EN, Platanius LC, Human SLFN5 is a transcriptional co-repressor of STAT1-mediated interferon responses and promotes the malignant phenotype in glioblastoma, *Oncogene.* 36 (2017) 6006–6019. doi:10.1038/onc.2017.205. [PubMed: 28671669]
- [14]. Mu Y, Lou J, Srivastava M, Zhao B, Feng X, Liu T, Chen J, Huang J, SLFN11 inhibits checkpoint maintenance and homologous recombination repair, *EMBO Rep.* 17 (2016) 94–109. doi:10.15252/embr.201540964. [PubMed: 26658330]
- [15]. Li M, Kao E, Malone D, Gao X, Wang JYJ, David M, DNA damage-induced cell death relies on SLFN11-dependent cleavage of distinct type II tRNAs, *Nat. Struct. Mol. Biol.* 25 (2018). doi: 10.1038/s41594-018-0142-5.
- [16]. Yang J-Y, Deng XY, Li YS, Ma XC, Feng JX, Yu B, Chen Y, Luo YL, Wang X, Chen ML, Fang ZX, Zheng FX, Li YP, Zhong Q, Kang TB, Song LB, Xu RH, Zeng MS, Chen W, Zhang H, Xie W, Gao S, Structure of schlafen13 reveals a new class of tRNA/rRNA-targeting RNase engaged in translational control, *Nat. Commun* 9 (2018). doi:10.1038/s41467-018-03544-x.
- [17]. Pisareva VP, Muslimov IA, Tcherepanov A, Pisarev AV, Characterization of novel ribosome-associated endoribonuclease SLFN14 from rabbit reticulocytes, *Biochemistry.* 54 (2015) 3286–3301. doi:10.1021/acs.biochem.5b00302. [PubMed: 25996083]
- [18]. Kovalenko PL, Basson MD, Schlafen 12 expression modulates prostate cancer cell differentiation, *J. Surg. Res* 190 (2014) 177–184. doi:10.1016/j.jss.2014.03.069. [PubMed: 24768141]
- [19]. Basson MD, Wang Q, Chaturvedi LS, More S, Vomhof-DeKrey EE, Al-Marsoumi S, Sun K, Kuhn LA, Kovalenko P, Kiupel M, Schlafen 12 interaction with serpinB12 and deubiquitylases drives human enterocyte differentiation, *Cell. Physiol. Biochem* 482 (2018) 1274–1290. doi: 10.1159/000492019.
- [20]. Bastian F, Parmentier G, Roux J, Moretti S, Laudet V, Robinson-Rechavi M, Bgee: Integrating and comparing heterogeneous transcriptome data among species, *Lect. Notes Comput. Sci.* (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 5109 LNBI (2008) 124–131. doi:10.1007/978-3-540-69828-9_12.
- [21]. Shaw D, Gohil K, Basson MD, Intestinal mucosal atrophy and adaptation, *World J. Gastroenterol* 18 (2012) 6357–6375. doi:10.3748/wjg.v18.i44.6357. [PubMed: 23197881]
- [22]. Lewis K, Lutgendorff F, Phan V, Söderholm J, Sherman P, McKay D, Enhanced translocation of bacteria across metabolically stressed epithelia is reduced by butyrate, *Inflamm. Bowel Dis* 16 (2009) 1138–1148. doi:10.1002/ibd.21177.
- [23]. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402. doi:10.1093/nar/25.17.3389. [PubMed: 9254694]
- [24]. Jelinska C, Conroy MJ, Craven CJ, Hounslow AM, Bullough PA, Waltho JP, Taylor GL, White MF, Ky F, Court F, Bank W, Sheffield S, Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging, *Structure.* 13 (2005) 963–971. doi:10.1016/j.str.2005.04.016. [PubMed: 16004869]
- [25]. Tanaka T, Padavattan S, Kumarevel T, Crystal structure of archaeal chromatin protein Alba2-double-stranded DNA complex from *Aeropyrum pernix* K1, *J. Biol. Chem* 287 (2012) 10394–10402. doi:10.1074/jbc.M112.343210. [PubMed: 22334696]
- [26]. Gubser C, Goodbody R, Ecker A, Brady G, O'Neill LAJ, Jacobs N, Smith GL, Goodbody R, Gubser C, O'Neill LAJ, Jacobs N, Smith GL, Camelpox virus encodes a schlafen-like protein that affects orthopoxvirus virulence, *J. Gen. Virol* 88 (2007) 1667–1676. doi:10.1099/vir.0.82748-0. [PubMed: 17485525]
- [27]. Lin Y, Sun L, Zhu D, Hu Z, Wang X, Du C, Wang Y, Wang X, Zhou J, Equine schlafen 11 restricts the production of equine infectious anemia virus via a codon usage-dependent mechanism, *Virology.* 495 (2016) 112–121. doi:10.1016/j.virol.2016.04.024. [PubMed: 27200480]

- [28]. Brady G, Boggan L, Bowie A, O'Neill LAJ, Schlafen-1 causes a cell cycle arrest by inhibiting induction of cyclin D1, *J. Biol. Chem* 280 (2005) 30723–30734. doi:10.1074/jbc.M500435200. [PubMed: 15946944]
- [29]. Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, Cen S, Guo F, Liang C, The interferon-inducible MxB protein inhibits HIV-1 infection, *Cell Host Microbe*. 14 (2013) 398–410. doi:10.1016/j.chom.2013.08.015. [PubMed: 24055605]
- [30]. Patel VB, Majumdar APN, Sanders MA, Yu Y, Patel BB, Nautiyal J, Kanwar SS, Oh P-S, Schlafen-3 decreases cancer stem cell marker expression and autocrine/juxtacrine signaling in FOLFOX-resistant colon cancer cells, *Am. J. Physiol. Liver Physiol* 301 (2011) G347–G355. doi: 10.1152/ajpgi.00403.2010.
- [31]. Patel VB, Yu Y, Das JK, Patel BB, Majumdar APN, Schlafen-3: A novel regulator of intestinal differentiation, *Biochem. Biophys. Res. Commun* 388 (2009) 752–756. doi:10.1016/j.bbrc.2009.08.094. [PubMed: 19703412]
- [32]. Sohn W-J, Kim D, Lee K-W, Kim M-S, Kwon S, Lee Y, Kim D-S, Kwon H-J, Novel transcriptional regulation of the schlafen-2 gene in macrophages in response to TLR-triggered stimulation, *Mol. Immunol* 44 (2007) 3273–3282. [PubMed: 17434208]
- [33]. van Zuylén WJ, Garceau V, Idris A, Schroder K, Irvine KM, Lattin JE, Ovchinnikov DA, Perkins AC, Cook AD, Hamilton JA, Hertzog PJ, Stacey KJ, Kellie S, Hume DA, Sweet MJ, Macrophage activation and differentiation signals regulate schlafen-4 gene expression: evidence for schlafen-4 as a modulator of myelopoiesis, *PLoS One*. 6 (2011) e15723. doi:10.1371/journal.pone.0015723. [PubMed: 21249125]
- [34]. Mavrommatis E, Arslan AD, Sassano A, Hua Y, Kroczyńska B, Plataniás LC, Expression and regulatory effects of murine schlafen (Slfn) genes in malignant melanoma and renal cell carcinoma, *J Biol Chem*. 288 (2013) 33006–33015. doi:10.1074/jbc.M113.460741. [PubMed: 24089532]
- [35]. De La Casa-Esperon E, From mammals to viruses: The Schlafen genes in developmental, proliferative and immune processes, *Biomol. Concepts* 2 (2011) 159–169. doi:10.1515/bmc.2011.018. [PubMed: 25962026]
- [36]. Chaturvedi L, Sun K, Walsh MF, Kuhn LA, Basson MD, The P-loop region of Schlafen 3 acts within the cytosol to induce differentiation of human Caco-2 intestinal epithelial cells, *Biochim. Biophys. Acta - Mol. Cell Res* 1843 (2014). doi:10.1016/j.bbamcr.2014.09.017.
- [37]. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N, ConSeq: The identification of functionally and structurally important residues in protein sequences, *Bioinformatics*. 20 (2004) 1322–1324. doi:10.1093/bioinformatics/bth070. [PubMed: 14871869]
- [38]. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N, ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, *Nucleic Acids Res*. 38 (2010) 529–533. doi:10.1093/nar/gkq399.
- [39]. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, *Nucleic Acids Res*. 44 (2016) 1–7. doi:10.1093/nar/gkw408. [PubMed: 26621913]
- [40]. Goldberg T, Hamp T, Rost B, LocTree2 predicts localization for all domains of life, *Bioinformatics*. 28 (2012) 458–465. doi:10.1093/bioinformatics/bts390.
- [41]. Rost B, Liu J, The PredictProtein server, *Nucleic Acids Res*. 31 (2003) 3300–3304. doi: 10.1093/nar/gkg508. [PubMed: 12824312]
- [42]. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L, Ashkenazy H, Punta M, Schlessinger A, Bromberg Y, Schneider R, Vriend G, Sander C, Ben-Tal N, Rost B, PredictProtein - an open resource for online prediction of protein structural and functional features, *Nucleic Acids Res*. 42 (2014) 337–343. doi:10.1093/nar/gku366.
- [43]. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V, A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core, *J. Mol. Biol* 430 (2017) 2237–2243. doi:10.1016/j.jmb.2017.12.007. [PubMed: 29258817]

