



Potential of known and short prokaryotic protein motifs as a basis for novel peptide-based antibacterial therapeutics: a computational survey

Heini Ruhanen^{1,2,3}, Daniel Hurley^{1,2,3}, Ambarnil Ghosh⁴, Kevin T. O'Brien^{1,2,3}, Catriona R. Johnston⁵ and Denis C. Shields^{1,2,3}*

¹ Complex and Adaptive Systems Laboratory, University College Dublin, Dublin, Ireland

² Conway Institute of Biomolecular and Biomedical Science, University College Dublin, Dublin, Ireland

³ School of Medicine and Medical Science, University College Dublin, Dublin, Ireland

⁴ Crystallography and Molecular Biology Department, Saha Institute of Nuclear Physics, Kolkata, India

⁵ Translational Research Institute, Woolloongabba, QLD, Australia

Edited by:

Nádia S. Parachin, Universidade de Brasília, Brazil

Reviewed by:

M. Pilar Francino, Center for Public Health Research, Spain

Luis C. N. Da Silva, Universidade Federal de Pernambuco, Brazil

Viji Sarojini, University of Auckland, New Zealand

Mrinal Bhawe, Swinburne University of Technology, Australia

*Correspondence:

Denis C. Shields, Complex and Adaptive Systems Laboratory, University College Dublin, Belfield Office Park-8, Dublin 4, Ireland
e-mail: denis.shields@ucd.ie

Short linear motifs (SLiMs) are functional stretches of protein sequence that are of crucial importance for numerous biological processes by mediating protein–protein interactions. These motifs often comprise peptides of less than 10 amino acids that modulate protein–protein interactions. While well-characterized in eukaryotic intracellular signaling, their role in prokaryotic signaling is less well-understood. We surveyed the distribution of known motifs in prokaryotic extracellular and virulence proteins across a range of bacterial species and conducted searches for novel motifs in virulence proteins. Many known motifs in virulence effector proteins mimic eukaryotic motifs and enable the pathogen to control the intracellular processes of their hosts. Novel motifs were detected by finding those that had evolved independently in three or more unrelated virulence proteins. The search returned several significantly over-represented linear motifs of which some were known motifs and others are novel candidates with potential roles in bacterial pathogenesis. A putative C-terminal G[AG].\$ motif found in type IV secretion system proteins was among the most significant detected. A KK\$ motif that has been previously identified in a plasminogen-binding protein, was demonstrated to be enriched across a number of adhesion and lipoproteins. While there is some potential to develop peptide drugs against bacterial infection based on bacterial peptides that mimic host components, this could have unwanted effects on host signaling. Thus, novel SLiMs in virulence factors that do not mimic host components but are crucial for bacterial pathogenesis, such as the type IV secretion system, may be more useful to develop as leads for anti-microbial peptides or drugs.

Keywords: short linear motifs (SLiMs), virulence factor, motif mimicry, antibacterial, bioinformatics, pathogen

INTRODUCTION

Short linear motifs (SLiMs) are functional microdomains in proteins that play a critical role in many distinct biological processes such as cell signaling and regulation, post-translational modifications, proteolytic cleavage, and protein trafficking (Davey et al., 2011b; Mooney et al., 2012). These motifs are typically found in eukaryotic disordered protein regions and vary in size from 3 to 12 amino acids (Fuxreiter et al., 2007). In general, SLiMs have less than five defined amino acid positions and frequently these positions have some degree of flexibility in amino acid composition. Their shortness makes them evolutionarily plastic, allowing them to evolve convergently in unrelated proteins. This can allow proteins to rapidly acquire new protein interaction functions (Nedeva and Russell, 2005; Diella et al., 2008; Davey et al., 2010, 2012b). Their short length also presents a challenge for SLiM discovery both experimentally and computationally, since there may be many false positive findings using both methods.

The presence of SLiMs in eukaryotes and viruses has been well-established. Several pioneering viral studies were crucial for the original characterization of SLiMs (Davey et al., 2011b). Viruses use SLiMs as a principal mechanism of hijacking cells by binding to host proteins and recruiting them to process viral proteins. A viral genome can contain various short motifs, many of which are necessary for the viral life cycle, providing a plethora of ways for the virus to take over the molecular machinery of the host cell (Kadaveru et al., 2008; Davey et al., 2011b). Like viruses, pathogenic bacteria are extremely proficient in intercepting host cell functions and in many cases it is still poorly understood how bacteria carry out the manipulation of the host cells. SLiMs have been documented in a number of cases to play a role in bacterial pathogenicity. However, bacterial linear motifs are not as well-characterized as in eukaryotes.

