

## Review

## Resources to Discover and Use Short Linear Motifs in Viral Proteins

Peter Hraber,<sup>1,\*</sup> Paul E. O'Maille,<sup>2</sup> Andrew Silberfarb,<sup>3</sup> Katie Davis-Anderson,<sup>4</sup> Nicholas Generous,<sup>5</sup> Benjamin H. McMahon,<sup>1</sup> and Jeanne M. Fair<sup>4</sup>

**Viral proteins evade host immune function by molecular mimicry, often achieved by short linear motifs (SLiMs) of three to ten consecutive amino acids (AAs). Motif mimicry tolerates mutations, evolves quickly to modify interactions with the host, and enables modular interactions with protein complexes. Host cells cannot easily coordinate changes to conserved motif recognition and binding interfaces under selective pressure to maintain critical signaling pathways. SLiMs offer potential for use in synthetic biology, such as better immunogens and therapies, but may also present biosecurity challenges. We survey viral uses of SLiMs to mimic host proteins, and information resources available for motif discovery. As the number of examples continues to grow, knowledge management tools are essential to help organize and compare new findings.**

### How Viruses Do More with Less

Viruses exploit host cellular processes to replicate, and have developed myriad ways to subvert host immune defenses. **Molecular mimicry** (see [Glossary](#)) is a common and effective strategy, enabling a pathogen to usurp host protein function by resemblance [1,2]. Molecular mimicry varies over a continuum, from one extreme that includes sequence and structural similarity (i.e., orthologs) of entire proteins, to another extreme of chemical similarity at only a few localized sites, as is the case for **short linear motifs (SLiMs)**. The growing body of literature on SLiMs indicates that some important virus–host interactions can be attributed to a few well-chosen AAs [3–7]. Rather than devote entire proteins to one function, SLiMs enable multifunctional viral proteins. Interactions between globular virus and host proteins have picomolar affinities, while SLiMs have micromolar binding affinities with globular host proteins [8]. Moderate binding affinity of SLiMs facilitates disruption of signaling interactions, rather than competing for stable formation of persistent protein complexes. Synthetic biology practitioners can benefit from an introduction to how SLiMs enable viral interference with host cell functions and computational resources available for SLiM analysis.

Viral SLiMs are potentially useful in synthetic biology, to provide a toolkit for new functions, for example, to modulate immune responses or to complement and interact with newly developed **adjuvants** in a synergistic manner [9]. Research efforts to develop broad-spectrum antiviral compounds or design broadly cross-protective vaccine immunogens benefit directly from knowledge of gene products, protein functions, and motifs involved with viral immune interference. **N-linked glycosylation** of the **PNG sequon** is a well-known example used by viral glycoproteins as camouflage against immune recognition [10,11]. The distribution of N-linked glycosylation sites has recently been recognized as essential for the design of immunogens to induce broadly cross-reactive immune protection against such challenging viruses as HIV-1 [12,13]. Motifs associated with cellular trafficking (localization, transport, secretion, and sequestration) are readily edited to modify where expression products go, and change interaction profiles with other proteins [14]. In addition to motifs that stabilize the structure of immunogens, such as trimerization ('foldon') [15–18] and dimerization [19,20] **domains**, motifs that interact with cellular processes for innate antiviral pathways could be used to enhance immunogenicity. While SLiMs in eukaryotic proteins have been discussed extensively, SLiM involvement in viral immunomodulation remains less thoroughly explored, and suggests new opportunities for use in engineered biotechnology applications.

The ability to transfer genetic components across species, or to introduce such components *de novo*, enables new functions. While such functions are generally well intended, some risk also exists for harmful effects. Subject to the technical advances of synthetic biology, such effects are not

### Highlights

Short linear motifs (SLiMs) are patterns of three to ten consecutive AAs used by eukaryotic cells for tasks that include: signaling, localization, degradation, and proteolytic cleavage.

Viruses use SLiMs to their advantage, including interference with antiviral innate immune pathways.

Viral SLiMs can tolerate mutations, evolve quickly to modify host interactions, and co-occur in a modular manner or involve multi-protein complexes.

SLiMs are useful in synthetic biology, where minor edits can alter target specificity, modulate persistence, reprogram interactions with cell-signaling domains, and alter protein function in myriad other ways.

Aside from possible beneficial uses, for example, to produce better immunogens and develop therapeutic interventions against infectious disease, SLiMs may help characterize new and emerging threats to global health.

<sup>1</sup>Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>2</sup>Biosciences Division, SRI International, 333 Ravenswood Ave, Menlo Park, CA 94025, USA

<sup>3</sup>Artificial Intelligence Center, SRI International, 333 Ravenswood Ave, Menlo Park, CA 94025, USA

<sup>4</sup>Biosecurity and Public Health, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>5</sup>Global Security Directorate, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

\*Correspondence: [pth@lanl.gov](mailto:pth@lanl.gov)



necessarily a taxonomically relevant property. It may be necessary to evaluate risks of new functions by other means than taxonomy or even protein functional evaluations. Instead, new methods are needed that assess functions at a finer resolution than the gene, whether by computational analysis or functional phenotypic assessments [21–23]. SLiM analysis might help with such assessments.

## Viral Immunomodulatory Proteins

Viral proteins can modulate immunity in several ways, which include: shutdown of host macromolecular synthesis, inhibiting antigen production or apoptosis, and interference with such processes as antigen presentation by MHC, natural killer (NK) cell function, antiviral cytokines, or interferon responses. Each of these processes involves coordination among multiple components in host cells. Viral interference with these functions is frequently attributed to entire proteins, but in some important cases has been localized to SLiMs.

## Signaling Interference

Because of their compact size, SLiMs are modular, rapidly evolvable sequence elements. Different instances of a given SLiM can vary in sequence while maintaining the overall functional profile, that is, the **regular expression** for the sequence motif, where a few positions are invariant while other positions tolerate numerous substitutions. Thus, partial sequence matches are sufficient for transient binding interactions with target domains, for example, signal transduction proteins. This observation led to the proposal of *ex nihilo* SLiM evolution – the evolution of a novel SLiM ‘from nothing’ – the appearance of a new functional module from a previously nonfunctional region of protein sequence [24]. Because hosts’ interaction networks are often conserved, SLiMs represent a significant vulnerability for opportunistic exploitation. These properties enable pathogens to acquire host-like SLiMs rapidly through *ex nihilo* convergent evolution, to rewire host interaction networks, and to acquire tropism and virulence traits needed for successful adaptation and propagation [25]. Over 200 motifs are known, with 2,400 validated instances, and many more motifs may await discovery [3,26]. Focus on viral motifs may reveal practical utility, to broaden the repertoire of tools available to reprogram molecular function in synthetic biology.

One example of how viral proteins use SLiMs to subvert host cell function is illustrated by Epstein–Barr virus (EBV), which persists in resting memory B cells of nearly all (>95%) individuals throughout their adult lives [5]. Latent membrane protein 1 (LMP1) is central to EBV persistence. The cytoplasmic tail of this membrane-bound protein includes PxxPxP and PxQxT motifs that recruit signaling proteins (JAK3, that is, Janus kinase 3, and several TRAFs, tumor necrosis factor receptor-associated factors, respectively) [5]. Together, the motifs mimic the cytoplasmic domain of CD40 to activate nuclear factor- $\kappa$ B via intermediates, including a third motif YYD\$ (where \$ denotes the C terminus), the TRADD (tumor necrosis factor receptor-associated death domain) binding domain. The overall result is that LMP1 inhibits apoptosis and infected B cell proliferation, to confer viral persistence [5]. Other examples of viral SLiM contributions to motif mimicry involved with immune function include: protein degradation, transcription, translation, and transport into and out of the nucleus [5].

## Viral Gene Ontologies

Given the continued growth of this field [27], established frameworks can manage and exploit this knowledge beyond catalogues of currently known motifs [5,7,28] or details on contributions of one viral protein (e.g., [14,29]). Work to use SLiMs in bioengineering can benefit from understanding viral protein function. This information is organized in viral knowledge bases, such as **ViralZone** (Box 1). Ontologies describe systematically the many different functional roles of viral proteins, including immune evasion. By promoting use of standard terms for relationships between concepts, an **ontology** arranges concepts into a framework that can be updated as knowledge grows. Protein function is captured broadly in such a framework, though the nuanced details of interactions with other molecules are not localized to domains or motifs.