- [44]. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410. doi:10.1016/S0022-2836(05)80360-2. [PubMed: 2231712]
- [45]. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL, Domain enhanced lookup time accelerated BLAST, *Biol. Direct* 7 (2012) 12. doi:10.1186/1745-6150-7-12. [PubMed: 22510480]
- [46]. Holm L, Laakso LM, Dali server update, *Nucleic Acids Res.* 44 (2016) W351–W355. doi:10.1093/nar/gkw357. [PubMed: 27131377]
- [47]. Laskowski RA, Chistyakov VV, Thornton JM, PDBsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids, *Nucleic Acids Res.* 33 (2005) 266–268. doi:10.1093/nar/gki001.
- [48]. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M, Pfam: The protein families database, *Nucleic Acids Res.* 42 (2014) 281–288. doi:10.1109/TCSVT.2017.2671899.
- [49]. Stark C, BioGRID: a general repository for interaction datasets, *Nucleic Acids Res.* 34 (2006) D535–539. doi:10.1093/nar/gkj109. [PubMed: 16381927]
- [50]. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, Bursteinas B, Bye-Ajee H, Cowley A, Da Silva A, De Giorgi M, Dogan T, Fazzini F, Castro LG, Figueira L, Garmiri P, Georghiou G, Gonzalez D, Hatton-Ellis E, Li W, Liu W, Lopez R, Luo J, Lussi Y, MacDougall A, Nightingale A, Palka B, Pichler K, Poggioli D, Pundir S, Pureza L, Qi G, Rosanoff S, Saidi R, Sawford T, Shypitsyna A, Speretta E, Turner E, Tyagi N, Volynkin V, Wardell T, Warner K, Watkins X, Zaru R, Zellner H, Xenarios I, Bougueleret L, Bridge A, Poux S, Redaschi N, Aimo L, ArgoudPuy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Boeckmann B, Bolleman J, Boutet E, Breuza L, Casal-Casas C, De Castro E, Coudert E, Cucho B, Doche M, Dornevil D, Duvaud S, Estreicher A, Famiglietti L, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Jungo F, Keller G, Lara V, Lemercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto T, Nospikel N, Paesano S, Pedruzzi I, Pilbout S, Pozzato M, Pruess M, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey AL, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Zhang J, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.* 45 (2017) D158–D169. doi:10.1093/nar/gkw1099. [PubMed: 27899622]
- [51]. Törönen P, Medlar A, Holm L, PANNZER2: a rapid functional annotation web server, *Nucleic Acids Res.* 46 (2018) W84–W88. doi:10.1093/nar/gky350. [PubMed: 29741643]
- [52]. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Philip E, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Thanki N, Weissig H, Westbrook JD, The Protein Data Bank: unifying the archive, *Nucleic Acids Res.* 30 (2002) 245–248. [PubMed: 11752306]
- [53]. Rose PW, Prli A, Altunkaya A, Bi C, Bradley AR, Christie CH, Di Costanzo L, Duarte JM, Dutta S, Feng Z, Green RK, Goodsell DS, Hudson B, Kalro T, Lowe R, Peisach E, Randle C, Rose AS, Shao C, Tao YP, Valasatava Y, Voigt M, Westbrook JD, Woo J, Yang H, Young JY, Zardecki C, Berman HM, Burley SK, The RCSB Protein Data Bank: integrative view of protein, gene and 3D structural information, *Nucleic Acids Res.* 45 (2017) D271–D281. doi:10.1093/nar/gkw1000. [PubMed: 27794042]
- [54]. Wu S, Zhang Y, MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information, *Proteins Struct. Funct. Genet* 72 (2008) 547–556. doi:10.1002/prot.21945. [PubMed: 18247410]
- [55]. Wallner B, Elofsson A, Pcons5: combining consensus, structural evaluation and fold recognition scores, *Bioinformatics.* 21 (2005) 4248–4254. doi:10.1093/bioinformatics/bti702. [PubMed: 16204344]
- [56]. Wallner B, Larsson P, Elofsson A, Pcons.net: protein structure prediction meta server, *Nucleic Acids Res.* 35 (2007) 369–374. doi:10.1093/nar/gkm319.
- [57]. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T, Protein structure homology modeling using SWISS-MODEL workspace, *Nat. Protoc* 4 (2009) 1–13. doi:10.1038/nprot.2008.197. [PubMed: 19131951]

- [58]. Xu D, Jaroszewski L, Li Z, Godzik A, FFAS-3D: Improving fold recognition by including optimized structural features and template re-ranking, *Bioinformatics*. 30 (2014) 660–667. doi: 10.1093/bioinformatics/btt578. [PubMed: 24130308]
- [59]. Laskowski RA, MacArthur MW, Moss DS, Thornton JM, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291. doi: 10.1107/S0021889892009944.
- [60]. Benkert P, Tosatto SCE, Schomburg D, QMEAN: a comprehensive scoring function for model quality assessment, *Proteins-Structure Funct. Bioinforma* 71 (2008) 261–277. doi:10.1002/prot.21715.
- [61]. Iyer LM, Leipe DD, Koonin EV, Aravind L, Evolutionary history and higher order classification of AAA+ ATPases, *J. Struct. Biol* 146 (2004) 11–31. doi:10.1016/j.jsb.2003.10.010. [PubMed: 15037234]
- [62]. Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, Chothia C, Gough J, SUPERFAMILY - Sophisticated comparative genomics, data mining, visualization and phylogeny, *Nucleic Acids Res.* 37 (2009) 380–386. doi:10.1093/nar/gkn762.
- [63]. Raney KD, Byrd AK, Aarattuthodiyil S, Structure and mechanisms of SF1 DNA helicases, *Adv Exp Med Biol.* 767 (2013) 17–46. doi:10.1007/978-1-4614-5037-5. [PubMed: 23161005]
- [64]. Sander C, Schneider R, Database of homology-derived protein structures and the structural meaning of sequence alignment, *Proteins Struct. Funct. Bioinforma* 9 (1991) 56–68.
- [65]. Niemann HH, Knetsch MLW, Scherer A, Manstein DJ, Kull FJ, Crystal structure of a dynamin GTPase domain in both nucleotide-free and GDP-bound forms, *EMBO J.* 20 (2001) 5813–5821. doi:10.1093/emboj/20.21.5813. [PubMed: 11689422]
- [66]. Bustelo XR, Sauzeau V, Berenjano IM, GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo, *Bioessays.* 29 (2007) 356–370. doi:10.1021/nl061786n.Core-Shell. [PubMed: 17373658]
- [67]. Vale RD, AAA proteins: lords of the ring, *J. Cell Biol* 150 (2000) 13–19. doi:10.1083/jcb.150.1.F13. [PubMed: 10893252]
- [68]. Koonin EV, The origin and early evolution of eukaryotes in the light of phylogenomics, *Genome Biol.* 11 (2010) 209. doi:10.1186/gb-2010-11-5-209. [PubMed: 20441612]
- [69]. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M, KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.* 44 (2016) D457–D462. doi:10.1093/nar/gkv1070. [PubMed: 26476454]
- [70]. Kanehisa M, Sato Y, Morishima K, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences, *J. Mol. Biol* 428 (2016) 726–731. doi: 10.1016/j.jmb.2015.11.006. [PubMed: 26585406]
- [71]. Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, Shu S, Neupane R, Cipriano M, Mancuso J, Tu H, Salamov A, Lindquist E, Shapiro H, Lucas S, Grigoriev IV, Cande WZ, Fulton C, Rokhsar DS, Dawson SC, The genome of *Naegleria gruberi* illuminates early eukaryotic versatility, *Cell.* 140 (2010) 631–642. doi:10.1016/j.cell.2010.01.032. [PubMed: 20211133]
- [72]. Anand R, Eschenburg S, Reubold TF, Crystal structure of the GTPase domain and the bundle signalling element of dynamin in the GDP state, *Biochem. Biophys. Res. Commun* 469 (2016) 76–80. doi:10.1016/j.bbrc.2015.11.074. [PubMed: 26612256]
- [73]. Askew YS, Pak SC, Luke CJ, Askew DJ, Cataltepe S, Mills DR, Kato H, Lehoczky J, Dewar K, Birren B, Silverman GA, SERPINB12 is a novel member of the human ov-serpin family that is widely expressed and inhibits trypsin-like serine proteinases, *J Biol Chem.* 276 (2001) 49320–49330. doi:10.1074/jbc.M108879200. [PubMed: 11604408]
- [74]. McGowan S, Buckle AM, Irving JA, Ong PC, Kan W, Henderson KN, Bulynko YA, Popova EY, Smith AI, Bottomley SP, Rossjohn J, Grigoryev SA, Pike RN, Whisstock JC, X-ray crystal structure of MENT: evidence for functional loop – sheet polymers in chromatin condensation, *EMBO J.* 25 (2006) 3144–3155. doi:10.1038/sj.emboj.7601201. [PubMed: 16810322]
- [75]. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, 42 (2014) 252–258. doi:10.1093/nar/gku340.

