Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Convergent evolution and mimicry of protein linear motifs in host–pathogen interactions

Lucía Beatriz Chemes[1], Gonzalo de Prat-Gay[1] and Ignacio Enrique Sánchez[2]

Pathogen linear motif mimics are highly evolvable elements that facilitate rewiring of host protein interaction networks. Host linear motifs and pathogen mimics differ in sequence, leading to thermodynamic and structural differences in the resulting protein–protein interactions. Moreover, the functional output of a mimic depends on the motif and domain repertoire of the pathogen protein. Regulatory evolution mediated by linear motifs can be understood by measuring evolutionary rates, quantifying positive and negative selection and performing phylogenetic reconstructions of linear motif natural history. Convergent evolution of linear motif mimics is widespread among unrelated proteins from viral, prokaryotic and eukaryotic pathogens and can also take place within individual protein phylogenies. Statistics, biochemistry and laboratory models of infection link pathogen linear motifs to phenotypic traits such as tropism, virulence and oncogenicity. *In vitro* evolution experiments and analysis of natural sequences suggest that changes in linear motif composition underlie pathogen adaptation to a changing environment.

Addresses
[1] Protein Structure-Function and Engineering Laboratory, Fundación Instituto Leloir and IIBBA-CONICET, Av. Patricias Argentinas 435, 1405 Buenos Aires, Argentina
[2] Protein Physiology Laboratory, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales and IQUIBICEN-CONICET, Universidad de Buenos Aires, Buenos Aires, Argentina

Corresponding authors: Chemes, Lucía Beatriz (lchemes@leloir.org.ar) and Sánchez, Ignacio Enrique (isanchez@qb.fcen.uba.ar)

Current Opinion in Structural Biology 2015, **32**:91–101

This review comes from a themed issue on **Sequences and topology**

Edited by **Anna Panchenko** and **Madan Babu**

http://dx.doi.org/10.1016/j.sbi.2015.03.004

Introduction

Linear motifs mediate a wide range of processes such as protein cleavage, protein degradation, post-translational modification, sub-cellular localization or binding to modular protein domains [1,2•] (Box 1). Linear motifs can confer functional diversity to splice variants [2•] and may act in coordination, forming higher order functional units or 'switches' [1,2•,3] (Box 1). Given their ubiquity, it is not surprising that aberrant function of linear motifs is linked to both genetic [2•] and infectious diseases [2•,4,5]. In the case of infectious diseases, it is widely recognized that many viral proteins harbor functional linear motifs that resemble those found in the host [4,5]. This special type of mimicry that takes place in host–pathogen interactions can be called linear motif mimicry (Box 1), and the individual examples may be called mimics. Beyond viruses, recent reviews describe a number of linear motif mimics in a diverse set of prokaryotic and eukaryotic pathogens [2•,6–9]. Pathogen mimics are widely distributed across the tree of life, including proteobacteria, spirochaetes, firmicutes, fungi, chromalveolata, excavata and animalia [2•,9–11]. Similar to viruses, these linear motif mimics have contributed to the discovery of new host linear motifs [12] and mediate a wide range of processes, such as adhesion to target cells, virulence protein secretion, perturbation of cell signaling, and inhibition of host defense systems [2•,9]. Despite the growing recognition of their relevance, we are only beginning to understand the evolutionary patterns and comparative properties of host motifs and their pathogen mimics. In the following, we briefly review the major advances in these areas.

Linear motif mimicry from a sequence viewpoint

Linear motif mimicry takes place when a pathogen protein harbors a linear motif instance also present in the host. However, we are only beginning to grasp how far this resemblance holds, as illustrated at different scales in Figures 1 and 2 [7]. We may first consider sequence mimicry in the linear motif and its vicinity. Different instances of a linear motif are often not homogeneous in sequence, while still in agreement with the regular expression (Figure 1a, top) (Box 1). For example, host and viral instances of the [LI]xCx[DE] retinoblastoma-binding motif may differ in the fixed positions at the beginning and end of the regular expression, in the wildcard positions interspersed in the regular expression and in the immediately adjacent positions (see also [1,13]) (Box 1). Taken as two separate groups, host and virus instances of a linear motif may present subtly different amino acid preferences, as illustrated by the sequence logos in Figure 1b. On a broader sequence context, the number of motif copies in a certain host or pathogen protein [14,15] and the positions flanking the motif [1,2•] may

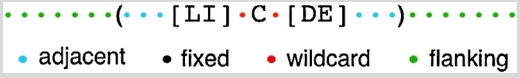
Box 1 Linear motif mimicry

Linear motif. Protein subsequence that modulates protein function. Typical linear motifs are shorter than 15 residues and embedded in intrinsically disordered regions.

Motif switch. Gain, loss or exchange of linear motif binding partners that leads to integrated regulation of protein interaction interfaces and function. This process often takes place through splicing or post-translational modificationof linear motifs.

Linear motif mimicry. A particular kind of molecular mimicry in which a pathogen harbors a linear motif resembling a linear motif found in the host. In most cases, the mimic functions through binding to the same binding site as the host linear motif, and gives the pathogen access to the motif-mediated systems of the host.

Regular expression. Structured sequence pattern that lists the allowed and forbidden amino acids at each position of a linear motif. Within regular expressions, we can define fixed and wildcard positions, and within the linear motif vicinity we can define adjacent and flanking positions.



Motif repertoire. Specific combination of linear motifs found in a protein.

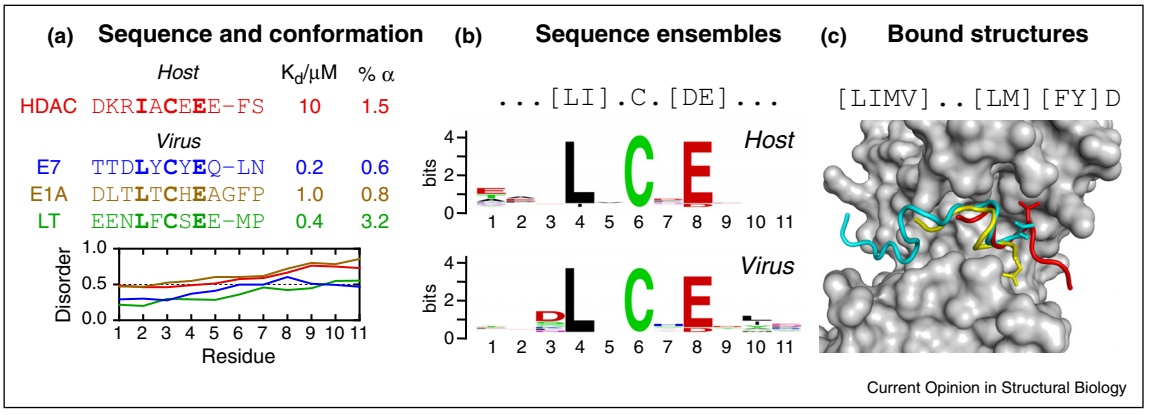
Host protein hijacking. Interaction between a host and pathogen protein that alters function of the host protein and leads to an increase in pathogen fitness.

also vary across instances. Finally, we should also consider that, as a consequence of convergent evolution of linear motifs (Box 2), the remainder of the pathogen protein usually does not present measurable sequence similarity to the host protein harboring the mimicked motif [16•]. Thus, the same linear motif may be associated with very different repertoires of other linear motifs and domains in each host and pathogen protein [17,18] (Figure 2a) (Box 1). To sum up, the sequence context of host linear motifs and their pathogen mimics can differ widely both at the local and whole-protein scales.