Most of the known instances of bacterial motifs are involved in pathogenicity including signals in effector proteins or host motif

mimicry (Cornelis and Van Gijsegem, 2000; Alto et al., 2006). The tripeptide RGD motif is a known host extracellular matrix adhesion factor that is also used by bacteria to attach onto host cells (Tegtmeyer et al., 2010; Zimmermann et al., 2010; Zhang et al., 2012). RGD based anticancer and antithrombotic drugs are currently being developed but their direct impact on limiting bacterial adhesion and infectivity has not been investigated. A second example of a bacterial motif is the EPIYA motif found in several bacterial type III or IV secretion system effector proteins, which mimics SH2 binding peptides of the host (Hayashi et al., 2013). A third example of a bacterial motif has evolved to antagonize host proteins, but does this using a motif for which there is no eukaryotic equivalent. This W...E motif (where "." indicates any amino acid) in bacterial effector proteins has been proposed to mimic host G-proteins (Alto et al., 2006; Jackson et al., 2008; Ham et al., 2009). Other motifs found in prokaryotes which are not simply mimicking known eukaryotic motifs play roles in transport, modification and proteolysis of the bacterial proteins (Table 1).

Since SLiMs are used in a plethora of cellular processes in eukaryotes and are utilized by both pathogenic bacteria and viruses, discovering and characterizing new linear motifs is of great importance. As well as shedding light on the mechanisms of fundamental cellular processes they also hold promise as future therapeutic targets. There is an urgent need for new classes of antimicrobial therapeutics that are effective against multidrug resistant bacteria. Conventional antibiotics are becoming increasingly ineffective against pathogenic bacteria, such as methicillin resistant *Staphylococcus aureus* (MRSA) which presents a severe threat to public health.

We were interested in whether SLiMs may be valuable when developing new antimicrobial peptides or drugs. Compared with recombinant proteins, the smaller size of peptides makes them easier to manufacture and deliver. The use of chemically synthesized peptides in pharmacological and clinical applications is relatively limited by their low systemic stability and high clearance, poor membrane permeability, negligible activity when administered orally and their high cost of manufacture in comparison to small chemical compounds. However, to date more than 100 peptide-based drugs have already reached the market and of these, the majority are at the smaller end of the size spectrum at 8–10 amino acids (Craik et al., 2013).

Here, we conducted a study to discover SLiMs computationally in bacterial virulence factor datasets. We surveyed the distribution of these novel motifs, and compared their distribution with that of known motifs observed in prokaryotic proteins. The list of motifs given here represents a useful resource for experimental scientists interested in targeting SLiMs that may be important for the pathogenicity of bacteria.

MATERIALS AND METHODS

We utilized data from a virulence factor database MvirDB (Lawrence Livermore National Laboratory), which integrates DNA and protein sequence information from Tox-Prot, SCORPION, the PRINTS database of virulence factors, VFDB, TVFac, Islander, ARGO, CONUS, KNOTTIN, a subset of VIDA and sequences derived by means of literature searches (Zhou et al., 2007). MvirDB can be accessed at <http://mvirdb.llnl.gov>.

The MvirDB browser tool was used to search the database to retrieve virulence factors by functional categories (Table 2) and to download sequences of interest. Protein sequence identifiers for the downloaded sequences for each functional category are available in Table S1.

The recovered protein sequences in each functional category thought to be associated with pathogenicity were searched for SLiMs using SLiMfinder (Davey et al., 2010) both locally, and on a webserver that is available at <http://bioware.ucd.ie>. The default settings provided in SLiMfinder without any extra masking were used in the analysis. This method finds sets of three or more unrelated proteins in a dataset of proteins that share a motif. Chemotaxis and enzyme protein sequence datasets were filtered to contain only sequences longer than 20 amino acids and lipoprotein and Exotoxin datasets sequences longer than 40 amino acids prior to the analysis.