- [76]. Terribilini M, Lee J-H, Yan C, Jernigan R, Honavar V, Dobbs D, Prediction of RNA-binding sites in proteins from amino acid sequence: a machine learning approach, *RNA*. 12 (2005) 1450–1462. doi:10.1261/rna.2197306.
- [77]. Marletaz F, Firbas P, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CD, de la Calle-Mustienes E, Bertrand S, Burguera D, van Heeringen SJ, Jimenez-Gancedo S, Aldea D, Marquez Y, Buono L, Kozmikova I, Louis A, Permanyer J, Acemel RD, Albuixech-Crespo B, Le Petillon Y, Florian AL, Subirana L, Farahani E, Albalat R, Aury JM, Benito-Gutierrez E, Canestro C, Castro F, D'Aniello S, Ferrier D, Huang S, Laudet V, Marais G, Pontarotti P, Robinson-Rechavi M, Schubert M, Seitz H, Somorjai I, Takahashi T, Tostivint H, Xu A, Yu J-K, Martinez-Morales JR, Crolius HR, Kozmik Z, Weirauch M, Garcia-Fernandez J, Lister R, Lenhard B, Holland P, Escriva H, Gomez-Skarmeta JL, Irimia M, Amphioxus functional genomics and the evolution of vertebrate regulatory traits, *Revis*. (2018). doi:10.1038/s41586-018-0734-6.
- [78]. Goyal M, Banerjee C, Nag S, Bandyopadhyay U, The Alba protein family: structure and function, *Biochim. Biophys. Acta* 1864 (2016) 570–583. doi:10.1016/j.bbapap.2016.02.015. [PubMed: 26900088]
- [79]. Aravind L, Iyer LM, Anantharaman V, The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism, *Genome Biol.* 4 (2003) R64. doi:10.1186/gb-2003-4-10-r64. [PubMed: 14519199]
- [80]. Bandyra KJ, Wandzik JM, Luisi BF, Substrate recognition and autoinhibition in the central ribonuclease RNase E, *Mol. Cell* 72 (2018) 275–285.e4. doi:10.1016/j.molcel.2018.08.039. [PubMed: 30270108]
- [81]. Callaghan AJ, Aurikko JP, Ilag LL, Günter Grossmann J, Chandran V, Kühnel K, Poljak L, Carpousis AJ, Robinson CV, Symmons MF, Luisi BF, Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E, *J. Mol. Biol.* 340 (2004) 965–979. doi:10.1016/j.jmb.2004.05.046. [PubMed: 15236960]
- [82]. Feuerstein R, Wang X, Song D, Cooke NE, Liebhaber SA, The LIM/double zinc-finger motif functions as a protein dimerization domain, *Proc. Natl. Acad. Sci* 91 (1994) 10655–10659. doi:10.1073/pnas.91.22.10655. [PubMed: 7938009]
- [83]. Krishna SS, Majumdar I, Grishin NV, Structural classification of zinc fingers, *Nucleic Acids Res.* 31 (2003) 532–550. doi:10.1093/nar/gkg161. [PubMed: 12527760]
- [84]. Laity JH, Lee BM, Wright PE, Zinc finger proteins: new insights into structural and functional diversity, *Curr. Opin. Struct. Biol.* 11 (2001) 39–46. doi:10.1016/S0959-440X(00)00167-6. [PubMed: 11179890]
- [85]. Weimann M, Grossmann A, Woodsmith J, Özkan Z, Birth P, Meierhofer D, Benlasfer N, Valovka T, Timmermann B, Wanker EE, Sauer S, Stelzl U, A Y2H-seq approach defines the human protein methyltransferase interactome, *Nat. Methods* 10 (2013) 339–342. doi:10.1038/nmeth.2397. [PubMed: 23455924]
- [86]. Irving JA, Pike RN, Lesk AM, Whisstock JC, Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function, *Genome Res.* 10 (2000) 1845–1864. doi:10.1101/gr.1478R. [PubMed: 11116082]
- [87]. Carbon S, Dietze H, Lewis SE, Mungall CJ, Munoz-Torres MC, Basu S, Chisholm RL, Dodson RJ, Fey P, Thomas PD, Mi H, Muruganujan A, Huang X, Poudel S, Hu JC, Aleksander SA, McIntosh BK, Renfro DP, Siegele DA, Antonazzo G, Attrill H, Brown NH, Marygold SJ, McQuilton P, Ponting L, Millburn GH, Rey AJ, Stefancsik R, Tweedie S, Falls K, Schroeder AJ, Courtot M, Osumi-Sutherland D, Parkinson H, Roncaglia P, Lovering RC, Foulger RE, Huntley RP, Denny P, Campbell NH, Kramarz B, Patel S, Buxton JL, Umrao Z, Deng AT, Alrohaif H, Mitchell K, Ratnaraj F, Omer W, Rodríguez-López M, Chibucos MC, Giglio M, Nadendla S, Duesbury MJ, Koch M, Meldal BHM, Melidoni A, Porras P, Orchard S, Shrivastava A, Chang HY, Finn RD, Fraser M, Mitchell AL, Nuka G, Potter S, Rawlings ND, Richardson L, Sangrador-Vegas A, Young SY, Blake JA, Christie KR, Dolan ME, Drabkin HJ, Hill DP, Ni L, Sitnikov D, Harris MA, Hayles J, Oliver SG, Rutherford K, Wood V, Bahler J, Lock A, De Pons J, Dwinell M, Shimoyama M, Laulederkind S, Hayman GT, Tutaj M, Wang SJ, D'Eustachio P, Matthews L, Balhoff JP, Balakrishnan R, Binkley G, Cherry JM, Costanzo MC, Engel SR, Miyasato SR, Nash RS, Simison M, Skrzypek MS, Weng S, Wong ED, Feuermann M, Gaudet P, Berardini TZ, Li D, Muller B, Reiser L, Huala E, Argasinska J, Arighi C, Auchincloss A, Axelsen K, Argoud-Puy G,

Bateman A, Bely B, Blatter MC, Bonilla C, Bougueleret L, Boutet E, Breuza L, Bridge A, Britto R, Hye-A-Bye H, Casals C, Cibrian-Uhalte E, Coudert E, Cusin I, Duek-Roggli P, Estreicher A, Famiglietti L, Gane P, Garmiri P, Georghiou G, Gos A, Gruaz-Gumowski N, Hatton-Ellis E, Hinz U, Holmes A, Hulo C, Jungo F, Keller G, Laiho K, Lemercier P, Lieberherr D, Mac-Dougall A, Magrane M, Martin MJ, Masson P, Natale DA, O'Donovan C, Pedruzzi I, Pichler K, Poggioli D, Poux S, Rivoire C, Roechert B, Sawford T, Schneider M, Speretta E, Shypitsyna A, Stutz A, Sundaram S, Tognolli M, Wu C, Xenarios I, Yeh LS, Chan J, Gao S, Howe K, Kishore R, Lee R, Li Y, Lomax J, Muller HM, Raciti D, Van Auken K, Berriman M, Stein L, Paul Kersey PW, Sternberg D. Howe, Westerfield M, Expansion of the Gene Ontology Knowledgebase and resources: the Gene Ontology Consortium, *Nucleic Acids Res.* 45 (2017) D331–D338. doi: 10.1093/nar/gkw1108. [PubMed: 27899567]

