Review

CellPress

How pathogens use linear motifs to perturb host cell networks

Allegra Via[1], Bora Uyar[2], Christine Brun[3,4,5], and Andreas Zanzoni[3,4]

[1] Department of Physics, Sapienza University, 00185 Rome, Italy
[2] Structural and Computational Biology, European Molecular Biology Laboratory, 69117 Heidelberg, Germany
[3] Inserm, UMR1090 TAGC, Marseille F-13288, France
[4] Aix-Marseille Université, UMR1090 TAGC, Marseille F-13288, France
[5] CNRS, Marseille F-13402, France

Molecular mimicry is one of the powerful stratagems that pathogens employ to colonise their hosts and take advantage of host cell functions to guarantee their replication and dissemination. In particular, several viruses have evolved the ability to interact with host cell components through protein short linear motifs (SLiMs) that mimic host SLiMs, thus facilitating their internalisation and the manipulation of a wide range of cellular networks. Here we present convincing evidence from the literature that motif mimicry also represents an effective, widespread hijacking strategy in prokaryotic and eukaryotic parasites. Further insights into host motif mimicry would be of great help in the elucidation of the molecular mechanisms behind host cell invasion and the development of anti-infective therapeutic strategies.

Pathogens, molecular mimicry, and linear motifs

Increasing evidence indicates that phylogenetically distant pathogens show remarkably similar *modi operandi* in host cell entry and subversion [1,2]. Commonalities are expected to occur particularly at the pathogen–host interface; notably in pathogen macromolecules involved in cytoadherence and/or cell penetration. In both cases, pathogens belonging to distant clades are confronted with similar host components and pathways.

Transient protein–protein interactions are often mediated by short stretches of contiguous amino acids known as SLiMs (Box 1) [3] that embody functions independently of a larger sequence and structure context. In many cases, the presence of a SLiM is sufficient to promote ligand binding, targeting, and control of protein stability and, more generally, to regulate several pathways, provided the motif is adequately exposed on the protein surface. SLiMs are mostly located in natively disordered protein regions and, if within folded domains, tend to reside in accessible loops [3], which are evolutionarily variable segments where motifs may appear or disappear as a result of single point mutations [4]. This plasticity of SLiMs makes them ideal elements to tune functionality in eukaryotic regulatory proteins. This adaptive process is facilitated by convergent evolution events, in which similar SLiMs may arise *de novo* in unrelated protein sequences. Including post-transcriptional modification sites, it has been estimated that more than 1 million linear motif functional sites may exist in the human proteome [5], a testament to the complexity of cell regulatory systems. Pathogenic organisms may, however, take advantage of host systems if their secreted proteins also contain convergently evolved SLiMs. These are the so-called 'mimicry motifs' (Box 2) – short stretches of amino acids that are similar, if not identical, to host SLiMs in both composition and function. Thus, motif plasticity represents a double-edged sword or, as nicely expressed by Davey *et al.* [6], creates an Achilles heel permitting pathogens to exploit the host cell machinery to interact with host pathways.

The occurrence of mimicry motifs has been extensively reported in viral proteomes [6–8]. As highlighted previously [6], these organisms, probably thanks to their quickly evolving genomes, have evolved numerous host-resembling SLiMs that facilitate their internalisation into the host cell and the manipulation of a wide range of host cellular pathways involved in immune response, cell cycle, and transcription regulation [6]. It is worth noting that experimentally validated examples of viral mimicry could be shown for one-third of the motif classes in the Eukaryotic Linear Motif (ELM) database [9] and recent computational studies identified a large number of potentially functional eukaryotic motifs in an exhaustive set of viral genomes [7,8].

The extensive use of viral motif mimicry suggests that this tactic could be a general practice not only in viruses but also among pathogens belonging to other taxonomic domains such as bacterial and eukaryotic parasites, especially if they invade host cells. For instance, Src homology 3 (SH3)-interacting proline-rich motifs (see Glossary) are found in proteins from various pathogen phyla, from bacteria to Apicomplexa, where they are generally used to interact with SH3 domains within the infected cell and modulate host signalling pathways [10].