The motifs identified by the SLiMfinder analysis were further examined for similarity to known SLiMs from literature motifs using CompariMotif, which takes two lists of protein motifs and compares them to each other, identifying and scoring similarities between short motifs in the sets (Edwards et al., 2008).

Motifs were visualized using the MEME Suite (Bailey et al., 2009), by taking a stretch of 10 amino acid residues containing the motif of interest from each protein sequence where the motif was found. MEME represents motifs as position dependent letter probability matrices which describe the probability of each possible letter at each position in the pattern. These are displayed as "sequence LOGOS," containing stacks of letters at each position in the motif. The total height of the stack is the "information content" of that position in the motif in bits. The height of the individual letters in a stack is the probability of the letter at that position multiplied by the total information content of the stack.

Datasets comprised of protein sequences obtained from UniProtKB that are predicted to be effector proteins from a selection of 60 organisms represented in the MvirDB were used to assess the distribution of prokaryotic protein motifs. The presence of both known and novel motifs in these datasets was investigated using the predictive computational tool SLiMsearch which can be used to determine the occurrences of predefined motifs in protein sequences (Davey et al., 2011a). Heat maps were generated to visualize the incidences of motifs in the protein datasets where the frequency of the heat map represents the logarithm of the normalized N_UPC (Number of incidences of a motif in an Unrelated Protein Cluster) value returned in the SLiMsearch results. The N_UPC for an individual motif in a specific organism was normalized by dividing the value by the total amount of UPCs (Unrelated Protein Clusters) in the specific organism and the average N_UPCs of a motif across all 60 organisms. For motifs where there were no incidences in a specific organism the frequency was set to an arbitrary value lower than the minimum actual observed value.

The organisms in Figures 2, 3 which cover the motif sequences were presented in a phylogenetic tree (Figure 4). The Taxonomic IDs for all the organisms are used as input in NCBI's Taxonomy Common Tree tool (<http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>). The "phenogram" taxonomic tree (*.phy format) obtained from the NCBI server was fed into

Table 1 | Examples of known instances of Short Linear Motifs in bacterial virulence factors.

Virulence factor	Motif	Function	References
ADHERENCE			
CagL, Mce	RGD*, NGR	An integrin binding cell adhesion motif	Conradi et al., 2012a,b; Zhang et al., 2012
CagL YadA	FEANE [SG][VI][AS][IVT]G..S	Participates in integrin binding Repeated collagen binding motif	Conradi et al., 2012b Tahir et al., 2000
InlJ, InlA, nanA, CspA	LPTG	Cell wall anchor motif	Harris et al., 2003; Sabet et al., 2005; Banerjee et al., 2010
Eno	[LF]Y[DNK]...[KG][KV]Y[VD]*	Plasminogen binding motif	Bergmann et al., 2003; Nogueira et al., 2012
CBPA-G, LytA-C	W[WFY][FY]...G.M	Repeated cholin (cell wall) binding motif	Garau et al., 2005
SpsA	YRNYPT	Host secretory immunoglobulin A (SIgA) and secretory component (SC) binding motif	Hammerschmidt et al., 2000
EFFECTOR			
SipA, SopA	DEVD*, [DSTE][^P][^DEWHFYC]D[GSAN]	Caspase 3 cleavage site motif	Srikanth et al., 2010; Dinkel et al., 2012
AvrPto, HopF2, AvrB, AvrRpm1	MG..C*.G... [STC], ^M{0,1}(G)[^EDRKHPFYW]..[STAGCN][^P]	N-myristoylation/S-palmitoylation motif	Shan et al., 2000; Robert-Seilaniantz et al., 2006; Dinkel et al., 2012; Hicks and Galan, 2013
WtsE, AvrE1, IpgB2, IpgB1, Map, EspM, EspT, SifA, SifB	W...E*	Host Rho GTPase activation/modulation motif	Alto et al., 2006; Ham et al., 2009
WtsE,	[LR][KQVS][KQLR][EST][GQR][FLKS][EGPK] [MLVAS][KNAL][SGIE]*	Putative endoplasmic reticulum membrane retention/retrieval motif	Ham et al., 2006, 2009
SifA, AnkB	CLCCFL*, (C)[^DENO][LIVM].\$	CAAX box, putative prenylation motif (addition of farnesyl or geranylgeranyl group)	Boucrot et al., 2003; Hicks et al., 2011; Dinkel et al., 2012
YopE, SptP, ExoS	G.LR...T(YopE*)	Arginine finger motif, essential for Rho GAP function of virulence factors	Black and Bliska, 2000; Wurtele et al., 2001
PopB, PopP2, AvrBs3	[^DE]((K[RK]))(RK)[KRP][KR][^DE], [KR][KR].[7,15][^DE]((K[KR]))(RK)(([^DE][KR]) [KR][^DE]))[^DE], [^DE]((K[RK]))(RK)(([^DE][KR]))((KR)[^DE]))((PKR) [I[^DE][DE]]), ((PKR).[0,1][^DE])((PKR))((K[RK]))(RK) (([^DE][KR]))((KR)[^DE]))[^DE]	Nuclear localization signal (NLS) motifs	Szurek et al., 2002; Deslandes et al., 2003; Dean, 2011; Dinkel et al., 2012
CagA, Tarp, AnkA, LspA	E[PNS][IV]Y[AEG]	Membrane targeting/phosphorylation motif	Higashi et al., 2005; Suzuki et al., 2009; Hayashi et al., 2013
SspH2, SseI	...GSGC...G(C)M[GS][CL][KP]C, ^M{0,1}G(C)..S[AKS]	S-palmitoylation motif	Hicks et al., 2011; Dinkel et al., 2012