Linear motif mimicry from a physicochemical viewpoint

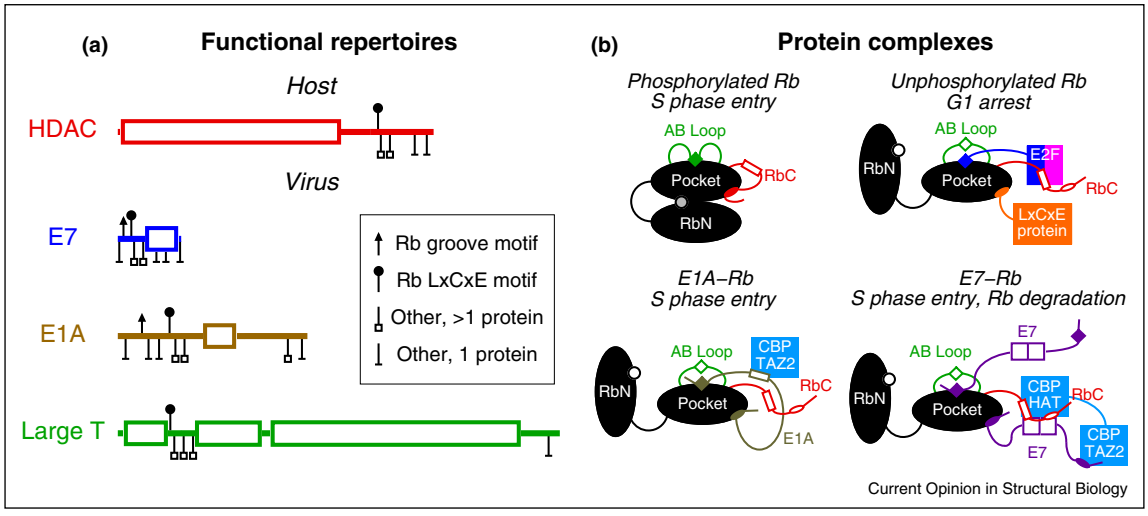
Differences in sequence between linear motif instances and their pathogen mimics has consequences at many levels. From a physicochemical viewpoint, it is useful to distinguish between the unbound and bound states of a linear motif. The unbound state is usually disordered, with low populations of conformations with local secondary structure. On the other hand, the bound state is usually well ordered and forms short-range interactions with a globular domain. Differences in sequence across linear motif instances can lead to different structures in the bound state. Figure 1c shows two host motifs and a pathogen mimic of the retinoblastoma-binding AB groove motif in complex with their target domain [19,20]. Although all three instances bind forming a helical conformation, there are observable differences in the flanking residues, in main chain conformation, and in the orientation of side chains corresponding to fixed positions in the regular expression. Compliance with a regular expression is also compatible with a range of structural propensities in the unbound state, such as local disorder (Figure 1a, bottom) and local structural propensities for alpha helix, beta sheet and/or polyproline type II helix conformations [21,22••,23]. Also, the unbound state of the motif can have structural propensities similar or different from the bound form [21,22••,23]. The structural context of a motif

Figure 1



Molecular mimicry of retinoblastoma-binding linear motifs in host–pathogen interactions. **(a)** Sequence and conformation. Differences in sequence [1], binding affinity [17,24,26], disorder propensity (IUPRED; URL: http://iupred.enzim.hu/, bottom) and average alpha helix propensity (AGADIR; URL: http://agadir.crg.es/), for one host (HDAC, red) and three viral instances (human papillomavirus type 16 E7, blue; adenovirus type 5 E1A, brown; SV40 polyomavirus Large T antigen, green) of the retinoblastoma-binding [LI]xCx[DE] motif. **(b)** Sequence ensembles. Regular expression for the retinoblastoma-binding [LI]xCx[DE] motif [1] (Box 1) and sequence logos (WEBLOGO; URL: http://weblogo.berkeley.edu/) for host and viral instances. Positions 3 and 10 are responsible for the higher affinity of viral instances [24]. **(c)** Structures in the bound state. Regular expression of the retinoblastoma-binding AB groove motif and structure of the E2F1 (host, cyan), RBloop (host, red) and E1A (viral, yellow) instances in complex with the retinoblastoma protein (gray surface) [1,20,31•]. Side-chains of the last motif-determining residue are also shown.

Figure 2



Diverse functional outputs of mimic repertoires in pathogen proteins targeting the host retinoblastoma protein. **(a)** Functional repertoires. Architecture of HDAC (host, red), papillomavirus E7 (blue), adenovirus E1A (brown) and polyomavirus Large T antigen (green). Boxes indicate globular domains, while lines indicate intrinsically disordered regions [1]. Flags indicate experimentally validated linear motifs [1,44,62••]. Circles: retinoblastoma-binding [LI]xCx[DE] motif. Triangles: retinoblastoma-binding AB groove motif. Squares: other motifs, present in more than one of these proteins. Perpendicular lines: other motifs, present in only one of these proteins. **(b)** Protein complexes. Large shapes: globular domains. Smaller shapes: linear motifs. Lines: intrinsically disordered regions. Top: retinoblastoma protein conformational substates in the host cell. Bottom: hijacking of the retinoblastoma protein by the adenovirus E1A and papillomavirus E7 proteins.

instance will be even more protein-dependent in the case of those motifs located in the loops of globular domains [2•]. Differences in sequence and structure are often coupled to differences in binding affinity and kinetics (Figure 1a, top). The changes in affinity can be traced to motif-determining residues [24], to wildcard, adjacent and flanking residues [17,24–26], to the unbound structure [22••] and/or to an increased number of copies of the motif [14,15]. The differences in affinity can lead to a binding hierarchy in which strong-binding proteins displace weaker-binding proteins. It has been proposed that pathogen linear motif mimics effectively compete with host proteins by means of tighter binding [4] (Figures 1a and 2b). In several cases, the pathogen linear motif mimics indeed showed a higher affinity that could be traced to the differences in sequence [15,24,27•] (Figure 1b). This is made possible by the fact that host interactions are not optimized for high binding affinity, allowing for the appearance of 'superbinders' that may have a physiological effect on the network [28]. Local protein concentration in the cell may affect binding as much as the measured differences in affinity [29]. This issue, which may affect competition between host linear motifs and their pathogen mimics, is still poorly understood. Therefore, current evidence shows that there are subtle yet significant differences in conformation and energetic properties between host linear motifs and their pathogen mimics.

Diverse functional outputs of mimic repertoires in pathogen proteins

Host–pathogen protein–protein interactions are usually multivalent, with several binding surfaces mediated by different linear motif mimics and/or globular domains (Figure 2a). This leads to multiprotein assemblies with complex architectures and distinct conformational substates (Figure 2b, top) [1,2•,30,31•,32–36], in which the role of a motif mimic in hijacking the host protein is highly pathogen dependent (Figure 2b, bottom) (Box 1). For example, the closed substate of the retinoblastoma cell cycle regulator is stabilized by phosphorylation events at the RbN domain (gray circle), the AB loop of the pocket domain (green diamond) and the RbC domain (red ellipse). The open substate is stabilized by multivalent binding of E2F/DP dimers and of proteins containing the [LI]xCx[DE] motif [31•]. Figure 2b bottom shows hijacking of the retinoblastoma protein by the adenovirus E1A and papillomavirus E7 proteins, which contain a similar set of mimics (Figure 2a). The monomeric E1A protein displaces the host targets of retinoblastoma by mimicking the [LI]xCx[DE] and AB groove motifs and recruits the CBP protein [30]. This ternary complex formed between the host CBP and retinoblastoma proteins and the adenovirus E1A protein shows either positive or negative cooperativity, depending on the available E1A interaction sites [37•]. On the other hand, the dimeric E7 protein displaces the host targets of

Box 2 Linear motif evolution

Evolutionary innovation. A specific feature that arises throughout evolution, and endows an organism with a qualitatively new ability.