GO is an authoritative resource for annotating functions of gene sequences [30,31]. An example of interest is ‘evasion or tolerance by virus of host immune response’ [32] ([www.ebi.ac.uk/QuickGO/](http://www.ebi.ac.uk/QuickGO/)

## Glossary

**Adjuvant:** an additive to a vaccine that promotes nonspecific immune responses. When administered together with an antigen, it induces more potent responses than the antigen alone.

**Autophagy:** an evolutionarily conserved degradation system for maintaining cellular homeostasis and innate immunity to clear pathogens from cells.

**Domain:** structure-based modular subunit of a protein, often with a specific function. Domains are generally larger and associated with structural subunits, while motifs are shorter and associated with intrinsically disordered protein regions.

**ELM:** the eukaryotic linear motif resource ([elm.eu.org](http://elm.eu.org)), a repository of known SLiMs; includes annotation from primary literature and information about experimental assays used.

**Glycosylation:** post-translational modification by host transferases linking sugar molecules to side chains of either nitrogen (N-linked) in asparagine or oxygen (O-linked) in serine or threonine. Viral proteins like HIV-1 envelope can be heavily glycosylated, providing camouflage against immune recognition, shifting as the PNG sequon mutates.

**GO:** Gene Ontology, an ontology for genetic products of any and all organisms, providing a framework to annotate gene and protein functions in genetic sequence databases and bioinformatics analysis procedures. GO includes multiple dimensions to capture biological complexity in adequate depth: molecular function, cellular component, and biological process. GO development is conducted by a consortium of research communities and databases, which regularly solicits input and feedback from the broader research community.

**Immune modulation:** interference with an immune-related process by a pathogen.

**Intrinsically disordered protein/domain:** regions of a protein sequence that are predicted or experimentally shown not to form consistent structure, e.g.,  $\alpha$  helices or  $\beta$  sheets. Such regions tend to be more accessible for interactions with other proteins.

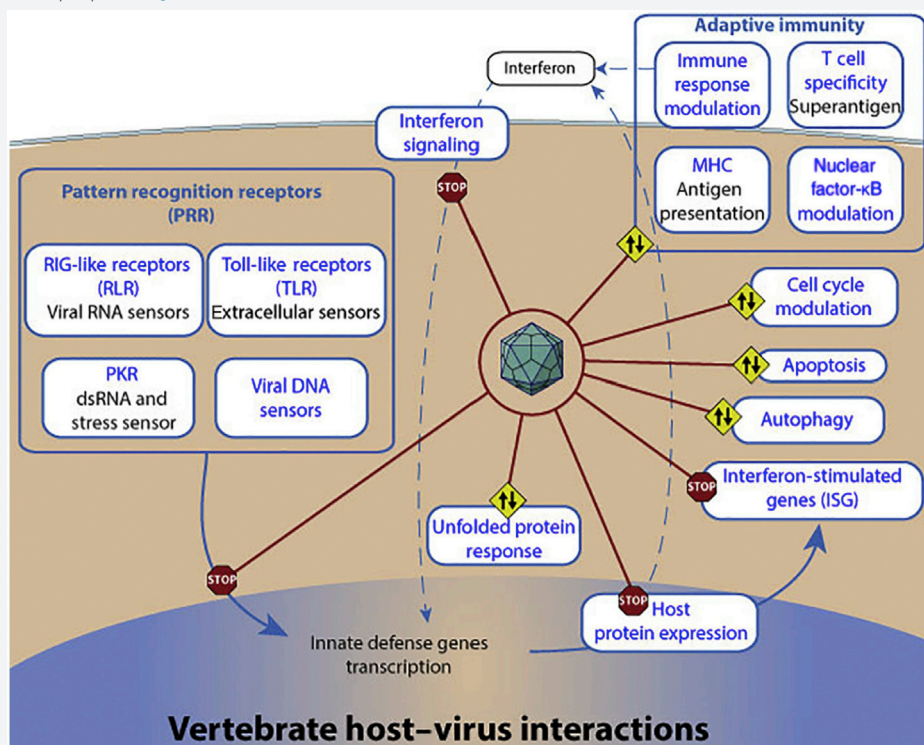
term/GO:0030683, Figure 1). Concepts are hierarchically organized, and include a definition, synonyms, and lists of parents and children. Functional annotation in GO reflects the diverse effects of viral proteins on immune interference.

Modulating **autophagy** is an example of recent advances in this research area [33,34]. A growing number of reports describe how virus proteins and SLiMs therein modulate autophagy to promote various aspects of their life cycle [34,35]. Both GO and ViralZone have developed concepts to detail

### Box 1. The ViralZone Ontology Outlines Viral Protein Function

A bridge that links literature reports to GO term annotation, ViralZone is an online knowledge base that contains 'textbook' information about viral taxonomy, replication, genome organization, and virion structure, and provides links to viral sequence data [62,63]. Importantly, ViralZone staff collaborate with the GO Consortium to define entries for virus-specific molecular functions [63–66]. ViralZone cross-references its keywords with GO Consortium terms and UniProt [67] identifiers. This makes it possible to search for viral proteins by their functional role.

ViralZone staff have developed GO concepts specifically for viruses, to represent the diversity of viral replication and processes involved with viral entry, replication, and egress [66]. ViralZone staff have also developed a detailed listing of virus–host interactions, with entries for 65 functions and 57 GO terms [64,65]. Each entry ('keyword') has a unique identifier. Unlike Enzyme Commission (EC) numbers [68], ViralZone IDs are arbitrary numbers and do not indicate position in the concept hierarchy. Instead, organization of the keyword hierarchy is provided online. The web address <https://viralzone.expasy.org/886> is an entry point into the ViralZone concept space (Figure 1).



Trends in Biotechnology

**Figure 1. ViralZone Summarizes What Is Known about Viral Proteins Involved in Virus–Host Interactions.**

Blue text indicates a link to more detail. Shown is the vertebrate host–virus interactions page [64]. Also available are summaries for invertebrate, plant, and bacterial host–virus interactions.

**Molecular mimicry:** structural similarity that enables repurposing or hijacking of molecular function by pathogens, such as viruses. Molecular mimicry varies in extent from entire globular proteins, to localized domains, down to short linear motifs.

**Motif class:** a regular expression that summarizes known variants to define a sequence motif.

**Motif instance:** a particular motif as found in a protein or translated genetic sequence from a specific organism, strain, or isolate.