- [88]. The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver LI, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G, Gene Ontology: tool for the unification of biology, *Nat. Genet* 25 (2000) 25–29. doi:10.1038/75556. Gene. [PubMed: 10802651]
- [89]. Kudithipudi S, Schuhmacher MK, Kebede AF, Jeltsch A, The SUV39H1 protein lysine methyltransferase methylates chromatin proteins involved in heterochromatin formation and VDJ recombination, *ACS Chem. Biol* 12 (2017) 958–968. doi:10.1021/acscchembio.6b01076. [PubMed: 28169523]
- [90]. Ait-Si-Ali S, Guasconi V, Fritsch L, Yahia H, Sekhri R, Naguibneva I, Robin P, Cabon F, Polesskaya A, Harel-Bellan A, A Suv39h-dependent mechanism for silencing S-phase genes in differentiating but not in cycling cells, *EMBO J.* 23 (2004) 605–615. doi:10.1038/sj.emboj.7600074. [PubMed: 14765126]
- [91]. Johnson WL, Yewdell WT, Bell JC, McNulty SM, Duda Z, O'Neill RJ, Sullivan BA, Straight AF, RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin, *Elife.* 6 (2017) 1–32. doi:10.7554/eLife.25299.
- [92]. Unoki M, Masuda A, Dohmae N, Arita K, Yoshimatsu M, Iwai Y, Fukui Y, Ueda K, Hamamoto R, Shirakawa M, Sasaki H, Nakamura Y, Lysyl 5-hydroxylation, a novel histone modification, by jumonji domain containing 6 (JMJD6), *J. Biol. Chem* 288 (2013) 6053–6062. doi:10.1074/jbc.M112.433284. [PubMed: 23303181]
- [93]. Chang B, Chen Y, Zhao Y, Bruick RK, JMJD6 Is a histone arginine demethylase, *Science* (80–.). 318 (2007) 444–447. <http://science.sciencemag.org/content/318/5849/444.abstract>.
- [94]. Webby CJ, Wolf A, Gromak N, Dreger M, Kramer H, Kessler B, Nielsen ML, Schmitz C, Butler DS, Yates JR, Delahunty CM, Hahn P, Lengeling A, Mann M, Proudfoot NJ, Schofield CJ, Böttger A, Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing, *Science* (80–.). 325 (2009) 90 LP–93. <http://science.sciencemag.org/content/325/5936/90.abstract>.
- [95]. Liu W, Ma Q, Wong K, Li W, Ohgi K, Zhang J, Aggarwal AK, Rosenfeld MG, Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release, *Cell.* 155 (2013) 1581–1595. doi:10.1016/j.cell.2013.10.056. [PubMed: 24360279]
- [96]. Kwok J, O'Shea M, Hume DA, Lengeling A, Jmjd6, a JmjC dioxygenase with many interaction partners and pleiotropic functions, *Front. Genet* 8 (2017) 1–19. doi:10.3389/fgene.2017.00032. [PubMed: 28179914]
- [97]. Hong X, Zang J, White J, Wang C, Pan C, Zhao R, Murphy RC, Interaction of JMJD6 with single-stranded RNA, *Proc. Natl. Acad. Sci* 107 (2010) 14568–14572. doi:10.1073/pnas.1008832107/-/DCSupplemental. www.pnas.org/cgi/doi/10.1073/pnas.1008832107. [PubMed: 20679243]
- [98]. Hahn P, Wegener I, Burrells A, Böse J, Wolf A, Erck C, Butler D, Schofield CJ, Böttger A, Lengeling A, Analysis of Jmjd6 cellular localization and testing for its involvement in histone demethylation, *PLoS One.* 5 (2010) e13769 10.1371/journal.pone.0013769. [PubMed: 21060799]
- [99]. Heim A, Grimm C, Müller U, Häußler S, Mackeen MM, Merl J, Hauck SM, Kessler BM, Schofield CJ, Wolf A, Böttger A, Jumonji domain containing protein 6 (Jmjd6) modulates splicing and specifically interacts with arginine-serine-rich (RS) domains of SR- and SR-like

- proteins, *Nucleic Acids Res.* 42 (2014) 7833–7850. doi:10.1093/nar/gku488. [PubMed: 24914048]
- [100]. Böse J, Gruber AD, Helming L, Schiebe S, Wegener I, Hafner M, Beales M, Köntgen F, Lengeling A, The phosphatidylserine receptor has essential functions during embryogenesis but not in apoptotic cell removal, *J. Biol* 3 (2004) 15. doi:10.1186/jbiol10. [PubMed: 15345036]
- [101]. Liu X, Kraus WL, Bai X, Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways, *Trends Biochem. Sci* 40 (2015) 516–525. doi:10.1016/j.tibs.2015.07.003. [PubMed: 26254229]
- [102]. Huang V, Li LC, Demystifying the nuclear function of argonaute proteins, *RNA Biol* 11 (2014) 18–24. doi:10.4161/rna.27604. [PubMed: 24384674]
- [103]. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström Di., Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF, Schramm A, Baker BJ, Spang A, Ettema TJG, Asgard archaea illuminate the origin of eukaryotic cellular complexity, *Nature*. 541 (2017) 353–358. doi:10.1038/nature21031. [PubMed: 28077874]
- [104]. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, Van Der Oost J, The evolutionary journey of argonaute proteins, *Nat. Struct. Mol. Biol* 21 (2014) 743–753. doi:10.1038/nsmb.2879. [PubMed: 25192263]
- [105]. Willkomm S, Makarova KS, Grohmann D, DNA silencing by prokaryotic argonaute proteins adds a new layer of defense against invading nucleic acids, *FEMS Microbiol. Rev.* 42 (2018) 376–387. doi:10.1093/femsre/fuy010. [PubMed: 29579258]
- [106]. Serrano L, Martínez-Redondo P, Marazuela-Duque A, Vazquez BN, Dooley SJ, Voigt P, Beck DB, Kane-Goldsmith N, Tong Q, Rabanal RM, Fondevila D, Muñoz P, Krüger M, Tischfield JA, Vaquero A, The tumor suppressor SirT2 regulates cell cycle progression and genome stability by modulating the mitotic deposition of H4K20 methylation, *Genes Dev.* 27 (2013) 639–653. doi:10.1101/gad.211342.112. [PubMed: 23468428]
- [107]. Volpe T, Martienssen RA, RNA interference and heterochromatin assembly, *Cold Spring Harb. Perspect. Biol* 3 (2011) 1–11. doi:10.1101/cshperspect.a003731.
- [108]. Bayne EH, White SA, Kagansky A, Bijos DA, Sanchez-Pulido L, Hoe KL, Kim DU, Park HO, Ponting CP, Rappsilber J, Allshire RC, Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity, *Cell*. 140 (2010) 666–677. doi:10.1016/j.cell.2010.01.038. [PubMed: 20211136]
- [109]. Sampey GC, Guendel I, Das R, Jaworski E, Klase Z, Narayanan A, Kehn-Hall K, Kashanchi F, Transcriptional gene silencing (TGS) via the RNAi machinery in HIV-1 infections, *Biology (Basel)*. 1 (2012) 339–369. doi:10.3390/biology1020339. [PubMed: 24832229]
- [110]. Li F, Huarte M, Zaratiegui M, Vaughn MW, Shi Y, Martienssen R, Cande WZ, Lid2 Is required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin, *Cell*. 135 (2008) 272–283. doi:10.1016/j.cell.2008.08.036. [PubMed: 18957202]
- [111]. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, Szpyt J, Tam S, Zarraga G, Pontano-Vaites L, Swarup S, White AE, Schweppe DK, Rad R, Erickson BK, Obar RA, Guruharsha KG, Li K, Artavanis-Tsakonas S, Gygi SP, Wade Harper J, Architecture of the human interactome defines protein communities and disease networks, *Nature*. 545 (2017) 505–509. doi:10.1038/nature22366. [PubMed: 28514442]
- [112]. Krcmery J, Gupta R, Sadleir RW, Ahrens MJ, Misener S, Kamide C, Fitchew P, Losordo DW, Crawford SE, Simon H-G, Loss of the cytoskeletal protein Pdlim7 predisposes mice to heart defects and hemostatic dysfunction, *PLoS One*. 8 (2013) e80809. doi:10.1371/journal.pone.0080809. [PubMed: 24278323]
- [113]. Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wühr M, Chick J, Zhai B, Kolippakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW, Gygi SP, The BioPlex network: a systematic exploration of the human interactome, *Cell*. 162 (2015) 425–440. doi:10.1016/j.cell.2015.06.043. [PubMed: 26186194]