(Continued)

Table 1 | Continued

Virulence factor	Motif	Function	References
ExoS, SopE	[FIV]..[FIV].[FIV]..[NC].[FIV]	Membrane localization motif (targets ExoS to the Golgi-endoplasmic reticulum)	Zhang and Barbieri, 2005
SopD2, SifA, SseJ, SspH2	WEK(IM)..FF	Translocation/late endocytic compartments targeting motif	Brown et al., 2006
ExoU	KAWRN	Plasma membrane localization/ubiquitylation motif	Rabin and Hauser, 2005; Stirling et al., 2006
SopE, BopE	GAG(AT)	Catalytic loop motif essential for guanine nucleotide exchange	Schlumberger et al., 2003
AvrPphB	GDK	Autoproteolytic cleavage motif	Downen et al., 2009
SopA, IpaH, SspH1	L...TC, C.D	E3 ubiquitin ligase motif	Zhang et al., 2006; Rohde et al., 2007
VirF	LP... ..L	F-box domain motif, mediates protein-protein interactions	Tzfira et al., 2004
VopL, VopF	[^R]..((.[ILMV])((.[ILMV].))[^P][^P][ILVM].{4,7}L((.[KR].)([NK])([VATIGS], [R]..[ILVMF][ILMV][^P][ILVM].{4,7}L((.[KR].)([NK])([VATI]	WH2-domain motif	Liverman et al., 2007; Dinkel et al., 2012
SpvC, OspF, VirA	[KR]{0,2}[KR].{0,2}[KR].{2,4}[ILVM].[ILVF]	D motif, Docking motif required for specific binding to MAPKs	Zhu et al., 2007; Dinkel et al., 2012
IpaA	L..AA..VA..V..LI..A.	Vinculin binding domain motif	Hamiaux et al., 2006
ExoS, ExoT	LLDALDLA	FAS (14-3-3 protein) binding motif, mediates activation of the ADPRT domain	Sun et al., 2004; Dean, 2011
EspF	[RKY]..P.P, P.P.[KR],...[PV]..P, KP.[QK]. ...	SH3 binding motif	Alto et al., 2007; Dinkel et al., 2012
Map, NleH1, EspI (NleA)	...[ST].[ACVILF]\$, ... [VLIFY].[ACVILF]\$,...[DE].[ACVILF]\$ (EspI*)	C-terminal PDZ1 binding motif	Lee et al., 2008; Martinez et al., 2010; Dinkel et al., 2012
TOXIN			
Listeriolysin O (LLO)	PPASP*	PEST-motif, involved in phagosomal escape of bacteria in infected cells	Lety et al., 2001
OTHER			
VirD4, VirB11, VirB4, SecA	G...GK[TS]*	Walker A motif, nucleotide-binding motif	Sato et al., 1996; Atmakuri et al., 2004
SecA	[RK]...G...L[VILFWYMC]{4,4}D	Walker B, nucleotide binding motif	Sato et al., 1996
MsbA, PiaA, PiuA	LSGGQ (PiaA*, PiuA*)	ABC-motif, ATP binding cassette transporter motif	Garmory and Titball, 2004; Buchaklian and Klug, 2006
EsxA, EsxB, esat6	W.G	W.G motif helps to create a shallow cleft structure and may represent a peptide recognition feature by which cargo proteins are acquired for transport	Burts et al., 2005; Sundaramoorthy et al., 2008