Here we highlight the growing number of motif mimicry examples being observed in pathogens belonging to the cellular domains of life. While the available data is much less than for viruses, it does raise the question of to what extent motif mimicry represents the expression of a

Corresponding author: Zanzoni, A. (zanzoni@tagc.univ-mrs.fr).
Keywords: molecular mimicry; short linear motifs; SLiMs; host–pathogen interactions; virulence.

0968-0004/
 http://dx.doi.org/10.1016/j.tibs.2014.11.001

Glossary

14-3-3 proteins: a family of evolutionarily conserved proteins that play a key role in multiple biological processes by interacting with a plethora of client proteins. They bind to phosphorylated serine or threonine residues.
CagL protein: a component of the type IV secretion system (T4SS) of the gastric pathogen *Helicobacter pylori*. It is a specialised adhesin of the T4SS pilus interacting via an RGD motif with host α5β3 and α5β1 integrins. It is required for pathogen adhesion to gastric epithelial cells.
***Campylobacter* invasion antigens (Cia):** proteins exported via a type III secretion system and delivered to the host cell to promote maximal cell invasion.
CD36: an integral membrane protein found on the surface of many cell types in vertebrate animals.
CLAVATA3/embryo surrounding region (ESR) (CLE)-related gene family: a gene family comprising numerous genes that contain conserved CLE domains in various plant species and plant-parasitic nematodes. Plant CLE genes encode small proteins with an N-terminal secretion signal peptide and a conserved 14-amino acid domain called the CLE motif at the C terminus. CLE proteins have roles in shoot, floral, and root meristem maintenance, organ size regulation, apical dominance, and vascular development.
C-Src kinase (Csk): an enzyme that phosphorylates tyrosine residues located in the C-terminal end of Src family kinases.
CT10 regulator of kinase (Crk) adaptor family: a family of important adaptor molecules that participate in diverse signalling pathways and localises to EPEC pedestals.
Cytotoxicity-associated immunodominant antigen (CagA): a phosphotyrosine-containing protein that is secreted by the *H. pylori* type IV secretion system. It induces morphological changes in the infected cell by interacting with proteins of the host's signalling pathways, including Grb2, Shp2, and Csk.
Dense granule protein 16 (GRA16): a dense granule protein that is exported through the *Toxoplasma gondii* vacuole membrane and reaches the host cell nucleus, where it positively modulates genes involved in cell cycle progression and the p53 tumour-suppressor pathway.
Dense granule protein 24 (GRA24): a *T. gondii* protein secreted from the PV to the host cell nucleus, where it activates host kinases using two high-affinity MAPK-docking motifs.
ERK2: a serine/threonine kinase that plays a critical role in the regulation of cell growth and differentiation.
EspF(U) (or TccP) effector protein: EHEC protein injected through a type III secretion system into host cells, where it stimulates actin polymerisation by activating host WASP proteins.
Exoenzyme S (ExoS): a *Pseudomonas aeruginosa* type III secretion effector targeting multiple substrates in the host. It exerts complex effects on eukaryotic cell function, including inhibition of DNA synthesis, alterations in cell morphology, microvillus effacement, and loss of cellular adherence.
Fusicoccin: a phytotoxic terpenoid secreted by the fungus *Phomopsis amygdali*. The terpenoid binds and stabilises the host H^+-ATPase–14-3-3 complex, thus irreversibly activating plasma membrane H^+-ATPase and inducing uncontrolled stomata opening.
Growth factor receptor-bound protein 2 (Grb2): an adaptor protein involved in signal transduction. It binds several membrane receptors and contains one SH2 domain and two SH3 domains.
Hrp outer protein Q (HopQ1): a type III effector secreted by *Pseudomonas syringae* effector protein. It enhances bacterial virulence and associates with host 14-3-3 proteins in a phosphorylation-dependent manner.
Inclusion membrane protein G (incG): one of the transmembrane proteins of the chlamydial inclusion, a vacuole in which the *Chlamydia trachomatis* developmental cycle occurs.
Integrins: a family of heterodimeric receptors that link the surface of cells to various extracellular membrane matrix components. They mediate the transduction of cell–extracellular membrane matrix signalling.
Neuronal Wiskott–Aldrich syndrome protein (N-WASP): a family of proteins involved in transduction of signals from receptors on the cell surface to the actin cytoskeleton.
Non-catalytic region of tyrosine kinase (Nck) adaptor protein: an adaptor protein involved in transducing signals from receptor tyrosine kinases to downstream signalling proteins. It contains SH2 and SH3 domains and interacts with the WASP–Arp2/3 complex to coordinate actin cytoskeletal remodelling.
OspF effector family: type III secretion system effectors that downregulate the host innate immune response.
p38 mitogen-activated protein kinases (MAPKs): a class of MAPKs that are responsive to various stress stimuli such as cytokines, lipopolysaccharides, UV light, heat, and osmotic shock. They are also involved in cell differentiation and apoptosis.
Pfg27: a *Plasmodium falciparum* sexual stage-specific protein involved in maintaining cell integrity during the uniquely long gametocytogenesis of the parasite.
Pilus-biogenesis factor (PiY1): an essential, calcium-dependent regulator of *P. aeruginosa* twitching/surface motility.
***P. falciparum* erythrocyte membrane protein 1 (PfEMP1):** a protein encoded by the var genes. It interacts with adhesion molecules such as ICAM-1, CD36, and TSP via various domains.
Protein toxin A (ToxA): a host-selective toxin that is internalised into wheat mesophyll cells via RGD motif-mediated interaction with host cell integrins.
Thrombospondin (TSP)-related anonymous protein (PfTRAP): a *P. falciparum* type 1 membrane protein that possesses multiple adhesive domains in its extracellular region. It is essential for sporozoite motility and for liver cell invasion.
Shp2: a protein tyrosine phosphatase encoded by the gene PTPN11 and involved in several intracellular signalling pathways. It contains two SH2 domains.
Src homology 2 (SH2) domain: conserved docking modules recognising and interacting with phosphorylated tyrosine residues. They are found in several intracellular signalling proteins.
Src homology 3 (SH3) domain: conserved docking modules recognising and interacting with polyproline motifs. They are found in several intracellular signalling proteins.
Translocated intimin receptor (Tir): one of the effectors delivered into host cells by the EPEC type III secretion system. It drives the major pathway responsible for regulating actin polymerisation in the host cell.
XopQ protein: a type III effector protein found in phytopathogens of the genus *Xanthomonas*.