- [114]. He C, Pillai SS, Taglini F, Li F, Ruan K, Zhang J, Wu J, Shi Y, Bayne EH, Structural analysis of Stc1 provides insights into the coupling of RNAi and chromatin modification, *Proc. Natl. Acad. Sci* 110 (2013) E1879–E1888. doi:10.1073/pnas.1212155110. [PubMed: 23613586]
- [115]. Kalantari R, Chiang CM, Corey DR, Regulation of mammalian transcription and splicing by nuclear RNAi, *Nucleic Acids Res.* 44 (2016) 524–537. doi:10.1093/nar/gkv1305. [PubMed: 26612865]
- [116]. Yang Y, Liu R, Qiu R, Zheng Y, Huang W, Hu H, Ji Q, He H, Shang Y, Gong Y, Wang Y, CRL4B promotes tumorigenesis by coordinating with SUV39H1/HP1/DNMT3A in DNA methylation-based epigenetic silencing, *Oncogene.* 34 (2015) 104–118. doi:10.1038/onc.2013.522. [PubMed: 24292684]
- [117]. O’Connell BC, Harper JW, Ubiquitin proteasome system (UPS): what can chromatin do for you?, *Curr. Opin. Cell Biol* 19 (2007) 206–214. doi:10.1016/j.ceb.2007.02.014. [PubMed: 17314036]
- [118]. Jackson S, Xiong Y, CRL4s: the CUL4-RING E3 ubiquitin ligases, *Trends Biochem Sci.* 34 (2010) 562–570. doi:10.1016/j.tibs.2009.07.002.CRL4s.
- [119]. de la Casa-Esperón E, Horizontal transfer and the evolution of host-pathogen interactions, *Int. J. Evol. Biol* 2012 (2012) 1–9. doi:10.1155/2012/679045.
- [120]. Fribourgh JL, Nguyen HC, Matreyek KA, Alvarez FJD, Summers BJ, Dewdney TG, Aiken C, Zhang P, Engelman A, Xiong Y, Structural insight into HIV-1 restriction by MxB, *Cell Host Microbe.* 16 (2014) 627–638. doi:10.1016/j.chom.2014.09.021. [PubMed: 25312384]
- [121]. Verzi MP, Shin H, He HH, Sulahian R, Meyer CA, Montgomery RK, Fleet JC, Brown M, Liu XS, Shivdasani RA, Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2, *Dev. Cell* 19 (2010) 713–726. doi:10.1016/j.devcel.2010.10.006. [PubMed: 21074721]
- [122]. Melén K, Julkunen I, Nuclear cotransport mechanism of cytoplasmic human MxB protein, *J. Biol. Chem* 272 (1997) 32353–32359. doi:10.1074/jbc.272.51.32353. [PubMed: 9405443]
- [123]. Melén K, Keskinen P, Ronni T, Sareneva T, Lounatmaa K, Julkunen I, Human MxB protein, an interferon-alpha-inducible GTPase, contains a nuclear targeting signal and is localized in the heterochromatin region beneath the nuclear envelope, *J Biol Chem* 271 (1996) 23478–23486. [PubMed: 8798556]
- [124]. Springhetti EM, Istomina NE, Whisstock JC, Nikitina T, Woodcock CL, Grigoryev SA, Role of the Mloop and reactive center loop domains in the folding and bridging of nucleosome arrays by MENT, *J. Biol. Chem* 278 (2003) 43384–43393. doi:10.1074/jbc.M307635200. [PubMed: 12930828]
- [125]. Bramkamp M, Structure and function of bacterial dynamin-like proteins, *Biol. Chem* 393 (2012) 1203–1214. doi:10.1515/hsz-2012-0185. [PubMed: 23109540]
- [126]. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M, The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts, *Mol. Cell* 46 (2012) 674–690. doi:10.1016/j.molcel.2012.05.021. [PubMed: 22681889]
- [127]. Kristensen AR, Gsponer J, Foster LJ, A high-throughput approach for measuring temporal changes in the interactome, *Nat. Methods* 9 (2012) 907–909. doi:10.1038/nmeth.2131. [PubMed: 22863883]
- [128]. McLysaght A, Baldi PF, Gaut BS, Extensive gene gain associated with adaptive evolution of poxviruses, *Proc. Natl. Acad. Sci* 100 (2003) 15655–15660. doi:10.1073/pnas.2136653100. [PubMed: 14660798]
- [129]. Missen MA, Haylock D, Whitty G, Medcalf RL, Coughlin PB, Stage specific gene expression of serpins and their cognate proteases during myeloid differentiation, *Br. J. Haematol* 135 (2006) 715–724. doi:10.1111/j.1365-2141.2006.06360.x. [PubMed: 17107353]
- [130]. Boeuf S, Steck E, Peltari K, Hennig T, Buneß A, Benz K, Witte D, Sülthmann H, Poustka A, Richter W, Subtractive gene expression profiling of articular cartilage and mesenchymal stem cells: serpins as cartilage-relevant differentiation markers, *Osteoarthr. Cartil* 16 (2008) 48–60. doi:10.1016/j.joca.2007.05.008. [PubMed: 17604188]

- [131]. Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PGW, Irving JA, Lomas DA, Luke CJ, Moyer RW, Pemberton PA, Donnell ER, Salvesen GS, Travis J, Whisstock JC, The serpins are an expanding superfamily of structurally similar but functionally diverse proteins: evolution, mechanism of inhibition, novel functions, and a revised nomenclature, *J. Biol. Chem* 276 (2001) 33293–33296. doi:10.1074/jbc.R100016200. [PubMed: 11435447]
- [132]. Grigoryev SA, Higher-order folding of heterochromatin: protein bridges span the nucleosome arrays, *Biochem. Cell Biol* 79 (2001) 227–41. doi:10.1139/bcb-79-3-227. [PubMed: 11467737]
- [133]. Hoffmeister H, Fuchs A, Erdel F, Pinz S, Grobner-Ferreira R, Bruckmann A, Deutzmann R, Schwartz U, Maldonado R, Huber C, Dendorfer A-S, Rippe K, Langst G, CHD3 and CHD4 form distinct NuRD complexes with different yet overlapping functionality, *Nucleic Acids Res.* 45 (2017) 10534–10554. doi:10.1093/nar/gkx711. [PubMed: 28977666]
- [134]. Pasqualato S, Renault L, Cherfils J, Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for “front-back” communication, *EMBO Rep.* 3 (2002) 1035–1041. doi:10.1093/embo-reports/kvf221. [PubMed: 12429613]
- [135]. Tanner NK, Linder P, DExD/H box RNA helicases: from generic motors to specific dissociation functions, *Mol Cell.* 8 (2001) 251–262. [PubMed: 11545728]
- [136]. Shah A, Rashid F, Awan HM, Hu S, Wang X, Chen L, Shan G, The DEAD-box RNA helicase DDX3 interacts with m6A RNA demethylase ALKBH5, *Stem Cells Int.* 2017 (2017). doi: 10.1155/2017/8596135.
- [137]. Holoch D, Moazed D, RNA-mediated epigenetic regulation of gene expression, *Nat. Rev. Genet* 16 (2015) 71–84. doi:10.1038/nrg3863. [PubMed: 25554358]

Highlights

- Schlafen12 is predicted to bind RNA and may act as a 3' endoribonuclease
- All schlafen N-terminal regions contain an Alba_2 domain, not a AAA P-loop ATPase
- The schlafen box is part of the Alba_2 domain and may form a dimer interface
- Schlafens have a central (group III) or C-terminal (group II) rhoGTPase-like domain
- Schlafen12 and its partners may function analogously to RITS-CLRC in yeast