*Proven role in virulence. Bold, a non-eukaryotic motif. "^" start of the protein or if in the middle of the motif sequence states which amino acids are excluded in the position, "\$" end of the protein, "." any amino acid, {} defines the length of a range in the motif sequence, [] defines which amino acids can occur at a given motif position, () marks positions of specific interest e.g., covalent modification or is used to group parts of the expression. Motif table modified from Dean (2011), additional motifs added from the literature.

Table 2 | Functional search terms used to retrieve and download protein sequences from virulence factor database MvirDBbrowser tool.

Virulence protein group	Number of sequences	Number of unrelated proteins
Adherence	749	181
Capsule	332	57
Chemotaxis	192	18
Effector	111	27
Endotoxin	66	27
Enzyme	647	121
Exotoxin	92	12
Lipoprotein	463	70
Motility	86	23
Siderophore	150	43
Type III secretion system	571	75
Type IV secretion system	181	38

Drawgram tree drawing program of Phylip package (version 3.695). Branches were colored according to the following scheme: Purple, High GC Gram+ bacteria; Blue, Firmicutes; Yellow, α -proteobacteria; Light Brown, β -proteobacteria; Dark Brown, ϵ -proteobacteria; Green, γ -proteobacteria (non-enterobacteria); Red, γ -proteobacteria (enterobacteria); Black, others (CFB).

RESULTS

Our objective was to discover novel SLiMs in non-homologous bacterial proteins with similar roles in virulence that may have functional importance in pathogenesis, and thus have potential to be developed into antimicrobial peptides or drugs. Our analysis returned both previously characterized and novel motifs in several different functional categories indicating the suitability of SLiMfinder for the analysis of bacterial sequence data as well as eukaryotic data. We focused on 12 groups of bacterial proteins with predefined roles in pathogenicity (Table 2). SLiMfinder identified numerous motifs among these proteins. Table 3 lists those with a p -value (Sig) less than 0.05. Bonferroni correction for significance with 12 search datasets would suggest that motifs with a Sig value of less than 0.004 are significant. Since pathogenesis proteins from bacteria often interact with host protein components, we examined whether any of the identified motifs showed similarity to known eukaryotic linear motifs, using the Comparimotif tool. However, we did not find any convincing similarities, in spite of the known occurrence of eukaryotic motifs in bacterial effector proteins. We also investigated if any of the motifs were known prokaryotic motifs identified in the literature.

KNOWN MOTIFS

Three of the motifs highlighted by SLiMfinder were previously known bacterial motifs. The most significant of these was the well-characterized prokaryotic N-terminal lipid modification [LVI][ASTVI][GAS]C motif that has been previously shown to be essential for the anchoring of bacterial proteins to the membrane surface (Braun and Rehn, 1969; Babu et al., 2006). The square brackets enclose alternative amino acids which are possible at that

position in the motif. This motif is present in a wide range of proteins across Gram-positive and Gram-negative bacteria and is a clear example of a motif that has convergently evolved in many unrelated proteins. It was found in numerous configurations in the lipoprotein dataset of which seven are listed in Table 3. This “lipobox” motif sequence is located at the C-terminal end of the signal peptide and the lipid-modifiable cysteine (+1 position) is invariant (Juncker et al., 2003). Lipid modification of this cysteine residue (*N*-acyl-S-diacylglyceryl-Cys) has been found to be an essential, ubiquitous, and unique bacterial post-translational modification. Such a modification allows anchoring of even highly hydrophilic proteins to the membrane surface leaving the rest of the protein to carry out a variety of relevant functions in the aqueous or aqueous-membrane interface (Juncker et al., 2003; Babu et al., 2006). Bacterial lipoproteins affect a wide range of mechanisms in virulence. They have been shown to play key roles in adhesion to host cells and in translocation of virulence factors into host cells (Kovacs-Simon et al., 2011). Furthermore, they are potent inducers of host inflammatory responses.