Regulatory evolution. Evolution of regulatory networks that takes place through changes in network connectivity.

Convergent evolution of linear motifs. Multiple independent appearances of a given motif that can occur both in unrelated proteins and in separate branches of a protein phylogeny.

dN/dS. Ratio of nonsynonymous versus synonymous substitutions at a given codon or gene region. Values of dN/dS > 1 are indicative of positive selection, and values of dN/dS < 1 are indicative of negative selection.

Positive selection. Evolutionary process that leads to an increase in the prevalence of traits that provide a selective advantage or increase fitness.

Negative-purifying selection. Evolutionary process that leads to a decrease in the prevalence of traits that diminish fitness or contributes to the maintenance of an existing trait.

Adaptive evolution. Episode of natural selection through which specific phenotypes that increase fitness become fixed and maintained in a population.

Phylogenetic reconstruction of linear motif history. Bioinformatic mapping of the evolutionary history of a linear motif within one or more protein families. Reconstructions are performed using a hypothesis on the protein phylogeny, together with a model for motif evolution.

Homologous proteins. Proteins that are related by common ancestry. Homologs found in different species are commonly referred to as orthologs, while those that arise after gene duplication are referred to as paralogs.

retinoblastoma by mimicking the [LI]xCx[DE] and AB groove motifs and binding to the RbC domain. E7 also recruits the CBP protein using the region harboring the [LI]xCx[DE] motif [33••]. However, while displacement of the E2F transcription factors by both viral proteins induces S-phase reentry, pathogen-specific effects include E1A-induced acetylation and E7 mediated degradation of the retinoblastoma protein [38]. This evidence implies that the output of an interaction mediated by a pathogen linear motif mimic depends crucially on the remainder of the pathogen protein. Therefore, molecular mimicry of a host linear motif by different pathogens does not necessarily lead to similar functional consequences [7].

The design of drugs against motif-mediated host pathogen interactions is complicated by mimicry, because inhibition of interactions key to the pathogen life cycle may also interfere with host protein–protein interactions [5]. However, the motif repertoires of host and pathogen proteins may be different enough to enable specific targeting, by using compounds that target multiple motifs in a pathogen protein. A recently designed proof-of-concept engineered protein uses this rationale, targeting a unique combination of linear motif mimics found in the HIV Nef protein [39]. The designed decoy protein effectively interferes with Nef function, suggesting that it successfully wraps the viral linear motif mimics and renders them inactive [39].

Linear motifs as key elements in regulatory evolution

Evolution is a complex process that takes place through the rising of '*evolutionary innovations*' [40], which can involve changes in molecule architecture, or in the regulatory networks connecting the molecules themselves [40] (Box 2). The latter process is known as *regulatory evolution* and is a prevalent evolutionary process, first identified in transcription factor binding sites [40,41] (Box 2). One striking example is predator evasion, where increases in expression levels of the Distal-Less transcription factor allow the formation of butterfly eyespots [40].

Strong parallels can be drawn between transcriptional *regulatory evolution* and the evolution of post-translational modification and protein–protein interaction networks, governed by the presence of highly evolvable short linear motifs [2•,41]. In agreement with this, computational analysis of interaction networks shows that interactions mediated by linear motifs are the most likely to be rewired [42]. The nature and structural context of linear motifs makes them strong candidates for rapidly evolvable elements. Globular domains evolve slowly and divergently through domain duplication or swapping [43]. In contrast, while disordered regions may face evolutionary constraints to preserve structural ensembles and binding sites [44,45], they are able to evolve quickly as they are not restrained to maintain a rigid structural fold [45]. The small size and degenerate nature of linear motifs implies that, similar to transcription factor binding sites, the genotypic changes required for their appearance or loss are small [2•,43], making their frequent appearance plausible. Following emergence, linear motif sequences that increase fitness can be positively selected and become fixed in the population [2•,43,46,47], or negatively selected in order to suppress deleterious interactions [48,49] (Box 2). This process can occur independently in unrelated sequences, leading to convergent evolution [2•] (Box 2).

Evolution of host linear motifs: The old and the new

Although it has been long proposed that linear motifs evolve quickly and convergently, and are subject to adaptive evolution [43] (Box 2), until recently these tenets were supported mainly on theoretical grounds. Pioneering studies were focused on evidence scoring the presence or absence of individual motifs or their occurrence in unrelated proteins. Initial analyses of the eukaryotic linear motif (ELM) database indicated that ~50% of motifs occur in at least two unrelated proteins [50•]. Several motifs showed more than two convergent

instances, with many examples presenting instances in over 10 unrelated proteins [50•]. A proteome-wide study of N-glycosylation sites across eukaryote species showed that many modification sites are shared by proteins that arose after the phylogenetic divergence between the corresponding lineages [51]. Finally, several examples document the acquisition of motifs after duplication in paralogs [34–36] (Box 2). Although these studies did not provide a direct proof, they also suggested convergent evolution of motifs as the most likely explanation.