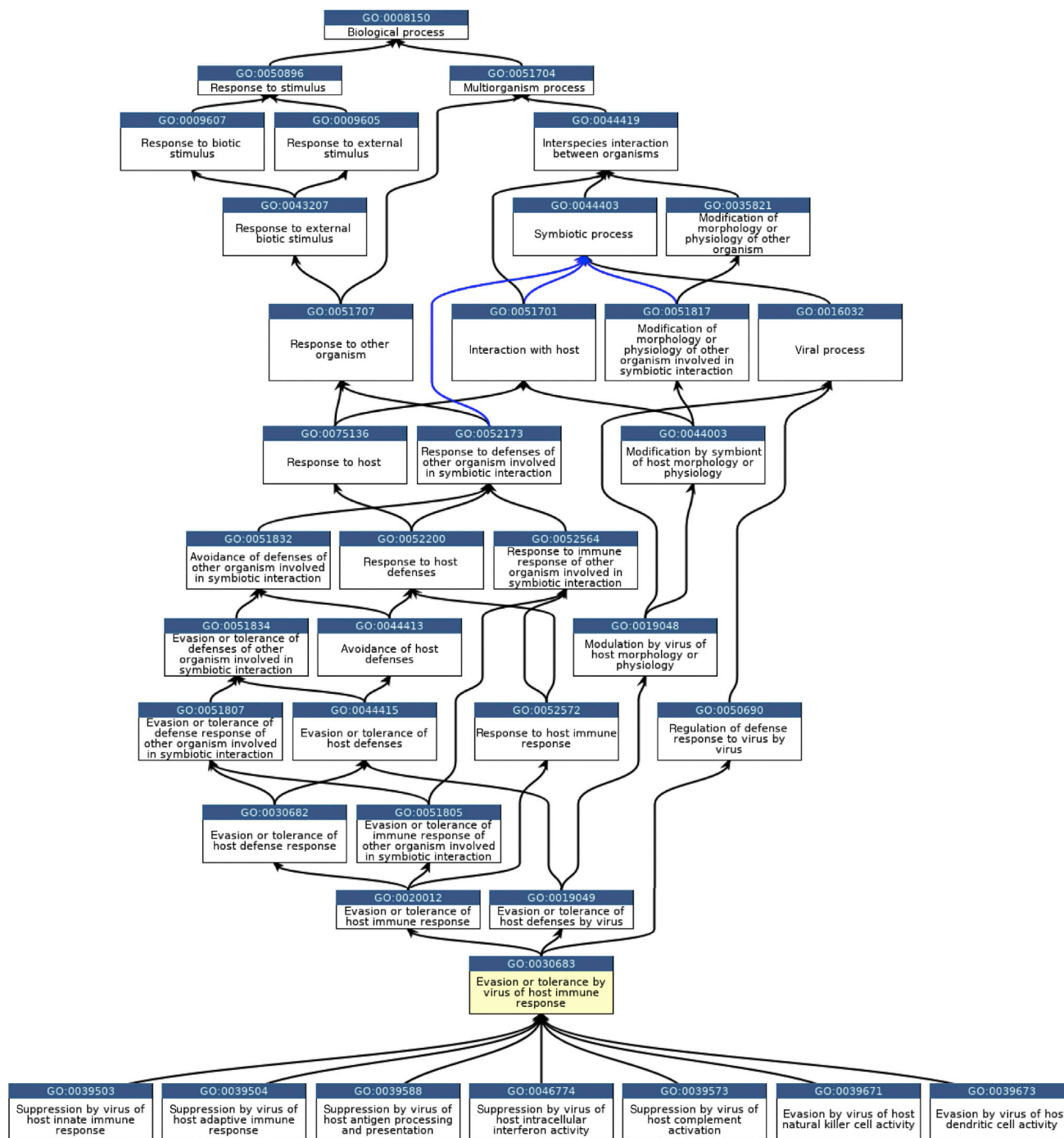
**Ontology:** a representation of domain-specific knowledge organized as concepts, their properties, and their relationships.

**PNG sequon:** three AA motif N [^P][ST], that is, asparagine, then any AA except proline, then serine or threonine, recognized by glycosyltransferase as a potential N-linked glycosylation site.

**Regular expression:** a string of characters that concisely represents many alternative sequence variants; may include wildcards to represent any character, groupings of possible characters, repetition, negation, start and end of sequence, etc.

**Short linear motif (SLiM):** also known as MiniMotifs or MoRFs (molecular recognition features). Frequently represented as regular expressions, typically three to ten AAs long.

**ViralZone:** a knowledge base (<https://viralzone.expasy.org>) that documents viral families, genome architectures, proteins, host–protein interactions, and an ontology for the functions of viral proteins.



Trends in Biotechnology

**Figure 1. Immune-Evasion Concept Hierarchy.**

Adapted from [www.ebi.ac.uk/QuickGO/term/GO:0030683](http://www.ebi.ac.uk/QuickGO/term/GO:0030683) [32]. Most arrows indicate 'is a' relations, where the hierarchy is refined by specialization. Blue arrows indicate 'part of' relations, which relate to the symbiont process as parts to a whole.

autophagy processes, including positive and negative regulation of xenophagy, the selective autophagy of pathogens [33–35].

Understanding the functional roles of SLiMs can help identify related mechanisms or processes, or possibly identify knowledge gaps where SLiMs may be posited but not yet identified. An overview of SLiM functions may also help to prioritize which are of greatest potential for use or abuse when artificially added to modify protein function. Databases and discovery tools are also useful to identify known and new SLiMs.

## Motif Databases

Identification of shared structural features across divergent protein families led to analysis and identification of modular protein domains. Protein domains are used to categorize protein function, and the InterPro database [36] ([www.ebi.ac.uk/interpro](http://www.ebi.ac.uk/interpro)) aggregates information at this within-protein level. The identification of protein domains led to recognition of SLiMs as compact, small-scale functional modules [37].

ELM is a database of eukaryotic motifs (Box 2), though its representation of viral–host interactions is not fully developed. At present, 264 ELM entries map to 648 GO terms. Most motifs map to multiple GO terms; the median is seven GO terms per motif and the maximum is 29 (MOD\_Plk\_1). Immune-associated function of the LIG\_IRF3\_LxIS\_1 motif is involved in signal transduction responses to pathogen-associated molecular patterns; this motif maps to 25 GO terms, but none occur in the ViralZone vocabulary. In total, only three ELM entries utilize GO terms from ViralZone: LIG\_BH\_BH3\_1, LIG\_HCF-1\_HBM\_1, and LIG\_Rb\_pABgroove\_1. These three motifs all map to the most general

### Box 2. Eukaryotic Linear Motif Resource Curates Confirmed Viral Motifs

ELM assigns motif classes to one of six functional categories [69,70]:

- (i) CLV, proteolytic cleavage sites;
- (ii) DEG, degradation sites, part of polyubiquitination;
- (iii) DOC, docking sites, involved in protein recruitment but not directly targeted by an active site;
- (iv) LIG, ligand binding sites, primarily for protein–protein interactions;
- (v) MOD, post-translational modification sites; and
- (vi) TRG, targeting sites for subcellular localization.

ELM has also spun-off several specialized databases: phospho.ELM for phosphorylation sites with experimental evidence [71], switches.ELM for conditional molecular switches, such as requiring that a site be modified [72], and iELM, with an emphasis on protein–protein interactions [73]. A detailed tutorial provides orientation for ELM use [74].

ELM documents each motif class with a concise description of its function. For example, one type of nuclear localization signal (NLS), TRG\_NLS\_Bipartite\_1 the ‘classic bipartite NLS’, which binds to importin- $\alpha$  for nuclear pore transfer and is utilized by the PB2 protein of influenza A, is documented here: [elm.eu.org/elms/TRG\\_NLS\\_Bipartite\\_1](http://elm.eu.org/elms/TRG_NLS_Bipartite_1). The abstract and functional site descriptions summarize what is known about the motif.

ELM provides a virus-specific summary at [elm.eu.org/viruses.html](http://elm.eu.org/viruses.html) and serves downloads in multiple formats from there or [elm.eu.org/downloads.html](http://elm.eu.org/downloads.html). ELM presently contains 53 viral motif classes and 246 viral motif instances. Viral motif instances are distributed over the functional categories as follows: 4.5% CLV, 1.6% DEG, 4.1% DOC, 47.8% LIG, 26.5% MOD, and 15.5% TRG. Of the 246 instances, the most common (50, or 20.3%) are N-linked glycosylation sites from six distinct viruses. The most commonly represented viruses with motif instances in ELM are HIV-1 (11.8% of all entries) and severe acute respiratory syndrome-related coronavirus, that is, SARS-CoV (5.7%), though again, these are dominated by the N-linked glycosylation sites: 25 of 29 HIV-1 entries and 13 of 14 from SARS-CoV. More viral motif classes and instances will surely be added to ELM in time.

GO term, GO:0019048 ('modulation by virus of host morphology or physiology', a synonym of 'virus–host interaction'). This underscores the prevalent mode of ELM motif discovery and annotation does not emphasize host–virus interactions, but rather systems-level interactions within eukaryotic cells. Thus, better integration of ViralZone–GO-term vocabulary with ELM or another domain-level representation of viral SLiMs is needed to promote potential utility for biotechnology.

### Integration of Resources

Despite not using the ViralZone ontology, ELM documents other **motif classes** with GO terms that refer to 'viral', 'virus', 'immune', or 'immunity' (Table 1). Because the focus is motif function in the host cell context, ELM does not directly indicate how viral immune interference results. Further, the relative lack of viral motifs in ELM does not indicate their absence *in vivo*, but rather the evidence-based requirement for ELM inclusion.

Related to the earlier observation, a review of how viruses use SLiMs to interfere with host cells [5] lists 52 examples that represent viral mimicry of host SLiMs (Table 2). Only 70% of these have corresponding ELM entries, though the SLiMs are known. The remaining 30% indicate that ELM does not fully capture all known viral motifs. This strong requirement by ELM for evidence-based motif classes and instances is not strictly a drawback. Indeed, the ELM creators are very aware that computational analysis alone is error prone and can yield misleading outcomes. In [38], they discuss this issue in depth, and recommend a workflow for SLiM discovery that culminates in experimental validation, whether *in vivo* or *in vitro*. Working with viral–host systems adds layers of difficulty to experimental motif validation, so it should not be surprising or to the detriment of available information resources that viral SLiMs are less thoroughly documented.