The second known motif identified was an N-terminal \wedge MK. $\{0,2\}$ K motif present in several search categories in varying configurations (Table 3, Adherence, Capsule, Enzyme, Lipoprotein, Siderophore, and Type IV secretion system). This motif representation indicates that the second K (lysine) may lay 0, 1, or 2 residues after the K that follows the initiator methionine. The “ \wedge ” symbol indicates the start of the protein, which is treated as a distinct character in motif discovery. SLiMfinder omits the M from the returned motif resulting in \wedge .K. $\{0,2\}$ K representation, since initiator methionines were deliberately masked out to avoid returning motifs reliant simply on the strong enrichment of M at the start of proteins. The \wedge MK. $\{0,2\}$ K motif is commonly found in bacterial signal peptides both in proteins that are targeted to the membrane and in secreted proteins (Juncker et al., 2003; Bagos et al., 2008). Both of the known motifs are presented as regular expressions in Figure 1, which provides some information on additional contextual preferences beyond the simple motif description. Signal peptides in bacteria are mainly divided into the secretory signal peptides that are cleaved by Signal Peptidase I and those cleaved by Signal Peptidase II which characterize the membrane-bound lipoproteins (Juncker et al., 2003; Bagos et al., 2008). The signal peptides in both classes of proteins in Gram-positive and Gram-negative bacteria are quite similar, sharing the N-terminal region which is characterized by presence of the positive amino acids at the start of the protein, as well as the preference for hydrophobic residues further along the signal peptide.

The third previously characterized bacterial motif returned in our analysis is the C-terminal KK\$ motif (where \$ indicates the end of the protein, and is treated as a distinct character in motif discovery) found in adherence and lipoprotein datasets (Table 3; Figure 1). This motif has been shown to play a role in plasminogen binding in *S. pyogenes* and *S. pneumoniae* α -enolase (Bergmann et al., 2003; Derbise et al., 2004; Itzek et al., 2010). Binding of plasminogen by α -enolase and its subsequent activation has been demonstrated to promote invasion of pathogenic bacteria and therefore represents an important determinant of virulence in invasive infection (Bergmann et al., 2003). Moreover, KK motifs close to the C-terminus are present in a family of

Table 3 | Significant motifs returned by SLiMfinder in each dataset (where probability <0.05).

Virulence protein group	Number of sequences	Motif pattern	Motif name	Information content	Occurrences	Unrelated proteins	Probability (Sig value)	References
(A) PREVIOUSLY DESCRIBED MOTIFS								
Adherence	749	^..K.{0,2}K	Signal peptide motif	3	151	47	0.00E+00	Juncker et al., 2003; Bagos et al., 2008
		^.{1,2}K.{1,2}L	Signal peptide motif	3	102	35	1.15E-10	Juncker et al., 2003; Bagos et al., 2008
		^.{1,2}K.{0,2}I	Signal peptide motif	3	91	34	1.41E-08	Juncker et al., 2003; Bagos et al., 2008
		^..K.{1}L	Signal peptide motif	2.77	58	21	5.80E-07	Juncker et al., 2003; Bagos et al., 2008
		^..K..S	Signal peptide motif	3	24	12	0.002	Juncker et al., 2003; Bagos et al., 2008
		KK\$	C-terminal KK	3	23	11	0.034	Bergmann et al., 2003; Itzek et al., 2010
Capsule	332	^.{1,2}K.{0,1}I	Signal peptide motif	3	55	18	6.96E-06	Juncker et al., 2003; Bagos et al., 2008
		^.{1,2}K.{0,1}..V	Signal peptide motif	4	16	7	0.002	Juncker et al., 2003; Bagos et al., 2008
		^.{1,2}K.{0,2}I	Signal peptide motif	3	59	18	0.003	Juncker et al., 2003; Bagos et al., 2008
		^..K.{1}LV	Signal peptide motif	2.63	39	13	0.003	Juncker et al., 2003; Bagos et al., 2008
		^.{1,2}K.{0,2}K	Signal peptide motif	3	52	17	0.006	Juncker et al., 2003; Bagos et al., 2008
Chemotaxis	192	–	–	–	–	–	–	–
Effector	111	–	–	–	–	–	–	–
Endotoxin	66	–	–	–	–	–	–	–
Enzyme	647	^.{1,2}KK	Signal peptide motif	3	27	14	9.52E-05	Juncker et al., 2003; Bagos et al., 2008
		^..K.{0,2}I	Signal peptide motif	3	30	13	0.042	Juncker et al., 2003; Bagos et al., 2008
Exotoxin	92	–	–	–	–	–	–	–
Lipoprotein	463	L.[AG]C[AGS]	Lipobox	3.4	78	30	0.00E+00	Braun and Rehn, 1969; Babu et al., 2006
		[FLV].L.[AG]C	Lipobox	3.4	136	24	0.00E+00	Braun and Rehn, 1969; Babu et al., 2006
		[1]LV.[AGS]C	Lipobox	2.27	370	53	0.00E+00	Braun and Rehn, 1969; Babu et al., 2006
		[AGS]C[AGS]	Lipobox	2.27	285	50	7.22E-15	Braun and Rehn, 1969; Babu et al., 2006
		^.{1,2}K.{0,2}K	Signal peptide motif	3	109	27	1.27E-09	Juncker et al., 2003; Bagos et al., 2008