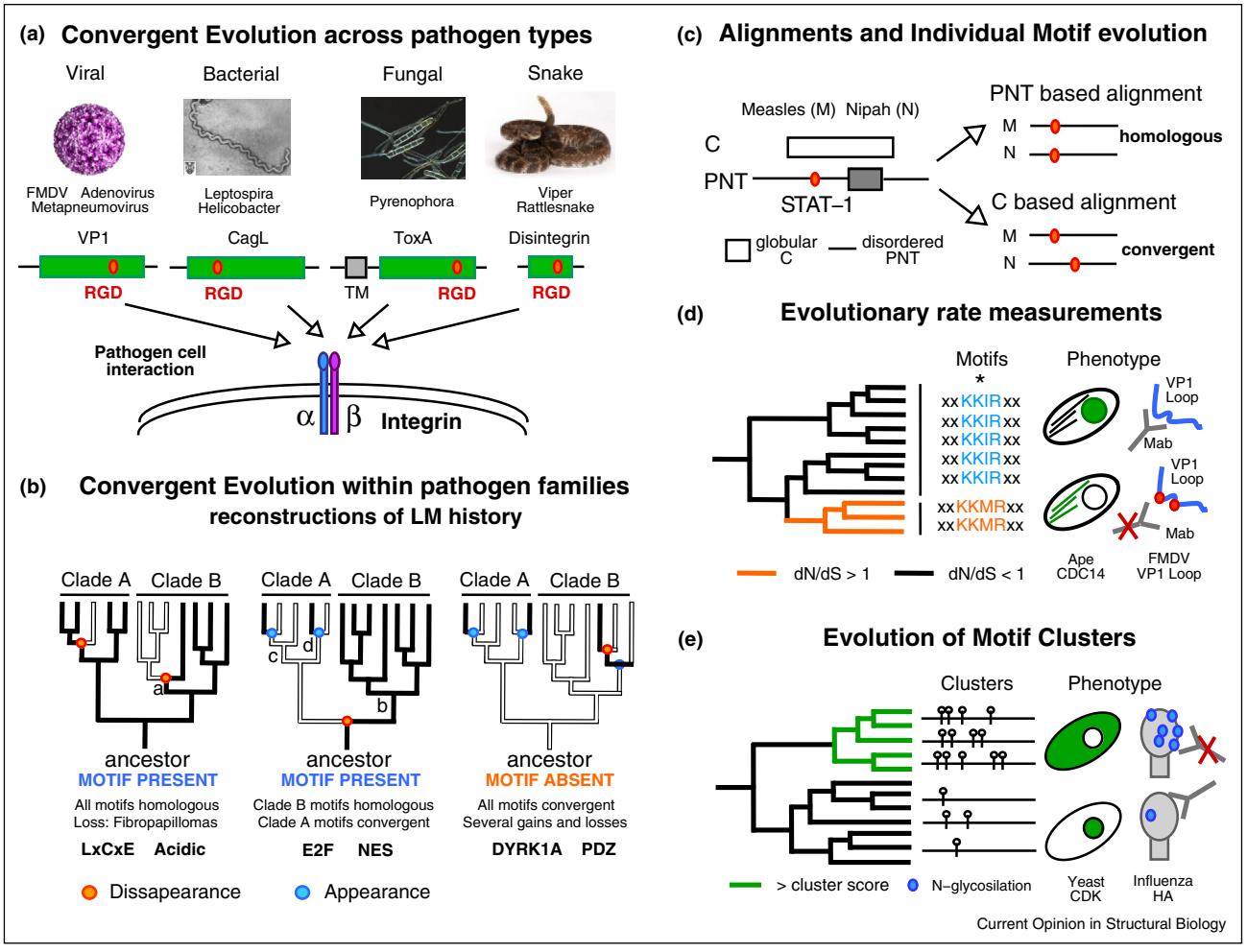
Additional evidence for linear motif convergence and plasticity can be provided by explicit modeling of linear motif evolution. For example, phylogenetic reconstructions of linear motif history can be used to test for convergence, while measurements of synonymous and nonsynonymous substitution rates allow for direct testing of positive and negative selection events and their relationship to adaptive evolution (Box 2). In the following, we review recent advances on host linear motif evolution, some of which are shown in Figure 3. Several studies have measured substitution rates in host linear motifs. Genome-wide measurements of evolutionary rate of yeast linear motifs have shown that linear motifs in paralogs undergo periods of accelerated evolution followed by functional divergence [52••]. Another study of paralogs of the hominid CDC14 protein identified short evolutionary periods of strong positive selection followed by purifying selection in a nuclear localization motif, leading to a phenotypic change from nuclear to microtubular association [47] (Figure 3d). Evolutionary analyses of extant and reconstructed sequences in the insect ENGRAILED selector protein showed the emergence of a specific GRO-interaction motif in the dipteran/lepidopteran lineage, which was maintained under strong purifying selection [53]. Host linear motifs may also evolve in response to pathogens. For example, it has been observed that the viral-sensing MAVS protein from primates ceased being a target for a viral protease upon mutation of the cleavage motif during an episode of positive selection [46]. However, the number of escape mutations may be limited by the need to maintain host interactions and functions [7]. Linear motifs may also undergo negative selection [49], which could function to suppress deleterious interactions [48]. Therefore, current evidence supports the notion that linear motifs are highly evolvable regulatory elements that play important roles in functional divergence.

Cumulative evidence on the evolution of linear motifs shows that several modes of conservation are possible. Some examples resemble the evolution of enzyme active sites, where a single copy of the sequence pattern is conserved across long evolutionary distances [54]. However, only 5% of instances in the ELM database are conserved from mammals through yeast [50•], and some motifs are found only in recent lineages [50•,53]. This implies that conservation of linear motifs should be studied together with the phylogenetic distribution of the motif. In other cases, the conserved feature is a motif cluster, defined as the presence of multiple copies of a motif, each of which can be found at variable positions of a defined protein region. For example, phosphorylation sites in yeast show strong conservation of motif clusters, with high evolutionary rates and turnover of individual motifs [41,55] (Figure 3e). Changes in motif clusters in specific lineages correlate with protein localization, suggestive of a functional role in diversification [41]. In a striking case, understanding motif conservation patterns may require knowledge of network evolution. In the calcineurin–kinase network, the conserved feature is the presence of a dual docking site on substrates of a specific kinase–phosphatase pair. However, intensive rewiring of the calcineurin network across evolution, leads to poor conservation of the docking motif in individual substrates, while maintaining kinase–phosphatase specificity [56•]. Therefore, studies of linear motif evolution require understanding of conservation patterns together with phylogeny and network connectivity.

Evolution of linear motifs in pathogens

Several properties of host–pathogen interactions suggest that they constitute a specific case of regulatory evolution. The interaction between a pathogen and its host leads to the establishment of a protein–protein interaction network in which pathogen proteins preferentially interact with highly connected host proteins (hubs) or host proteins that connect functional modules (bottlenecks) [6,16•], causing an extensive amount of rewiring. Several lines of evidence indicate that interactions of pathogen proteins are based to a large extent on linear motif mimics. First, different viruses often target the same host proteins, indicative of similar mechanisms of interaction [4–6,16•,18]. Second, evolutionary constraints on genome size lead to a high density of known [4,16•] and putative [57••] linear motifs. Third, the presence of linear motifs correlates with the preferential targeting of human proteins containing linear motif-binding domains [16•]. Moreover, the prevalence of disordered regions in viral proteins [58] and the fast genome mutation and evolutionary rates [59] are thought to endow viral linear motifs with increased evolutionary plasticity relative to host motifs [60]. Scoring of non-synonymous mutations in modeled populations of HIV-1 genome variants showed conservation of the original motif repertoire in 50% of sequences, while new motif variants were generated [57••], suggesting that linear motifs are robust yet highly evolvable elements. Although the evolutionary patterns of pathogen linear motifs are only beginning to be characterized, most studies suggest that viral linear motifs allow rewiring of virus–host interactions [61] and participate in viral adaptation.

Figure 3



Evolution of host and pathogen linear motifs. **(a)** Convergent evolution across pathogen types. Viral, bacterial, fungal and animal pathogens share proteins containing the RGD linear motif, which mediates interactions with α/β integrin receptors in the host cell [2•,4,9,10]. **(b)** Convergent evolution within a viral family. Linear motif history reconstructions based on phylogenetic analysis reveals different evolutionary patterns. Left panel: All instances of a conserved linear motif are homologous, and a single appearance event occurred, in this case exemplified in the ancestor protein. The Rb-binding LxCxE and Acidic motifs in the papillomavirus (PV) E7 protein present this behavior [62••]. Middle panel: Loss and re-emergence of a linear motif present in deep phylogenetic branches can occur, giving rise to instances that are homologous (i.e. all instances within Clade B) and to instances that are convergent (i.e. instances in Clade A compared to all others). The E2F-mimic motif and NES nuclear localization signal in E7 present this behavior [62••]. Right panel: Motifs can emerge several times in recent branches of the phylogeny, giving rise to motifs that are all convergent, and to a low overall prevalence and conservation of the motif. The DYRK1A phosphorylation site and the PDZ motifs in E7 follow this behavior [62••]. **(c)** Alignments and individual motif evolution. The paramyxovirus Measles and Nipah PNT regions present a STAT-1 binding motif (red sphere) [63••]. Alignment based on the disordered PNT region predicts conserved positioning of the STAT-1 motif, suggesting homologous descent. However, this region is difficult to align due to low general sequence conservation. Conversely, alignment based on the overlapping highly conserved globular C protein region and re-translation to the PNT coding frame, reveals different locations for each STAT-1 motif, indicating independent origins [63••]. **(d)** Measurements of linear motif evolutionary rates. Phylogenetic analysis can reveal periods of positive selection (dN/dS > 1) (Box 2), leading to changes in linear motif composition. This can give rise to changes in protein localization (from nuclear to microtubular) as for the Ape ancestor CDC14retro paralog [47], or to the development of antibody resistance, as seen for the antigenic loop region of the FMDV VP1 capsid protein [76]. **(e)** Evolution of motif clusters. Clusters, and not individual motif positions may be conserved. The expansion of motif clusters can give rise to changes in protein localization, as seen with CDK sites in the yeast mcm3 protein [41], or to phenotypic changes such as the development of immunotolerance in Influenza strains by acquisition of N-glycosylation sites in the hemagglutinin (HA) antigenic protein [68,69].