Searching arbitrary sequences for **motif instances** is computationally straightforward. Box 3 provides an example of motif searching with ELM, which might facilitate comparative analysis of two related proteins from different species of human herpesvirus (HSV). Resources such as InterPro and UniProt are able to perform similar assessments, but give broader, domain-level representations with less functional detail than the SLiM searches enabled by ELM. Reports in the primary literature take a

**Table 1. Number of ELM Viral Motif Instances with Virus-Related or Immune-Related GO Terms**

Regular expression	Viral instances	Motif	Role
<b>a</b>	15	LIG_Rb_LxCxE_1	Binds retinoblastoma B pocket
.P[TS]AP.	13	LIG_PTAP_UEV_1	UEV domain binding PTAP motif
[LM]YP.[LI]	11	LIG_LYPXL_S_1	Endosomal sorting of membrane proteins
[FVILMY].FG[DES]F	8	LIG_G3BP_FGDF_1	Binds Ras GTPase activating SH3 domain
<b>b</b>	5	LIG_KLC1_WD_1	Binds kinesin light chain TPR region
[LM]YP.[LI]	2	LIG_LYPXL_L_2	Endosomal sorting of membrane proteins
[DE]H.Y	1	LIG_HCF-1_HBM_1	Binds transcriptional coactivator HCF-1
..L.I(S)	1	LIG_IRF3_LxIS_1	Interferon regulatory factor 3 binding site
.[VILM].[LM][FY]D.	1	LIG_Rb_pABgroove_1	Binds retinoblastoma AB groove
<b>c</b>	0	LIG_BH_BH3_1	Binds BH domains to inhibit apoptosis
EP[ILY][TAG]	0	LIG_CSK_EPIYA_1	Binds C-terminal Src kinase SH2 domain

<sup>a</sup>[(DEST)]^(.){0,4}[LI].C.E.{1,4}[FLMIVAWPHY].{0,8}[(DEST)]\$)

<sup>b</sup>[LMTAFSR][^KRG]W[DE].{3,5}[LIVMFPA]

<sup>c</sup>....[LIFVYMT][ASGC][^P]{2}[^P]{2}[IVMTL][GACS][D][^P][FVLM]

Table 2. Examples of Viral Proteins That Mimic Host SLiMs, updated from [5]

Host target	Viral protein	Virus	Motif <sup>a</sup>	ELM
CDH1	E1	BPV	KEN	DEG_APCC_KENbox_2
Phosphodegron FBW7	LT	SV40	TPxxE	DEG_SCF_FBW7_1
Phosphodegron $\beta$ TrCP1	Vpu	HIV	DSGxxS	DEG_SCF_TrCP1_1
SIAH1	ORF45	KSHV	PxAxV	DEG_SIAH_1
Tankyrase	EBNA1	EBV	RxxPDG	DOC_ANK_TNKS_1
Cyclins	E1	HPV	RxLF	DOC_CYCLIN_1
PP1	$\gamma$ 134.5	HSV	RVxF	DOC_PP1
Calcineurin	p12	HTLV1	SPxLxLT	DOC_PP2B_1
USP7	EBNA1	EBV	PxE[ <sup>^</sup> P]xS[ <sup>^</sup> P]	DOC_USP7_MATH_2
14-3-3	Rep68	AAV	RSxSxP	LIG_14-3-3_CanoR_1
Clathrin heavy chain	HDAg-L	HDV	LFxAD	LIG_Clathr_ClatBox_1
TR	E1A	Adenovirus	LxxLxxxL	LIG_CORNBOX
CtBP SDB	E1A	Adenovirus	PxDLS	LIG_CtBP_PxDLS_1
Dynein light chain 8	P	Rabies	KxTQT	LIG_Dynein_DLC8_1
ALIX	Gag	HIV	LYPxxxL	LIG_LYPXL_L_2
BS69	EBNA2	EBV	PxLxP	LIG_MYND_1
PDZ domain	E6	HPV	TxV\$	LIG_PDZ_Class_1
Tsg101	Gag	HIV	PTAP	LIG_PTAP_UEV_1
RB (pocket region)	E7	HPV	LxCxE	LIG_RB_LxCxE_1
RB (E2F competition)	E1A	Adenovirus	LxxLYD	LIG_RB_pABgroove_1
Integrin $\alpha$ 5 $\beta$ 3	VP1	FMDV	RGD	LIG_RGD
SH2 domain	stpC	HVS	YxxV	LIG_SH2
SH3 domain	Nef	HIV	PxxP	LIG_SH3_2
TRAF2	LMP1	EBV	PxQxT	LIG_TRAF2_1
TRAF6	U(L)37	HSV	PxExxE	LIG_TRAF6
Syk	LMP2A	EBV	Yxx $\phi$ Yxx $\phi$	LIG_TYR_ITAM
Farnesyltransferase	HDAg-L	HDV	Cxxx\$	MOD_CAAXbox
Oligosaccharyltransferase	E1	HCV	N[ <sup>^</sup> P][ST]	MOD_N-GLC_1
N-myristoyltransferase	G9R	Vaccinia	<sup>^</sup> GxxxS	MOD_Nmyristoyl
PIAS1	IE1	HCMV	IKxE	MOD_SUMO
AP-2 $\mu$	Env	SIV	Yxx $\phi$	TRG_ENDOCYTIC_2
COP1	E3	Adenovirus	KK	TRG_ER_diLys_1

(Continued on next page)

Table 2. Continued

Host target	Viral protein	Virus	Motif <sup>a</sup>	ELM
ERD2	ctxA	Phage CTX	KDEL	TRG_ER_KDEL_1
AP-1	Nef	HIV	ExxxLL	TRG_LysEnd_APpAcLL_1
NESc	Rev	HIV	Ψ-rich	TRG_NES_CRM1_1
NLS, bipartite	PB2	Influenza	KR-rich	TRG_NLS_Bipartite_4
NLS, monopartite	LT	SV40	KR-rich	TRG_NLS_MonoCore_1
CtBP NDB	E1A	Adenovirus	RxxTG	
p300/CBP	E1A	Adenovirus	FxDxxxL	
Caspases	NS1	ADV	DxxD↓	
NEDD4	VP40	Ebola	PPxY	
SEC24C	VP40	Ebola	LxMVI	
JAK	LMP1	EBV	PxxPxP	
TRADD	LMP1	EBV	YYD\$	
Elongin C	Vif	HIV	SLxxxLxxl	
PACS1	Nef	HIV	EEEE	
HCF	VP16	HSV	EHxY	
NoLS	γ134.5	HSV	KR-rich	
Furin	Spike	IBV	R↓S	
PKR	NS1	Influenza	IMxKN	
H2A–H2B	LANA	KSHV	MxLRSG	
Palmitoyl acyltransferase	G	Rabies	CC	

<sup>a</sup>The down arrow (↓) indicates a cleavage site; φ (phi) represents a site occupied by a hydrophobic [VILFWYM] and Ψ (Psi) an aliphatic [VILM] AA. Other motif symbols are regular expression terms. For example, ^ indicates sequence N terminus, but in brackets indicates negation. That is, [^P] indicates any AA except proline. Motifs stated do not necessarily correspond to the general motif patterns currently in ELM.

different approach, by marking SLiMs in a protein alignment, which includes orthologues to mark conservation (e.g., Figure 3 in [39]). The ELM-generated report combines predicted SLiMs with information from annotated domains and local disorder predictions, for a perspective that complements the other approaches.

Clearly, false positives are inevitable among SLiM search results. This makes it necessary to filter for the most significant and informative outcomes. This leads to consideration of *in silico* (computational) methods for SLiM evaluation. A highly recommended, authoritative review of SLiM discovery techniques, from an author of the SLiMSuite software package, discusses motif identification techniques in depth [40].