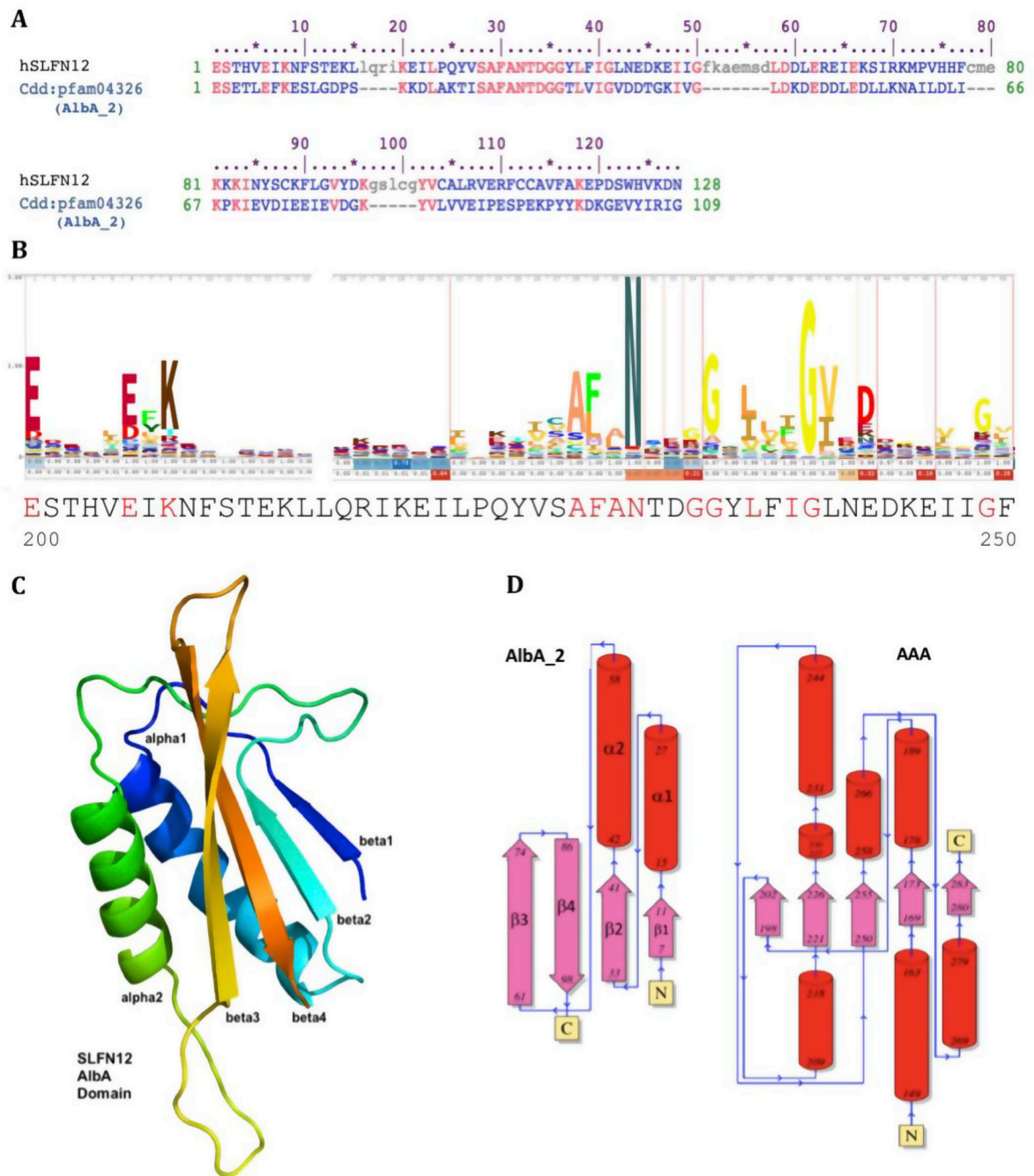


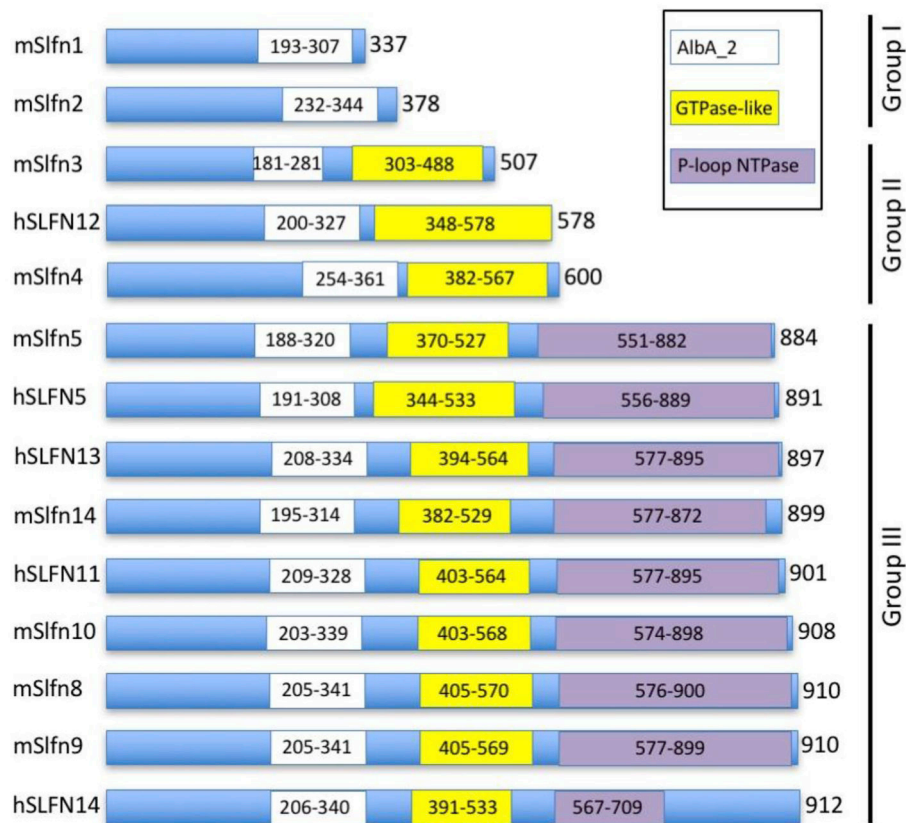
Fig. 1. Alba_2 and AAA domain analysis in hSlfn12.

(A) Sequence match between hSLFN12 residues 200–327 and the consensus sequence for Alba_2 domains (<https://pfam.xfam.org/family/PF04326>), including the Slfn box motif (ILPQYVS**AFANTDGGYLF**IGLNE) on the first line. Identical residues are shown in red, and short insertions in the hSLFN12 sequence appear in lowercase. These insertions occur at the N-termini of α helix 1 and β strand 4 and in the loop connecting β strand 2 and α helix 2 in hSLFN12, shown in the upper left corner of panel (C).

(B) Hidden Markov model weblogo representation showing sequence variation in the highly conserved region of AlbA_2 domains (<https://pfam.xfam.org/family/pf04326#tabview=tab4>). The height of each amino acid type is proportional to its dominance in that sequence position. hSLFN12 residues matching the dominant conserved residues in AlbA_2 domains are highlighted in red underneath the weblogo, showing their close match.

(C) The 3-dimensional fold of an AlbA_2 domain, corresponding to residues 203–308 in the homology modeled structure of hSLFN12, rendered from the atomic coordinates in Supplementary Data 1.

(D) Comparison of the prototypic AlbA_2 domain topology (PDB entry 3u6y) and a prototypic AAA P-loop NTPase domain (PDB entry 3n70), rendered by PDBsum. These two domains differ in α helix and β strand content and connectivity, and in anti-parallel versus parallel β sheet formation. Thus, a given region cannot form both a AAA domain and an AlbA_2 domain. Based on the sequence match shown in panels (A) and (B) and the structural match in rSlfn13 (PDB 5yd0), AlbA_2 is the correct annotation for this region in Slfns.