Convergent evolution of pathogen linear motifs

Increasing evidence points to the prevalence of convergent evolution of pathogen linear motifs. Scoring of linear motif occurrences in viral proteomes shows that ~30% of motifs in the ELM database can be found in proteins from unrelated viruses [4], with several instances such as the

[LI]xCx[DE] and PDZ motifs, showing over 10 convergent occurrences [4,18]. This is considered to be a lower estimate [4], since many un-annotated instances remain to be identified [57••]. Remarkably, many motifs have been convergently evolved across pathogens groups, including the integrin binding RGD motif and the 14-3-3 motifs present in viral, prokaryotic and eukaryotic pathogens [2•,4,9,10] (Figure 3a). The growing description of motif mimics in non-viral pathogens suggests convergent evolution of motif mimicry is pervasive.

As a consequence of the high evolutionary plasticity of linear motifs, two copies of the same motif within homologous pathogen proteins may also be the result of convergent evolution (Box 2). A recent study performed a comprehensive phylogenetic reconstruction of linear motif history in 217 sequences from the papillomavirus E7 oncoprotein [62••]. Papillomaviruses co-evolved with amniote hosts, providing with a well-established phylogeny covering ~350 million years. High variability was observed in linear motif evolutionary behavior (Figure 3b). Some highly conserved motifs appeared only once in papillomavirus evolution, with all instances being homologous [62••] (Figure 3b, left panel). However, four linear motifs within E7 showed multiple independent appearance events in deep and recent branches (Figure 3b, middle and right panels), providing direct evidence for convergent evolution within a viral phylogeny [62••].

Distinguishing convergently evolved from homologous instances within viral families requires careful construction of viral phylogenies and alignments. In a striking example, analysis of the disordered Paramyxovirus P protein PNT region based on an alignment of the overlapping structured C protein ORF showed that the Nipah and Measles STAT-1 motifs, considered to be homologous, have appeared independently two times [63••] (Figure 3c). Advanced alignment techniques performed on the disordered Paramyxovirus P protein N terminal region also helped determine that the soyuz1 motif present throughout the *Paramyxovirinae* sub-family probably evolved by homologous descent [64]. Therefore, while careful studies are needed, current evidence indicates that convergent evolution of pathogen linear motifs operates across pathogen types, as well as within individual phylogenies.

Pathogen linear motifs and adaptive evolution

A growing body of evidence indicates that pathogen linear motifs can be viewed as examples of *evolutionary innovations* [40]. Linear motif repertoire has been linked to changes in viral phenotypic traits such as virulence [65–67], persistence [68,69,70••] tropism [62••], oncogenicity [26,71,72], and response to therapy [73]. The association may be inferred from purely statistical correlation between motif repertoire and phenotypic traits [62••,67,68,73], such as the correspondence between the motif repertoire in the HIV proteome and response of patients to antiretroviral therapy [73]. The association may be further strengthened by experimental demonstration of biochemical properties modulated by the motif [26,69,71,72]. For example, specific substitutions in E7 proteins of high-risk oncogenic types, compared to low-risk types, lead to 30-fold increases in the thermodynamic and kinetic stability of the complexes formed between E7 and the retinoblastoma cell cycle regulator [26]. If possible, the association between motifs and phenotypes should also be studied using laboratory models of infection [65,66]. An interesting case is the influenza NS1 protein, where grafting a PDZ binding motif from the highly pathogenic H1N1 and H5N1 strains to the wild type virus results in decreased survival in a mice model system [65].

Emergence and fixation of pathogen linear motifs may be studied by measuring nonsynonymous (dN) and synonymous (dS) substitution rates in naturally occurring sequences. A positive selection event of motif emergence is characterized by a high dN/dS ratio. Few such studies have been reported to date. Analysis of extant genotypes of the hepatitis E virus genome reported high nonsynonymous to synonymous ratios for 4–10 codons in a motif-rich polyproline region [74•]. Genotypes with different host range show varying degrees of sequence divergence, suggesting that these events may be adaptive. Conversely, fixation of the motif by purifying selection is associated with a low dN/dS ratio, as observed for linear motifs in the hepatitis C virus core protein [75].

Adaptive evolution events may be recreated in the laboratory by *in vitro* evolution techniques with rapidly evolving viruses. This allows a greater experimental control of variables such as selection pressure. An interesting model is the Foot and Mouth Disease Virus (FMDV) capsid VP1 protein, which contains a variable loop harboring both an antigenic determinant and an integrin-binding RGD motif. *In vitro* selection of viruses able to replicate in media containing anti-VP1 antibodies resulted in the accumulation of mutations in the antigenic VP1 region. Analysis of this region showed that the mutated positions presented high dN/dS values, indicating positive selection [76] (Figure 3d). Instead, when the same viruses were selected for their ability to replicate in the presence of soluble integrin decoys, mutations were identified in the integrin-binding RGD motif [70••]. Another means of studying adaptive changes in viral linear motifs is through mutation experiments. In this case, reversion of inactivating mutations can be observed during *in vitro* evolution, as shown for the Hepatitis C virus NS5B retinoblastoma-binding [LI]xCx[DE] motifs [77]. In sum, studies of pathogen linear motif evolution can be linked to phenotype using different experimental and computational approaches. Future studies combining multiple techniques may

reveal to what extent linear motif evolution contributes to pathogen adaptation.

Coevolution of pathogen linear motifs

Pathogen linear motifs may function in a coordinate manner, forming higher-order functional units or switches [3]. This phenomenon may be detected through co-occurrence and coevolution patterns in single or in multiple protein families and tested in biochemical and functional studies. An interesting example regarding single protein families is the papillomavirus E7 protein, where the motif pairs LxCxE-CKII and CKII-Acidic co-occur significantly more often than expected [62••] and two sequence positions of the LxCxE-CKII pair show a coevolution signal [44]. Moreover, the CKII and acidic motifs are functionally coupled to the LxCxE motif to increase Rb binding affinity [26] and the absence of the LxCxE-Acidic pair is associated to a change in tissue tropism [62••]. The LxCxE-CKII-Acidic module is present in the E7 protein ancestor and conserved by purifying selection (Figure 3b). In the antigenic region of Influenza HA antigens, glycosylation motifs are found in multiple copies, with copy number correlating with antigenicity [68,69] (Figure 3e). Regarding unrelated viral families, large scale studies of many viral proteomes found over 200 viral linear motif pairs that occur more often than expected and in multiple, unrelated viral families [57••], suggesting that in this case motif switches have evolved convergently.

Host defense mechanisms can constrain linear motif mimicry

The response of the host to infection can strongly influence the evolution of pathogen linear motifs. The adaptive immune system relies on the recognition of conserved pathogen immunogenic sequences [78]. For example, it has been shown that linear motif mimics in the HIV Nef protein are preferential targets for the adaptive immune response [79], suggesting that many immunogenic epitopes may correspond to linear motif mimics. Viruses often evade the adaptive immune system by depleting immunogenic sequence motifs from their proteome [80]. This need for immune evasion may limit the use of linear motif mimicry by pathogens, although in some cases evasion may be compatible with the conservation of linear motif mimics [76].