## Motif Discovery

Methods for SLiM discovery can be divided into two broad classes: (i) *de novo* discovery of new SLiMs, and (ii) instance prediction to find new occurrences of known SLiMs. There are currently at least eight software packages available to discover new SLiMs and 40 packages (25 stand-alone programs or servers and two software suites that consist of multiple tools: SLiMSuite [40], which includes ten utilities, and MEME [41], which consists of five tools for SLiM instance detection). Though MEME was

developed for discovery of DNA sequence motifs, it generates ungapped, profile-based motifs using the expectation–maximization (EM) algorithm. No single method is inherently better than the rest, but the choice of which to use depends on several factors, such as the input sequence data, whether one sequence, an alignment, or a collection of nonhomologous sequences [40]. To illustrate the diversity of motif discovery methods, this section mentions only a handful of the software tools available (Table 3). Readers seeking to learn more about the full set of alternatives are strongly encouraged to consult [40], particularly Tables 1, 4, and 5 therein. Another helpful resource (Table 1 in [38]) lists online motif discovery bioinformatics services.

### De Novo Methods

Several alternative approaches for discovery of new motifs have been advanced. Edwards and Palopoli [40] review the alternatives in depth, discussing their merits and drawbacks. Briefly, they can be divided into alignment-based and alignment-free methods. An alignment-based approach looks for conserved sites among homologous sequences, but can be misled by high sequence conservation in globular domains. A program called SLiMPrints works around this with a specialized approach to model substitutions [42]. SLiMPrints uses a statistical model of relative local conservation, which looks for clusters of overly constrained sites in a window of about 30 AAs, using IUPred scores (intrinsically unordered prediction; see later) to weigh sites in **intrinsically disordered protein** regions more heavily than sites in globular (ordered) regimes [42].

In contrast, alignment-free methods look for enrichment of amino acid patterns in proteins that are expected by other means to perform similar motif-related roles, for example, by GO category annotations or protein–protein interaction (PPI) data, that is, via databases that capture experimental evidence for protein colocalization and functional interactions. An important caveat is that to assume such sequences are independent could yield spurious enrichment of shared patterns, so alignment-free methods need to compensate for evolutionary constraints at the domain level, rather than for full-length homologous proteins. The development of such corrections and their relative advantages are detailed in [40]. Some programs (e.g., SLiMDisc [43], SLiMfinder [44,45], and DILIMOT [46,47]) produce regular expressions that compensate for phylogenetic relatedness, while others (MEME suite, GLAM2 [48], and NestedMICA [49,50]) produce probabilistic profiles. For more discussion of these and issues of concern for computational motif discovery, see [40].

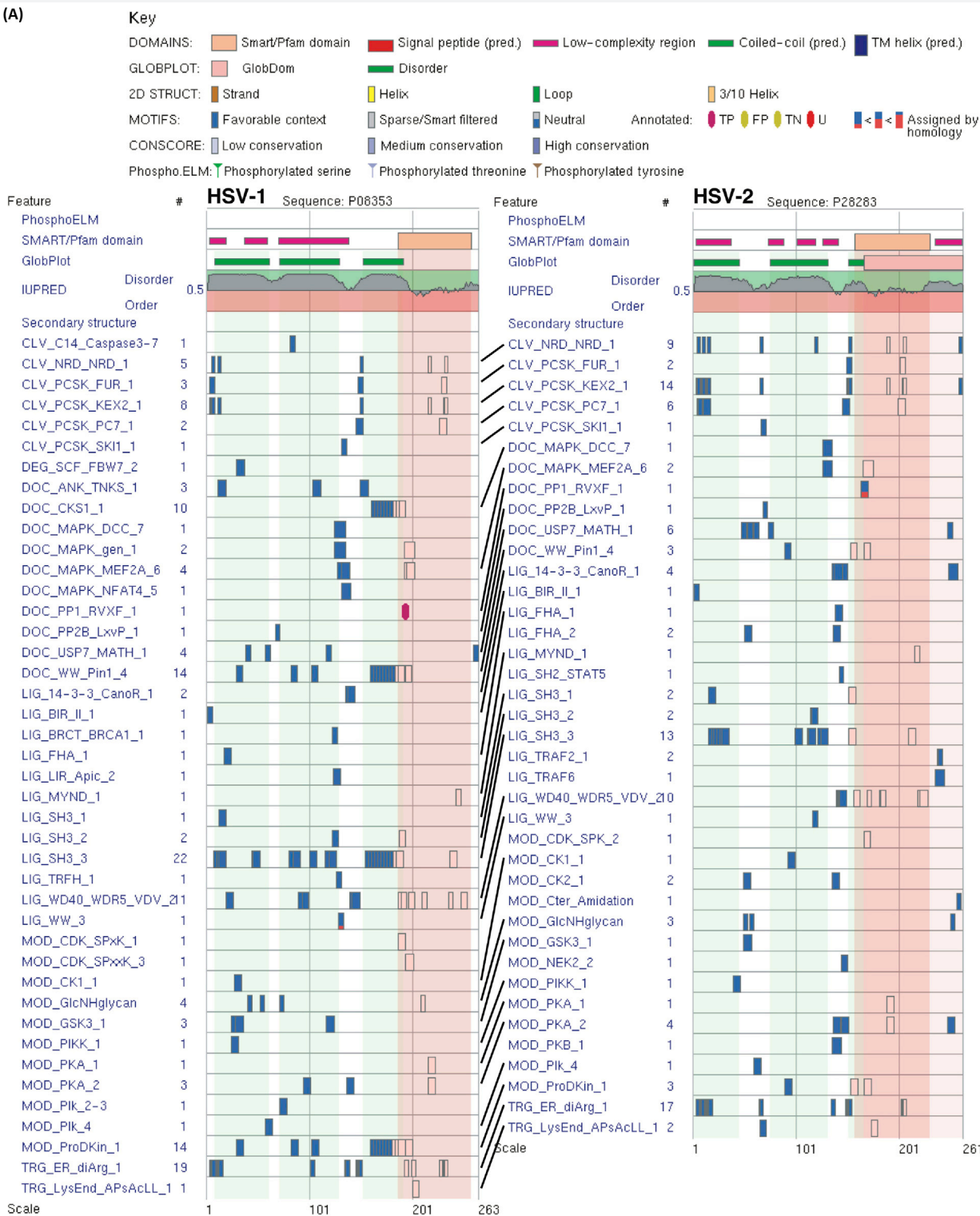
### Instance Detection

Filtering methods control high false positive rates from SLiM instance detection. Structural information, whether known or predicted, can be used for filtering. Box 3 illustrates how ELM filters results to