**Fig. 2.**

The schlafen family three-domain structure. The domains in murine and human Slfns are indicated by residue ranges corresponding to their AlbA_2, GTPase-like, and C-terminal P-loop NTPase domains. The AlbA_2 domain (Pfam motif PF04326) is found in all Slfns and encompasses the RNase active site in rSlfn13 and hSLFN14, which is conserved in hSLFN12. The GTPase-like domain following the AlbA_2 domain is present in all Group II-III Slfns, as detected by high (39%) sequence identity with residues 349–578 in hSLFN12, which are homologous to rho family GTPase D2VNU5_NAEGR (Section 3.2). The conserved SWADL motif in group II-III Slfns is central to the GTPase-like domain (Section 3.2). The group III Slfn C-terminal P-loop NTPase domain (SCOP motif SSF52540; <http://supfam.org>) [62], also annotated as a AAA_22 domain (Pfam PF13401) by CDD search in full results mode, is consistent with the ATPase motifs in superfamily I helicases [63] identified for group III Slfns [4]. UniProt identifiers of the sequences used to define these domains and residue ranges are: mSlfn1: Q9Z0I7; mSlfn2: Q9Z0I6; mSlfn3: Q9Z0I5; hSLFN12: Q8IYM2.1; mSlfn4: Q3UV66; mSlfn5: Q8CBA2.2; hSLFN5: Q08AF3.1; hSLFN13: Q68D06.1; mSlfn14: V9GXG1.1; hSLFN11: Q7Z7L1.2; mSlfn10: AAP30073.1 (<https://www.ncbi.nlm.nih.gov/genbank/>); mSlfn8: B1ARD8; mSlfn9: B1ARD6; and hSLFN14: P0C7P3.

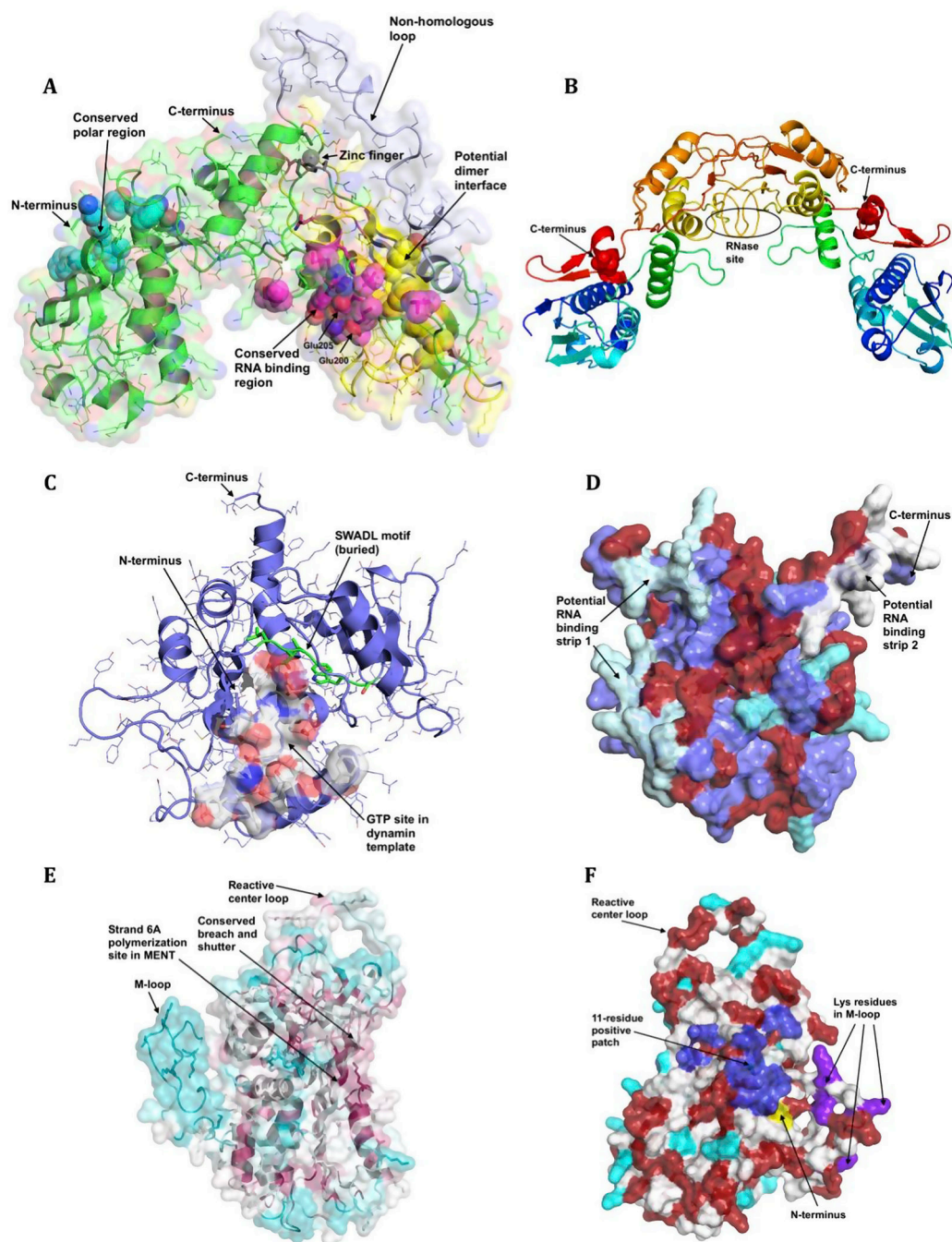
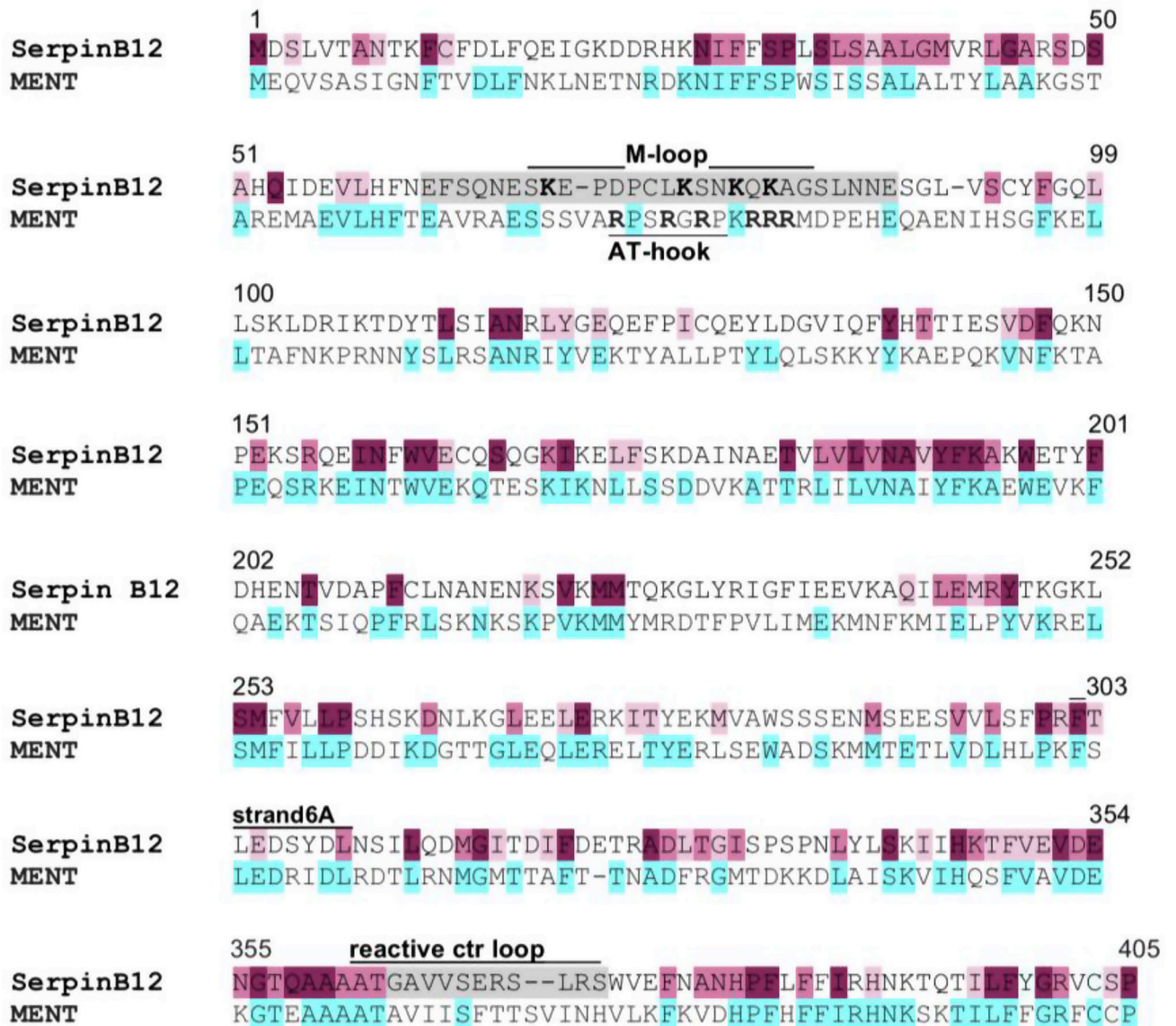


Fig. 3. Structural models of hSLFN12 and hSerpB12.