Technical challenges

Evolutionary analyses of viral linear motifs may be hampered by lack of knowledge of phylogenetic relationships between and within viral families [57••], by poor sampling of viral phylogenies, and by difficulties in producing reliable alignments of disordered regions [63••,64]. Moreover, many motifs can appear in different numbers or change their location in the sequence while preserving functionality, making their conservation and evolutionary rate hard to score by classical alignment-based techniques [41,50•,57••]. This can lead to erroneous reports that the linear motif is not conserved, while it is in fact conserved at a different location in each ortholog. Finally, detection algorithms face difficulties in distinguishing true positive from false positive instances for degenerate motifs [57••]. In many cases, careful sequence alignment [44,57••,63••], advanced statistics [81] or motif scoring by sequence scanning [41,57••,62••] can overcome these issues.

Conclusions and perspectives

Different experimental and computational approaches show that pathogen linear motifs are highly plastic elements whose evolution can be understood within the framework of *regulatory evolution*. Convergent evolution of motifs across pathogen groups facilitates targeting of similar host processes, and changes in motif repertoires give rise to new phenotypes and can be linked to pathogen adaptive evolution. While most current insights have been gained from studies of viral pathogens, it is likely that many of the mechanisms discussed here also play a role in the evolution of pathogenic bacterial and eukaryotic linear motifs. However, our review of molecular mimicry in pathogen linear motifs suggests that much remains to be learned. We know little about the functional implications of variability in linear motif sequence beyond the regular expression, in binding thermodynamics and kinetics and in conformational properties in the unbound and bound states. Motif repertoires are just starting to be explored through the concept of motif switches, with growing consensus that many pathogen motifs function and evolve in a cooperative manner. We foresee that detailed studies of mimicry, in combination with phylogenetic tools, will make an important contribution to the study of linear motif evolution in pathogens.

Conflict of interest statement

None declared.

Acknowledgements

The authors thank L.G. Alonso, P. Beltrao, N.E. Davey, D.U. Ferreiro, A.M. Moses, M.G. Noval, N. Palopoli and R. Sieira for critical reading of the manuscript. This work was supported by ANPCyT PICT 2012-2550 grant (to I.E.S.) and by ANPCyT PICT 2013-1985 grant (to L.B.C). All authors are CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) researchers.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Kruger D, Grebnev G, Kuban M *et al.*: **The eukaryotic linear motif resource ELM: 10 years and counting**. *Nucleic Acids Res* 2014, **42**:D259-D266.

2. • Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ, Davey NE: **Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation**. *Chem Rev* 2014, **114**:6733-6778.

A thorough review of recent advances in the field of linear motifs, including their evolutionary properties and motif mimicry by viral and non-viral pathogens.

3. Van Roey K, Gibson TJ, Davey NE: **Motif switches: decision-making in cell regulation**. *Curr Opin Struct Biol* 2012, **22**:378-385.

4. Davey NE, Trave G, Gibson TJ: **How viruses hijack cell regulation**. *Trends Biochem Sci* 2011, **36**:159-169.

5. Kadaveru K, Vyas J, Schiller MR: **Viral infection and human disease – insights from minimotifs**. *Front Biosci* 2008, **13**:6455-6471.

6. Dyer MD, Murali TM, Sobral BW: **The landscape of human proteins interacting with viruses and other pathogens**. *PLoS Pathog* 2008, **4**:e32.

7. Elde NC, Malik HS: **The evolutionary conundrum of pathogen mimicry**. *Nat Rev Microbiol* 2009, **7**:787-797.

8. Ruhanen H, Hurley D, Ghosh A, O'Brien KT, Johnston CR, Shields DC: **Potential of known and short prokaryotic protein motifs as a basis for novel peptide-based antibacterial therapeutics: a computational survey**. *Front Microbiol* 2014, **5**:4.

9. Via A, Uyar B, Brun C, Zanzoni A: **How pathogens use linear motifs to perturb host cell networks**. *Trends Biochem Sci* 2015, **40**:36-48.

10. Zhang L, Zhang C, Ojcius DM, Sun D, Zhao J, Lin X, Li L, Li L, Yan J: **The mammalian cell entry (Mce) protein of pathogenic Leptospira species is responsible for RGD motif-dependent infection of cells and animals**. *Mol Microbiol* 2012, **83**:1006-1023.

11. Tonelli RR, Giordano RJ, Barbu EM, Torrecilhas AC, Kobayashi GS, Langley RR, Arap W, Pasqualini R, Colli W, Alves MJ: **Role of the gp85/trans-sialidases in *Trypanosoma cruzi* tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo**. *PLoS Negl Trop Dis* 2010, **4**:e864.

12. Safari F, Murata-Kamiya N, Saito Y, Hatakeyama M: **Mammalian Pragmin regulates Src family kinases via the Glu-Pro-Ile-Tyr-Ala (EPIYA) motif that is exploited by bacterial effectors**. *Proc Natl Acad Sci U S A* 2011, **108**:14938-14943.

13. Evans P, Sacan A, Ungar L, Tozeren A: **Sequence alignment reveals possible MAPK docking motifs on HIV proteins**. *PLoS ONE* 2010, **5**:e8942.

14. Lee CW, Ferreon JC, Ferreon AC, Arai M, Wright PE: **Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation**. *Proc Natl Acad Sci U S A* 2010, **107**:19290-19295.

15. Aitio O, Hellman M, Kazlauskas A, Vingadassalom DF, Leong JM, Saksela K, Permi P: **Recognition of tandem PxxP motifs as a unique Src homology 3-binding mode triggers pathogen-driven actin assembly**. *Proc Natl Acad Sci U S A* 2010, **107**:21743-21748.

16. • Garamszegi S, Franzosa EA, Xia Y: **Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human–virus protein–protein interaction networks**. *PLoS Pathog* 2013, **9**:e1003778.

The study of host–pathogen protein interaction networks at domain resolution reveals that viral proteins preferentially target human proteins containing linear motif-binding domains. This result confirms the relevance of linear motif mimics in virus-host interactions from a network perspective.

17. Chemes LB, Sanchez IE, Smal C, de Prat-Gay G: **Targeting mechanism of the retinoblastoma tumor suppressor by a prototypical viral oncoprotein. Structural modularity, intrinsic disorder and phosphorylation of human papillomavirus E7**. *FEBS J* 2010, **277**:973-988.

18. de Souza RF, Iyer LM, Aravind L: **Diversity and evolution of chromatin proteins encoded by DNA viruses**. *Biochim Biophys Acta* 2010, **1799**:302-318.

19. Zanier K, Charbonnier S, Sidi AO, McEwen AG, Ferrario MG, Poussin-Courmontagne P, Cura V, Brimer N, Babah KO, Ansari T *et al.*: **Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins**. *Science* 2013, **339**:694-698.

20. Liu X, Marmorstein R: **Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor**. *Genes Dev* 2007, **21**:2711-2716.

21. Noval MG, Gallo M, Perrone S, Salvay AG, Chemes LB, de Prat-Gay G: **Conformational dissection of a viral intrinsically disordered domain involved in cellular transformation**. *PLOS ONE* 2013, **8**:e72760.

22. •• Borcherds W, Theillet FX, Katzer A, Finzel A, Mishall KM, Powell AT, Wu H, Manieri W, Dieterich C, Selenko P *et al.*: **Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells**. *Nat Chem Biol* 2014, **10**:1000-1002.

Biophysical and in cell studies show how wildcard residues of a linear motif can modulate its conformation in the unbound form and, as a consequence, the affinity for its target and the functional output of the interaction in the cell.

23. Uversky VN: **Intrinsically disordered proteins may escape unwanted interactions via functional misfolding**. *Biochim Biophys Acta* 2011, **1814**:693-712.