general strategy used by pathogens to 'sneak' into host cell regulatory networks.

Exploratory literature searches revealed that motif mimicry events occur in both prokaryotic and eukaryotic parasites (Table 1), where they represent the etiological agents of distinct diseases in both plants and animals (Table 2). Mimicry motifs facilitate adhesion to host cells and the release of pathogenic effectors that can ultimately reach specific intracellular target sites and perturb interaction networks. Interestingly, among the representative examples of pathogen mimicry of host SLiMs reported in Table 1, there are some – such as Arg–Glu–Asp (RGD), mitogen-activated protein kinase (MAPK) docking, and SH3 domains or 14-3-3-binding motifs – that occur in proteins of pathogens spanning different phyla (Figures 1 and 2) and are also observed in viruses [6] (Box 3).

We discuss examples from the growing literature on the subject, revealing how even evolutionarily distant pathogens exploit molecular mimicry of host-like motifs to assist in host–pathogen interactions.

Mimicry motifs in prokaryotic pathogens

RGD

The extracellular tripeptide motif RGD plays a crucial role in cell adhesion by mediating the interaction of several extracellular glycoproteins, such as thrombospondins (TSPs) and cell adhesion receptors, with members of the integrin superfamily [11].

Several pathogenic bacteria are known to possess RGD motifs in surface proteins that allow them to establish intimate contacts with host tissues, an essential preliminary step of infection, favouring either pathogen internalisation or the injection of specific protein effectors into host cells [12]. For instance, the *Helicobacter pylori* CagL protein, a component of the type IV secretion system, has an RGD sequence that mediates the interaction with the host α5β3 and α5β1 integrins and is required for pathogen adhesion to gastric epithelial cells [13,14]. Other examples of RGD mimicry by pathogenic bacteria were identified in the calcium-dependent pilus biogenesis factor (PiY1) [15] of *Pseudomonas aeruginosa*, an opportunistic pathogen of immunocompromised individuals, and in the lipoprotein T of *Mycoplasma conjunctivae*, which is responsible for infectious keratoconjunctivitis in domestic sheep [16].