(A) The N-terminal domain of hSLFN12, with main chain in ribbons and the AlBA_2 domain in yellow. Potential functional regions are highlighted by spheres:

- Conserved RNA binding site (magenta) including 3'-endoribonuclease residues E200 and E205 (red).
- Conserved polar pocket (cyan) formed by (53)NSGGG and (106)FVKSW epitopes.
- Potential dimer or protein-protein interface (yellow).
- Conserved zinc finger ligands (dark gray).

- Non-homologous loop, residues 155–180 (pale gray).
 - C-terminus marks where the GTPase-like domain of hSLFN12 attaches.
- (B) Dimer model of hSLFN12, based on the association of AlbA domains in RNaseE. Monomers are rainbow-colored from N-terminus (blue) to C-terminus (red). The left monomer in the dimer is rotated counter-clockwise by $\sim 120^\circ$ degrees relative to the view in (A).
- (C) The C-terminal rho GTPase-like domain of hSLFN12. Residues corresponding to the dynamin GDP site are highlighted by solvent-accessible molecular surface (C in white, N in blue, and O in red). The SWADL motif (green tubes) is buried.
- (D) Favorable sites for RNA binding in the GTPase-like domain. RNA propensities are favorable for Arg, Lys, and His (> 0.50 propensity) and unfavorable for Asp, Glu, Ala, Cys, Phe, Leu, Val, and Ile (< -0.50 ; red) [76]. Favorable surfaces for RNA binding are labeled as strip 1 and strip 2. The upper region of strip 1 (pale aqua) involves R376, R378, H379, H380, R387, R397, K398, and R426 in loops between β strands. The lower region of strip 1 is formed by R526, K528, K532, K536, K539, R540, K542, and R545 in a helix-turn-helix motif. The GDP site is behind the left edge of the domain, separated from strip 1 by 15 \AA . Strip 2 (white) is formed by K565, K568, K569, K572, R575 and R576 in the C-terminal helix. Additional Arg, Lys and His residues appear in cyan. Residues neither favorable nor unfavorable for RNA binding are in blue. View is rotated by 90° about the vertical axis relative to (C).
- (E) The hSerpB12 structure modeled by homology to the crystal structure of chromatin-remodeling MENT. Residues are colored according to conservation among 150 serpins by ConSurf (blue, least conserved, to burgundy, most conserved). Conserved residues are focused in the breach and shutter regions contributing to the conformational transition of the major β sheet between the native and inhibitory (cleaved) forms in ovalbumin-family serpins.
- (F) Favorable sites for RNA binding to hSerpB12, shown on the solvent-accessible molecular surface. Residues favorable for nucleic acid binding are colored purple for four Lys residues in the M-loop, dark blue for the 11-residue positive patch above the N-terminus, and cyan for other Arg, Lys, and His residues. Residues unfavorable for nucleic acid binding appear in red, with all other residues in white. This view is rotated by 180° about the vertical axis in (E).



Residue conservation scale defined by ConSurf: 7 8 9 (most conserved)

Boldface: positively charged residues in M-loop (also known as CD loop)

Gray: loops in model likely to be flexible due to absence of defined coordinates in MENT

Aqua: residues conserved between serpinB12 and MENT

Fig. 5. Residues conserved between hSerpinB12 and human MENT (cyan), a chromatin remodeling serpin, compared with residue conservation across 150 serpin homologs (highly conserved residues in dark red), as determined by ConSurf.

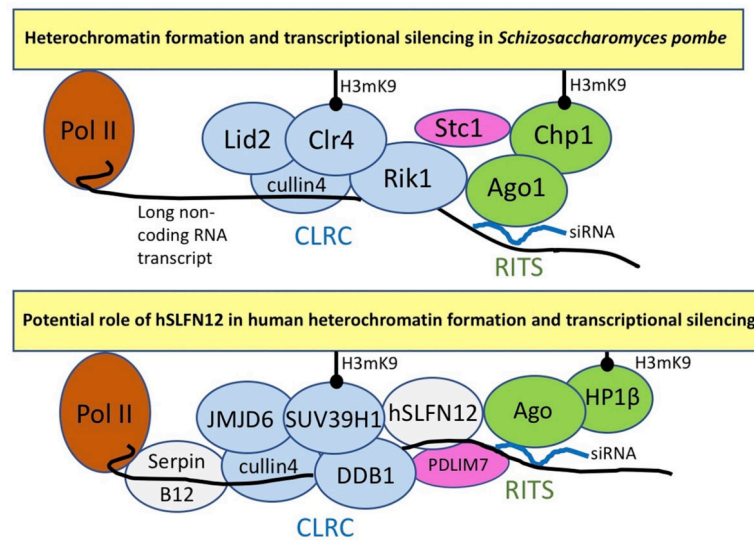


Fig. 6. Model of interactions between hSLFN12 and its binding partners, JMJD6, SUV39H1, PDLIM7, and hSerpinB12 by analogy with a characterized transcriptional regulatory complex in fission yeast. The top panel depicts the *S. pombe* RITS-CLRC complex, where CLRC is a cullin4-dependent E3 ubiquitin ligase/histone H3K9 methyltransferase complex essential for heterochromatin assembly by RNA interference (adapted from [107] and [137]). Here, heterochromatin is shown schematically as a yellow strip, with H3mK9 histone sites appearing as black lollipops. Protein correspondences between hSLFN12 partners (lower panel) and the *S. pombe* complexes (upper panel) are indicated by their similar placement and color in the schematic, and are based on the following information: human JMJD6 is a Lid2 homolog, human SUV39H1 is a Clr4 homolog, human DDB1 is a Rik1 homolog, human HP1β is a Chp1 homolog, and human PDLIM7 and yeast Stc1 are both LIM domain proteins. Based on co-immunoprecipitation, hSLFN12 is part of a complex with hSerpinB12, which also binds pol II, the CHD4 chromodomain/DNA helicase in the NuRD histone deacetylase, and E3 ligase components, according to BioGrid. hSerpinB12 is predicted to bind nucleic acids as well, based on features shared with MENT. An N-terminal Slfn-like Alba_2 domain with Slfn box is fused to an Ago protein in archaeal and bacterial proteins, indicating that Slfns and Ago work together in some contexts including chromatin maintenance in archaea. The roles of proteins in the depicted hSLFN12 complex are consistent with additional BioGrid data, including SUV39H1 binding to HP1β (CBX1), DDB1 and cullin4; HP1β binding to cullin 4; and PDLIM7 binding to the components of several E3 ligases as well as actin, which shuttles RITS components within the cell. This model provides a framework for testing hSLFN12 and hSerpinB12 roles in transcriptional regulation associated with cell differentiation, coupling two levels of gene silencing: blocking transcriptional access to genes via heterochromatin formation, and silencing nascent transcripts with siRNA. Interestingly, a top structural ortholog for hSLFN12 is a recG helicase-like protein (hATPase_c4; pfam13749; PDB entry 3lmm) from *Corynebacterium diphtheriae*, which we found to overlay well with the helicase, F-box, and

tip of the cullin domain in an E3 ligase complex with known 3D structure that is similar to CLRC, the SKP1-Cul1-Fbox E3 ligase/helicase (PDB entry 1ldk).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Summary of 3D structural models of hSLFN12 and hSerpinB12 and their stereochemical quality

Protein model	Modeling software	Modeling approach	PDB template structure	Template protein	Model QMEAN Z-score	Model-template sequence identity	Modeled residues	Procheck stereochemistry G-factor	% Residues in most-favored (ϕ, ψ) regions
N-terminal domain of hSLFN12	SwissModel	Homology	5y40 chain A	rSlfn13; ribonuclease	-2.3#	38%	12-344	0.0 (nothing unusual)	85%
C-terminal domain of hSLFN12	MUSTER	Fold recognition	1jx2 chain B	Dynammin; GTPase	-5.7##	14%	348-578	0.0 (nothing unusual)	85%
hSerpinB12	FFAS	Homology	2h4r chain A	MEN1; chromatin condensation	-3.0#	43%	1-405	0.13 (favorable)	91%

On edge of crystal structure distribution

On far edge of crystal structure distribution