24. Singh M, Krajewski M, Mikolajka A, Holak TA: **Molecular determinants for the complex formation between the retinoblastoma protein and LXCXE sequences**. *J Biol Chem* 2005, **280**:37868-37876.

25. Stein A, Aloy P: **Contextual specificity in peptide-mediated protein interactions**. *PLoS ONE* 2008, **3**:e2524.

26. Chemes LB, Sanchez IE, de Prat-Gay G: **Kinetic recognition of the retinoblastoma tumor suppressor by a specific protein target**. *J Mol Biol* 2011, **412**:267-284.

27. • Aitio O, Hellman M, Skehan B, Kesti T, Leong JM, Saksela K, Permi P: **Enterohaemorrhagic *Escherichia coli* exploits a tryptophan switch to hijack host f-actin assembly**. *Structure* 2012, **20**:1692-1703.

Structural and thermodynamic analysis of a host–pathogen protein–protein interaction shows that bacterial and cellular linear motifs interact with a host SH3 domain in a structurally similar fashion, but the bacterial motif presents a higher affinity due to a tryptophan residue in a position adjacent to the motif.

28. Kaneko T, Huang H, Cao X, Li X, Li C, Voss C, Sidhu SS, Li SS: **Superbinder SH2 domains act as antagonists of cell signaling**. *Sci Signal* 2012, **5**:ra68.

29. Gibson TJ, Seiler M, Veitia RA: **The transience of transient overexpression**. *Nat Methods* 2013, **10**:715-721.

30. Ferreon JC, Martinez-Yamout MA, Dyson HJ, Wright PE: **Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein**. *Proc Natl Acad Sci U S A* 2009, **106**:13260-13265.

31. • Burke JR, Hura GL, Rubin SM: **Structures of inactive retinoblastoma protein reveal multiple mechanisms for cell cycle control**. *Genes Dev* 2012, **26**:1156-1166.

Structural studies of the cell cycle regulator protein retinoblastoma reveal how the transition of the protein from an inactive to an active conformation is the integrated output of several interactions with linear motifs harbored by other proteins and by Retinoblastoma itself and regulated by phosphorylation.

32. Fuxreiter M, Tompa P: **Fuzzy complexes: a more stochastic view of protein function**. *Adv Exp Med Biol* 2012, **725**:1-14.

33. •• Jansma AL, Martinez-Yamout MA, Liao R, Sun P, Dyson HJ, Wright PE: **The high-risk HPV16 E7 oncoprotein mediates interaction between the transcriptional coactivator CBP and the retinoblastoma protein pRb**. *J Mol Biol* 2014, **426**:4030-4048.

Nuclear magnetic resonance unveils how the linear motif mimics in the papillomavirus E7 protein mediate the interaction with the host proteins Retinoblastoma and CBP. Remarkably, two E7 molecules take part in the interaction, with different structural roles.

34. Boutros R, Lobjois V, Ducommun B: **CDC25 phosphatases in cancer cells: key players? Good targets?**. *Nat Rev Cancer* 2007, **7**:495-507.

35. Besson A, Dowdy SF, Roberts JM: **CDK inhibitors: cell cycle regulators and beyond**. *Dev Cell* 2008, **14**:159-169.

36. Primorac I, Musacchio A: **Panta rhei: the APC/C at steady state**. *J Cell Biol* 2013, **201**:177-189.

37. • Ferreon AC, Ferreon JC, Wright PE, Deniz AA: **Modulation of allostery by protein intrinsic disorder**. *Nature* 2013, **498**:390-394.
The interaction between the adenovirus E1A protein and the host retinoblastoma and CBP proteins can have either positive or negative cooperativity depending on the available interaction sites, showing how the presence of multiple linear motif mimics in a pathogen protein may lead to context-specific functional outputs.

38. Felsani A, Mileo AM, Paggi MG: **Retinoblastoma family proteins as key targets of the small DNA virus oncoproteins**. *Oncogene* 2006, **25**:5277-5285.

39. Breuer S, Schievink SI, Schulte A, Blankenfeldt W, Fackler OT, Geyer M: **Molecular design, functional characterization and structural basis of a protein inhibitor against the HIV-1 pathogenicity factor Nef**. *PLoS ONE* 2011, **6**:e20033.

40. Wagner A: *The Origin of Evolutionary Innovations: A Theory of Transformative Change in Living Systems*. Oxford: Oxford Press; 2011.

41. Moses AM, Liku ME, Li JJ, Durbin R: **Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites**. *Proc Natl Acad Sci U S A* 2007, **104**:17713-17718.

42. Beltrao P, Serrano L: **Specificity and evolvability in eukaryotic protein interaction networks**. *PLoS Comput Biol* 2007, **3**:e25.

43. Neduva V, Russell RB: **Linear motifs: evolutionary interaction switches**. *FEBS Lett* 2005, **579**:3342-3345.

44. Chemes LB, Glavina J, Alonso LG, Marino-Buslje C, de Prat-Gay G, Sanchez IE: **Sequence evolution of the intrinsically disordered and globular domains of a model viral oncoprotein**. *PLoS ONE* 2012, **7**:e47661.

45. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW: **Evolution and disorder**. *Curr Opin Struct Biol* 2011, **21**:441-446.

46. Patel MR, Loo YM, Horner SM, Gale M Jr, Malik HS: **Convergent evolution of escape from hepaciviral antagonism in primates**. *PLoS Biol* 2012, **10**:e1001282.

47. Rosso L, Marques AC, Weier M, Lambert N, Lambot MA, Vanderhaeghen P, Kaessmann H: **Birth and rapid subcellular adaptation of a hominoid-specific CDC14 protein**. *PLoS Biol* 2008, **6**:e140.

48. Tan CS, Pasculescu A, Lim WA, Pawson T, Bader GD, Linding R: **Positive selection of tyrosine loss in metazoan evolution**. *Science* 2009, **325**:1686-1688.

49. Via A, Gherardini PF, Ferraro E, Ausiello G, Scalia Tomba G, Helmer-Citterich M: **False occurrences of functional motifs in protein sequences highlight evolutionary constraints**. *BMC Bioinform* 2007, **8**:68.

50. • Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ: **Attributes of short linear motifs**. *Mol Biosyst* 2012, **8**:268-281.
The authors study the conservation patterns, sequence composition and functional features of linear motifs from the ELM database, and provide the first estimate of the degree to which motif instances occur in unrelated protein sequences.

51. Zielinska DF, Gnad F, Schropp K, Wisniewski JR, Mann M: **Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery**. *Mol Cell* 2012, **46**:542-548.

52. •• Nguyen Ba AN, Strome B, Hua JJ, Desmond J, Gagnon-Arsenault I, Weiss EL, Landry CR, Moses AM: **Detecting functional divergence after gene duplication through evolutionary changes in posttranslational regulatory sequences**. *PLoS Comput Biol* 2014, **10**:e1003977.
This study by the Moses group undertakes a comprehensive analysis of the rates of evolution of linear motifs in the yeast genome, identifying evolutionary patterns for the diversification of these sites in paralogs. The study shows that changes in short linear motifs may underlie functional divergence following genome duplication.

53. Hittinger CT, Carroll SB: **Evolution of an insect-specific GROUCHO-interaction motif in the ENGRAILED selector protein**. *Evol Dev* 2008, **10**:537-545.

54. Bruning JB, Shamoo Y: **Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1**. *Structure* 2004, **12**:2209-2219.

55. Holt LJ, Tuch BB, Villen J, Johnson AD, Gygi SP, Morgan DO: **Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution**. *Science* 2009, **325**:1682-1686.