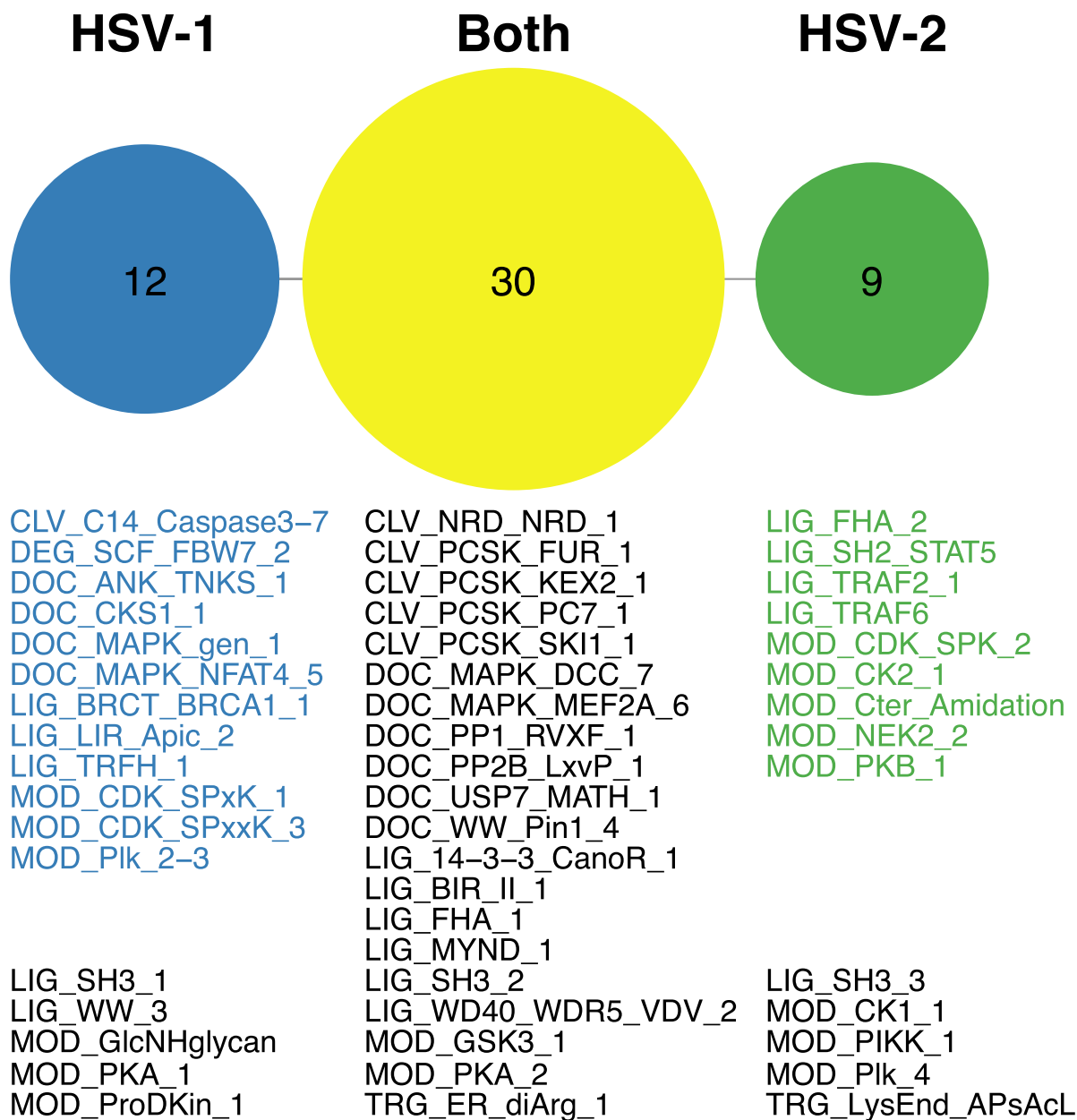
#### Box 3. Motif Searches Enable Comparison of Neurovirulence Proteins

HSV-1 and HSV-2 virulence factor ICP34.5 assists in viral immune evasion by molecular mimicry. HSV-1 neurovirulence protein ICP34.5, encoded by the  $\gamma$ 34.5 gene, initiates immune interference by binding and sequestering cellular proteins that would stimulate autophagy, translational arrest, and type I interferon responses. HSV-1 ICP34.5 binds TANK-binding kinase 1 (TBK1) to prevent type I interferon induction [75], Beclin-1 to prevent autophagy [76], and both PP1 $\alpha$  and eIF2 $\alpha$  to overcome translational arrest [77]. HSV-2  $\gamma$ 34.5 contains an intron not present in HSV-1, and up to four isoforms of HSV-2 ICP34.5 are known [78]. Full-length HSV-2 ICP34.5 has conserved PP1 $\alpha$  and eIF2 $\alpha$ -binding domains, but lacks TBK1 and Beclin-1 binding domains [79]. Additional HSV-1 motifs influence intracellular localization [80], virion maturation, and egress [81], not yet characterized in HSV-2. HSV-2 is recognized as more virulent than HSV-1, but both can cause neuropathology, including viral encephalitis and meningitis [82]. To attenuate virulence, ICP34.5 is routinely deleted or inactivated when making HSV-1 constructs for oncolytic therapy [82]. Both are the same length and share domain structures, and partially share SLiM compositions (Figure 1). Identifying differences in SLiMs from each could provide clues for more detailed experimental investigations to understand ICP34.5 virulence determinants and host protein targets.

Searching ELM for motif instances in HSV-2 ICP34.5 (UniProt P28283) gives 127 instances of 39 motif classes, before filtering to exclude globular domains and other likely false hits. Filtering leaves 97 instances, which cover almost all of this 261 AA protein except for a predicted globular domain of 73 AAs. A similar query for the HSV-1 protein (UniProt P08353) gives 158 instances of 42 distinct motifs, and 114 instances of 36 motifs when filtered. Most of the ELM motif hits are probably false positives, if only because the motif expressions overlap. The patterns (.RK)(RR[ $\wedge$ KR]), [KR]R., and R...[KR]R., RRGPRRR, MSRRR, and RR all match sites in the first ten AA positions of the HSV-2 query protein sequence, a low-complexity region dominated by positively charged AAs. The HSV-1 version contains a 30 amino acid ATP triplet repeat at sites 160–190, not found in the HSV-2 protein. ELM matches four motifs here, only one of which (DOC\_CKS1\_1) is unique to HSV-1.



(B)



Trends in Biotechnology

**Figure 1. Viral SLiM Comparisons Using ELM. (A) ELM-generated plots of motif locations in neurovirulence factor ICP34.5 from HSV-1 and HSV-2 [74,83]. (B) Summary of motifs found in one or both.**

exclude a region predicted to fold as globular protein. These predictions were made by SMART (simple modular architecture research tool) [51] and Pfam [52] domain matches, corroborated by GlobPlot [53]. Another widely used approach is to identify regions of local disorder, where protein structure is not clearly defined, making that region accessible to interact with other proteins. IUPred [54] is commonly used for this task, though the choice of parameter settings and how to interpret results varies. ELM results include an IUPred disorder score and a simple cutoff of 0.5 to define the disorder transition. Above this value, local protein structure is considered accessible for interaction with other proteins.

Scoring schemes filter for statistical enrichment of motif instances. An approach of filtering by homology [55] seems inappropriate for use to detect virus interactions with host proteins, as it may exclude nonhomologous regions with motifs that do interact, yielding false negatives. Regardless, failure to consider evolutionary relatedness among sequences being searched could introduce bias due to common ancestry, rather than independence, among sequences.

A simple approach to instance prediction is a stand-alone program called ShettiMotif [56]. It was used to scan 2251 protein sequences from 11 Poxviridae genomes (an average of 205 proteins per poxvirus) for low-complexity regions and regular expressions defined by PROSITE. The approach compared numbers of proteins per genome that carry each motif, and doubtlessly includes many motif instances that are not functional as SLiMs. Also, shorter motifs occur more frequently than longer motifs [3,27], partly due to chance alone. Regardless, systematic error may be considered a source of background noise across the large number of proteins in 11 viral proteomes, each having different host specificities, to enable somewhat meaningful comparisons, in such a 'statistical genomics' approach [3]. The comparisons could be more meaningful if false positive motif instances were reduced.

Becerra *et al.* developed another approach to instance counting [57], which involves comparison with a null distribution from permuting primary sequence and testing for presence of the motif in the permuted sequences. A motif is considered rare and therefore significantly unlikely to occur by chance if it is present at or below some cutoff frequency. Restricting the sequence region that is used for permutation testing, such as by use of structural considerations, can further focus the search. Indeed, such a hybrid filtering approach was described recently and evaluated on the HIV-1 proteome [57]. Following methods described in an earlier study [58], Becerra *et al.* used IUPred with a modified, window-based scoring procedure to identify intrinsically disordered protein regions, and tested for statistical rarity below 1% of 1000 shuffled variants. The approach further considered conservation above 70% in a set of aligned sequences, though combining three filtering criteria was too stringent and excluded all motif candidates [57].

## Motif-Specific Databases

While algorithmic approaches seek to identify a broad range of SLiM types, more specialized resources have emerged to track the distribution of a particular SLiM in viral proteins. For example,

**Table 3. Computational Resources for SLiM Discovery**

Program	Website	Refs
SLiMSuite	<a href="https://github.com/slimsuite/SLiMSuite">https://github.com/slimsuite/SLiMSuite</a>	[40]
SLiMfinder	<a href="https://www.slimsuite.unsw.edu.au/servers/slimfinder.php">https://www.slimsuite.unsw.edu.au/servers/slimfinder.php</a>	[45]
DILIMOT	<a href="https://dilimot.russelllab.org">https://dilimot.russelllab.org</a>	[46,47]
GLAM2	<a href="https://meme-suite.org/tools/glam2">https://meme-suite.org/tools/glam2</a>	[48]
NestedMica	<a href="https://www.mybiosoftware.com/nestedmica-0-8-0-motif-finder.html">https://www.mybiosoftware.com/nestedmica-0-8-0-motif-finder.html</a>	[49,50]
ShettiMotif	<a href="https://sites.google.com/site/haithamsobhy/ShettiMotif_V1.zip">https://sites.google.com/site/haithamsobhy/ShettiMotif_V1.zip</a>	[56]

## Outstanding Questions

What incentives or community-based activities would best enable integration of specialized viral gene ontologies with databases of motif classes and instances? To validate new SLiMs, what are the needed procedures, which simultaneously minimize spurious results and make the most promising candidates available to use and discover new instances?