56. • Goldman A, Roy J, Bodenmiller B, Wanka S, Landry CR, Aebersold R, Cyert MS: **The calcineurin signaling network evolves via conserved kinase–phosphatase modules that transcend substrate identity**. *Mol Cell* 2014, **55**:422-435.
In this striking example, the authors describe the evolution of the calcineurin network, showing how conservation of kinase/phosphatase pairs can subsist while substrates present high turnover. Network evolution and identity are shaped by the gain and loss of dual kinase–phosphatase docking sites, which are poorly conserved, and have appeared recently in evolution.

57. •• Hagai T, Azia A, Babu MM, Andino R: **Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions**. *Cell Rep* 2014, **7**:1729-1739.
A computational analysis of over 2000 viruses identifies many linear motif pairs present in different viruses, indicative of convergent evolution of motif switches. A simple evolutionary model shows that over 50% of the possible missense mutations in the HIV genome would change the linear motif repertoire of the virus, suggesting that motif evolution may allow fast rewiring of host–virus interactions.

58. Pushker R, Mooney C, Davey NE, Jacque JM, Shields DC: **Marked variability in the extent of protein disorder within and between viral families**. *PLOS ONE* 2013, **8**:e60724.

59. Duffy S, Shackelton LA, Holmes EC: **Rates of evolutionary change in viruses: patterns and determinants**. *Nat Rev Genet* 2008, **9**:267-276.

60. Daugherty MD, Malik HS: **Rules of engagement: molecular insights from host–virus arms races**. *Annu Rev Genet* 2012, **46**:677-700.

61. Wei P, Wong WW, Park JS, Corcoran EE, Peisajovich SG, Onuffer JJ, Weiss A, Lim WA: **Bacterial virulence proteins as tools to rewire kinase pathways in yeast and immune cells**. *Nature* 2012, **488**:384-388.

62. •• Chemes LB, Glavina J, Faivovich J, de Prat-Gay G, Sanchez IE: **Evolution of linear motifs within the papillomavirus E7 oncoprotein**. *J Mol Biol* 2012, **422**:336-346.
This study presents the first phylogenetic reconstruction of linear motif history for a viral protein, documenting convergent evolution as well as high evolutionary plasticity of linear motifs within the E7 papillomavirus phylogeny, which are linked to phenotypic changes and likely involved in adaptive evolution.

63. •• Lo MK, Sogaard TM, Karlin DG: **Evolution and structural organization of the C proteins of paramyxovirinae**. *PLOS ONE* 2014, **9**:e90003.
Using parallel alignments of the same genomic region from the disordered PNT region, or from an alternative overlapping ORF from the globular C protein within the paramyxovirus genome, the authors show how the evolutionary history of motifs can be erroneously inferred. This study is a striking example of the difficulties arising when trying to study conservation of linear motifs based on alignments of disordered regions.

64. Karlin D, Belshaw R: **Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of Mononegavirales phosphoproteins**. *PLoS ONE* 2012, **7**:e31719.

65. Jackson D, Hossain MJ, Hickman D, Perez DR, Lamb RA: **A new influenza virus virulence determinant: the NS1 protein four C-terminal residues modulate pathogenicity**. *Proc Natl Acad Sci U S A* 2008, **105**:4381-4386.

66. Prehaud C, Wolff N, Terrien E, Lafage M, Megret F, Babault N, Cordier F, Tan GS, Maitrepierre E, Menager P *et al.*: **Attenuation of**

rabies virulence: takeover by the cytoplasmic domain of its envelope protein. *Sci Signal* 2010, **3**:ra5.

67. Yang CW: **A comparative study of short linear motif compositions of the influenza A virus ribonucleoproteins**. *PLoS ONE* 2012, **7**:e38637.

68. Igarashi M, Ito K, Kida H, Takada A: **Genetically destined potentials for N-linked glycosylation of influenza virus hemagglutinin**. *Virology* 2008, **376**:323-329.

69. Abe Y, Takashita E, Sugawara K, Matsuzaki Y, Muraki Y, Hongo S: **Effect of the addition of oligosaccharides on the biological activities and antigenicity of influenza A/H3N2 virus hemagglutinin**. *J Virol* 2004, **78**:9605-9611.

70. •• Lawrence P, LaRocco M, Baxt B, Rieder E: **Examination of soluble integrin resistant mutants of foot-and-mouth disease virus**. *Virol J* 2013, **10**:2.

This *in vitro* evolution study of the FMDV virus shows how specific selective pressures, in this case soluble integrin decoys, lead to the generation of resistant viruses that harbor compensatory changes in sequence composition of the VP1 protein variable loop region containing the integrin-binding RGD motif. This example illustrates plasticity and fast adaptation of viruses based on mutations in cell attachment linear motifs.

71. Pim D, Bergant M, Boon SS, Ganti K, Kranjec C, Massimi P, Subbaiah VK, Thomas M, Tomaic V, Banks L: **Human papillomaviruses and the specificity of PDZ domain targeting**. *FEBS J* 2012, **279**:3530-3537.

72. Neveu G, Cassonnet P, Vidalain PO, Rolloy C, Mendoza J, Jones L, Tangy F, Muller M, Demeret C, Tafforeau L *et al.*: **Comparative analysis of virus-host interactomes with a mammalian high-throughput protein complementation assay based on Gaussia princeps luciferase**. *Methods* 2012, **58**:349-359.

73. Dampier W, Evans P, Ungar L, Tozeren A: **Host sequence motifs shared by HIV predict response to antiretroviral therapy**. *BMC Med Genomics* 2009, **2**:47.

74. • Purdy MA, Lara J, Khudyakov YE: **The hepatitis E virus polyproline region is involved in viral adaptation**. *PLOS ONE* 2012, **7**:e35974.

This study analyzes evolutionary rates of a motif-rich polyproline region in the Hepatitis E ORF1 region. The authors describe high non synonymous to synonymous ratios in this region ($dN/dS > 1$ for 4–10 codons) and differences in the extent of sequence divergence for viral genotypes adapted to different host ranges.

75. Dolan PT, Roth AP, Xue B, Sun R, Dunker AK, Uversky VN, LaCount DJ: **Intrinsic disorder mediates hepatitis C virus core – host cell protein interactions**. *Protein Sci* 2015, **24**:221-235.

76. Fares MA, Moya A, Escarmis C, Baranowski E, Domingo E, Barrio E: **Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens**. *Mol Biol Evol* 2001, **18**:10-21.

77. McGivern DR, Villanueva RA, Chinnaswamy S, Kao CC, Lemon SM: **Impaired replication of hepatitis C virus containing mutations in a conserved NS5B retinoblastoma protein-binding motif**. *J Virol* 2009, **83**:7422-7433.

78. Hertz T, Nolan D, James I, John M, Gaudieri S, Phillips E, Huang JC, Riadi G, Mallal S, Jojic N: **Mapping the landscape of host–pathogen coevolution: HLA class I binding and its relationship with evolutionary conservation in human and viral proteins**. *J Virol* 2011, **85**:1310-1321.

79. Sarmady M, Dampier W, Tozeren A: **Sequence- and interactome-based prediction of viral protein hotspots targeting host proteins: a case study for HIV Nef**. *PLoS ONE* 2011, **6**:e20735.

80. He L, De Groot AS, Gutierrez AH, Martin WD, Moise L, Bailey-Kellogg C: **Integrated assessment of predicted MHC binding and cross-conservation with self reveals patterns of viral camouflage**. *BMC Bioinform* 2014, **15(Suppl)**:S1.

81. Davey NE, Haslam NJ, Shields DC, Edwards RJ: **SLiMFinder: a web server to find novel, significantly over-represented, short protein motifs**. *Nucleic Acids Res* 2010, **38**:W534-W539.