How prevalent are immunomodulatory motifs in viruses, relative to the prevalence of entire viral proteins dedicated to this specialized function? To what extent can viral immunomodulatory function be localized to a motif or domain, or is the larger whole-protein context necessary for function?

What types of motif interactions are most common and important for viral immune modulation? Are they accurately reflected in the current literature and databases, or do many new types of motifs still await discovery?

Is there sound support for an enrichment (or scarcity) of particular motif types in certain viral classes (i.e., Baltimore classification)? If so, what does this reveal about commonalities among viral replication strategies and potential for broad-spectrum antiviral treatments?

What host countermeasures are involved with overcoming viral immune interference, either in general against many viruses, or in particular, against specific taxa? Do these countermeasures explain some of the crosstalk among interactions in antiviral innate signaling pathways?

The distribution of glycosylation sites on enveloped viruses can be extremely variable, even within one host. Such variation may impede bioengineering specific constructs. How are SLiMs influenced by dynamic evolutionary processes? What fitness costs are associated with SLiM evolution?

iLIR@viral is a web resource dedicated to detecting LIR motif-containing proteins in viruses [59]. LC3-interacting regions (LIR motifs) are SLiMs that mediate protein–protein interactions involved in autophagy, as used by influenza A virus M2 protein to subvert autophagy and maintain virion stability [60]. Using curated text mining analysis and position-specific scoring matrices, iLIR@viral analyzed 16 609 reviewed viral sequences available from UniProt across 2569 individual viral species and found that 15 589 viral sequences contain LIR motifs. While many predicted instances may represent false positives, the enrichment of LIR motifs in viral sequences is consistent with viral adaptation to host xenophagy [35]. Curiously, ELM currently lists the LIR motif as a candidate, rather than an accepted motif class.

## Concluding Remarks

Embedding SLiMs into engineered constructs may enable specific effects on cellular immune processes, for applications that include targeted drug delivery, pathogen-specific adjuvants, potent and broadly effective immunogens, transformational medical countermeasures, and improved design of vectors for gene therapy. SLiM modularity may enable easy ways to reprogram protein function with a few localized modifications. To realize the potential utility of SLiMs in synthetic biology, more research is needed to expand and integrate our collection of knowledge on viral SLiMs (see Outstanding Questions).

Detecting SLiMs in variant sequences may help to identify functional innovation or changes in virulence, in a manner that does not rely strictly on functional assessment at the whole-gene level, to identify how sequence-specific variation may interact with host responses. This may be particularly useful and important to understand new variants and assess the risk that they may spread and cause harmful effects on human health or agricultural interests. Such knowledge is needed in an era where synthetic biology may introduce new risks for biological error and biological terror. Detecting and understanding SLiM variants can help to reduce such risks and identify newly emerging threats to global health and security because watch lists for harmful organisms to ensure public safety by preventing access to select known risks may be inadequate [21–23].

SLiMs in viral proteins can interact in many different ways with host proteins to modulate immune responses. A motif may be necessary but not sufficient for any inferred function. The simplest case is where a viral SLiM interacts directly with a host protein to yield an immunomodulated phenotype. More elaborate cases are known, such as the multifunctional proteins E1A (EBV), Nef (HIV-1), and ICP34.5 (HSV). Computational prediction of SLiM classes and new instances is a process, which involves experimental confirmation and validation. High-throughput methods for experimental assessment of protein interactions are useful to validate computational predictions [38,61], and more assays are needed to evaluate functional and phenotypic effects of adding or deleting SLiMs.

## Acknowledgments

LA-UR-18-26869. Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the US Department of Energy under contract 89233218CNA000001. We gratefully acknowledge funding support from the Functional Genomic and Computational Assessment of Threats (Fun GCAT) program of the Intelligence Advanced Research Projects Agency, Office of the Director of National Intelligence. The views and conclusions contained herein are those of the authors, and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Agency (IARPA), Los Alamos National Laboratory (LANL), Department of Energy (DOE), SRI International, or the US Government.

## References

1. Garamszegi, S. *et al.* (2013) Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human–virus protein–protein interaction networks. *PLoS Pathog.* 9, e1003778

What specific constraints limit SLiM evolvability?

What strategies are most effective to advance knowledge of viral immunomodulatory SLiMs in the design of vaccines and therapies to promote global health? For example, can some viral peptides be useful as adjuvants?

2. Guven-Maiorov, E. et al. (2016) Pathogen mimicry of host protein–protein interfaces modulates immunity. *Semin. Cell Dev. Biol.* 58, 136–145
3. Hagai, T. et al. (2014) Use of host-like peptide motifs in viral proteins is a prevalent strategy in host–virus interactions. *Cell Rep.* 7, 1729–1739
4. Davey, N.E. et al. (2011) SLIMSearch 2.0: biological context for short linear motifs in proteins. *Nucleic Acids Res.* 39, W56–W60
5. Davey, N.E. et al. (2011) How viruses hijack cell regulation. *Trends Biochem. Sci.* 36, 159–169
6. Davey, N.E. et al. (2012) Attributes of short linear motifs. *Mol. Biosyst.* 8, 268–281
7. Via, A. et al. (2015) How pathogens use linear motifs to perturb host cell networks. *Trends Biochem. Sci.* 40, 36–48
8. Van Roey, K. et al. (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* 114, 6733–6778
9. Tom, J.K. et al. (2019) Applications of immunomodulatory immune synergies to adjuvant discovery and vaccine development. *Trends Biotechnol.* 37, 373–388
10. Crispin, M. and Doores, K.J. (2015) Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design. *Curr. Opin. Virol.* 11, 63–69
11. Crispin, M. et al. (2018) Structure and immune recognition of the HIV glycan shield. *Annu. Rev. Biophys.* 47, 499–523
12. Wagh, K. et al. (2018) Completeness of HIV-1 envelope glycan shield at transmission determines neutralization breadth. *Cell Rep.* 25, 893–908
13. Seabright, G.E. et al. (2019) Protein and glycan mimicry in HIV vaccine design. *J. Mol. Biol.* 431, 2223–2247
14. Pawlak, E.N. and Dikeakos, J.D. (2015) HIV-1 Nef: a master manipulator of the membrane trafficking machinery mediating immune evasion. *Biochim. Biophys. Acta* 1850, 733–741
15. Li, J. et al. (2013) Immunogenicity and protection efficacy of monomeric and trimeric recombinant SARS coronavirus spike protein subunit vaccine candidates. *Viral Immunol.* 26, 126–132
16. Lai, R.P. et al. (2014) A fusion intermediate gp41 immunogen elicits neutralizing antibodies to HIV-1. *J. Biol. Chem.* 289, 29912–29926
17. Sliopen, K. et al. (2015) Immunosilencing a highly immunogenic protein trimerization domain. *J. Biol. Chem.* 290, 7436–7442
18. Wohlbold, T.J. et al. (2015) Vaccination with soluble headless hemagglutinin protects mice from challenge with divergent influenza viruses. *Vaccine* 33, 3314–3321
19. Ruffini, P.A. et al. (2010) Human chemokine MIP1 $\alpha$  increases efficiency of targeted DNA fusion vaccines. *Vaccine* 29, 191–199
20. Poggianella, M. et al. (2015) Dengue E protein domain III-based DNA immunisation induces strong antibody responses to all four viral serotypes. *PLoS Negl. Trop. Dis.* 9, e0003947
21. DiEuliis, D. et al. (2017) Options for synthetic DNA order screening, revisited. *mSphere* 2, e00319-17
22. Wintle, B.C. et al. (2017) A transatlantic perspective on 20 emerging issues in biological engineering. *Elife* 6, e30247
23. National Academies of Sciences, Engineering, and Medicine et al. (2018) *Biodefense in the Age of Synthetic Biology*, The National Academies Press
24. Davey, N.E. et al. (2015) Short linear motifs – *ex nihilo* evolution of protein regulation. *Cell Commun. Signal.* 13, 43
25. Chemes, L.B. et al. (2015) Convergent evolution and mimicry of protein linear motifs in host–pathogen interactions. *Curr. Opin. Struct. Biol.* 32, 91–101
26. Seo, M.-H. and Kim, P.M. (2018) The present and the future of motif-mediated protein–protein interactions. *Curr. Opin. Struct. Biol.* 50, 162–170
27. Tompa, P. et al. (2014) A million peptide motifs for the molecular biologist. *Mol. Cell* 55, 161–169
28. Sobhy, H. (2016) A review of functional motifs utilized by viruses. *Proteomes* 4, 3
29. King, C.R. et al. (2018) Hacking the cell: network intrusion and exploitation by adenovirus E1A. *mBio* 9, e00390-18
30. Huntley, R.P. et al. (2014) The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res.* 43, D1057–D1063
31. (Dessimoz, C., and Skunca, N. eds) (2017). *The Gene Ontology Handbook (Methods in Molecular Biology, 1st edn)* Humana Press
32. Binns, D. et al. (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25, 3045–3046
33. Denny, P. et al. (2018) Exploring autophagy with Gene Ontology. *Autophagy* 14, 419–436
34. Wang, Y. et al. (2018) Autophagy in negative-strand RNA virus infection. *Front. Microbiol.* 9, 1–13
35. Choi, Y. et al. (2018) Autophagy during viral infection – a double-edged sword. *Nat. Rev. Microbiol.* 16, 341–354
36. Mitchell, A.L. et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360
37. Puntervoll, P. et al. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* 31, 3625–3630
38. Gibson, T.J. et al. (2015) Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun. Signal.* 13, 42
39. Rojas, M. et al. (2015) An eIF2 $\alpha$ -binding motif in protein phosphatase 1 subunit GADD34 and its viral orthologs is required to promote dephosphorylation of eIF2 $\alpha$ . *Proc. Natl. Acad. Sci. U. S. A.* 112, E3466–E3475
40. Edwards, R.J. and Palopoli, N. (2015) Computational prediction of short linear motifs from protein sequences. In *Computational Peptidology (Methods in Molecular Biology)* (Zhou, P. and Huang, J. eds), pp. 89–141, Springer
41. Bailey, T.L. et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208
42. Davey, N.E. et al. (2012) SLIMPrints: conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. *Nucleic Acids Res.* 40, 10628–10641
43. Davey, N.E. et al. (2006) SLIMDisc: short, linear motif discovery, correcting for common evolutionary descent. *Nucleic Acids Res.* 34, 3546–3554
44. Edwards, R.J. et al. (2007) SLIMFinder: a probabilistic method for identifying over-represented, convergently evolved, short linear motifs in proteins. *PLoS One* 2, e967
45. Davey, N.E. et al. (2010) SLIMFinder: a web server to find novel, significantly over-represented, short protein motifs. *Nucleic Acids Res.* 38, W534–W539
46. Neduva, V. et al. (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.* 3, e405
47. Neduva, V. and Russell, R.B. (2006) DILIMOT: discovery of linear motifs in proteins. *Nucleic Acids Res.* 34, W350–W355
48. Frith, M.C. et al. (2008) Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput. Biol.* 4, e1000071

49. Down, T.A. and Hubbard, T.J. (2005) NestedMICA: sensitive inference of over-represented motifs in nucleic acid sequence. *Nucleic Acids Res.* 33, 1445–1453
50. Doğruel, M. et al. (2008) NestedMICA as an *ab initio* protein motif discovery tool. *BMC Bioinformatics* 9, 19
51. Letunic, I. et al. (2014) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260
52. Finn, R.D. et al. (2016) InterPro in 2017–beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199
53. Linding, R. et al. (2003) GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31, 3701–3708
54. Dosztányi, Z. et al. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21, 3433–3434
55. Dinkel, H. and Sticht, H. (2007) A computational strategy for the prediction of functional linear peptide motifs in proteins. *Bioinformatics* 23, 3297–3303
56. Sobhy, H. (2017) A bioinformatics pipeline to search functional motifs within whole-proteome data: a case study of poxviruses. *Virus Genes* 53, 173–178
57. Becerra, A. et al. (2017) Prediction of virus–host protein–protein interactions mediated by short linear motifs. *BMC Bioinformatics* 18, 163
58. Hagai, T. et al. (2011) Intrinsic disorder in ubiquitination substrates. *J. Mol. Biol.* 412, 319–324
59. Jacomin, A.-C. et al. (2017) iLIR@viral: a web resource for LIR motif-containing proteins in viruses. *Autophagy* 13, 1782–1789
60. Beale, R. et al. (2014) A LC3-interacting motif in the influenza A virus M2 protein is required to subvert autophagy and maintain virion stability. *Cell Host Microbe* 15, 239–247
61. Blikstad, C. and Ivarsson, Y. (2015) High-throughput methods for identification of protein–protein interactions involving short linear motifs. *Cell Commun. Signal.* 13, 38
62. Hulo, C. et al. (2010) ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* 39, D576–D582
63. Masson, P. et al. (2012) ViralZone: recent updates to the virus knowledge resource. *Nucleic Acids Res.* 41, D579–D583
64. Masson, P. et al. (2014) An integrated ontology resource to explore and study host–virus relationships. *PLoS One* 9, e108075
65. Foulger, R. et al. (2015) Representing virus–host interactions and other multi-organism processes in the Gene Ontology. *BMC Microbiol.* 15, 146
66. Hulo, C. et al. (2017) The ins and outs of eukaryotic viruses: knowledge base and ontology of a viral infection. *PLoS One* 12, e0171746
67. The UniProt Consortium. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169
68. McDonald, A.G. and Tipton, K.F. (2014) Fifty-five years of enzyme classification: advances and difficulties. *FEBS J.* 281, 583–592
69. Dinkel, H. et al. (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* 42, D259–D266
70. Dinkel, H. et al. (2016) ELM 2016–data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.* 44, D294–D300
71. Dinkel, H. et al. (2011) Phospho.ELM: a database of phosphorylation sites – update 2011. *Nucleic Acids Res.* 39, D261–D267
72. Van Roey, K. et al. (2013) The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal.* 6, rs7
73. Weatheritt, R.J. et al. (2012) iELM – a web server to explore short linear motif-mediated interactions. *Nucleic Acids Res.* 40, W364–W369
74. Gouw, M. et al. (2017) Exploring short linear motifs using the ELM database and tools. *Curr. Protoc. Bioinformatics* 58, 8.22.1–8.22.35
75. Verpoeten, D. et al. (2009) Control of TANK-binding kinase 1-mediated signaling by the  $\gamma_134.5$  protein of herpes simplex virus 1. *J. Biol. Chem.* 284, 1097–1105
76. Orvedahl, A. et al. (2007) HSV-1 ICP34.5 confers neurovirulence by targeting the Beclin 1 autophagy protein. *Cell Host Microbe* 1, 23–35
77. Zhang, C. et al. (2008) A conserved domain of herpes simplex virus ICP34.5 regulates protein phosphatase complex in mammalian cells. *FEBS Lett.* 582, 171–176
78. Korom, M. et al. (2014) Up to four distinct polypeptides are produced from the  $\gamma_34.5$  open reading frame of herpes simplex virus 2. *J. Virol.* 88, 11284–11296
79. Davis, K.L. et al. (2014) Herpes simplex virus 2 ICP34.5 confers neurovirulence by regulating the type I interferon response. *Virology* 468–470, 330–339
80. Mao, H. and Rosenthal, K.S. (2002) An N-terminal arginine-rich cluster and a proline-alanine-threonine repeat region determine the cellular localization of the herpes simplex virus type 1 ICP34.5 protein and its ligand, protein phosphatase 1. *J. Biol. Chem.* 277, 11423–11431
81. Jing, X. et al. (2004) Replication of herpes simplex virus 1 depends on the  $\gamma_134.5$  functions that facilitate virus response to interferon and egress in the different stages of productive infection. *J. Virol.* 78, 7653–7666
82. Wilcox, D.R. and Longnecker, R. (2016) The herpes simplex virus neurovirulence factor  $\gamma_34.5$ : revealing virus–host interactions. *PLoS Pathog.* 12, e1005449
83. Gould, C.M. et al. (2010) ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.* 38, D167–D180