(Continued)

Table 3 | Continued

Virulence protein group	Number of sequences	Motif pattern	Motif name	Information content	Occurrences	Unrelated proteins	Probability (Sig value)	References
		L.{1,2}GC.{0,1}A	Lipobox	4	41	15	1.65E-09	Braun and Rehn, 1969; Babu et al., 2006
		A.{0,2}L..C.{0,2}S	Lipobox	4	68	19	3.02E-09	Braun and Rehn, 1969; Babu et al., 2006
		^..K..[ILV]	Signal peptide motif	2.63	66	19	4.36E-09	Juncker et al., 2003; Bagos et al., 2008
		^..K..[FLV]	Signal peptide motif	2.63	65	17	1.07E-07	Juncker et al., 2003; Bagos et al., 2008
		[ILV]..C.[AGS]	Lipobox	2.27	217	36	4.11E-06	Braun and Rehn, 1969; Babu et al., 2006
		KK\$	C-terminal KK	3	22	9	0.016	Bergmann et al., 2003; Itzek et al., 2010
Motility	86	–	–	–	–	–	–	–
Siderophore	150	^[KR]I	Signal peptide motif	2.77	12	7	0.024	Juncker et al., 2003; Bagos et al., 2008
Type III secretion	571	–	–	–	–	–	–	–
Type IV secretion	181	^..K[KR]	Signal peptide motif	2.77	29	9	0.003	Juncker et al., 2003; Bagos et al., 2008
		^..K..[FIL]	Signal peptide motif	2.63	27	10	0.025	Juncker et al., 2003; Bagos et al., 2008
(B) NOVEL MOTIFS								
Adherence	749	LPG.Y		4	37	12	0.012	
Capsule	332	G.S..M.L		4	15	7	0.029	
Chemotaxis	192	E..Q..I[AG].I		4.77	22	5	0.004	
		E..Q..I[IV]..I		3.77	24	7	0.02	
Effector	111	^..I..{0,1}N		3	21	6	0.012	
		[LV].PY		2.77	46	11	0.042	
		^..I[ST]		2.77	17	6	0.049	
Endotoxin	66	–	–	–	–	–	–	–
Enzyme	647	A.I.PVL		5	14	7	0.019	
		VSIL.S		5	11	7	0.049	
Exotoxin	92	–	–	–	–	–	–	–
Lipoprotein	463	ML..C		3	14	7	0.017	
Motility	86	–	–	–	–	–	–	–
Siderophore	150	I.K..G		3	28	17	0.044	
		GYP.TP		5	5	4	0.052	

(Continued)

Table 3 | Continued

Virulence protein group	Number of sequences	Motif pattern	Motif name	Information content	Occurrences	Unrelated proteins	Probability (Sig value)	References
Type III secretion	571	–		–	–	–	–	
Type IV secretion	181	G[AG].\$		2.77	19	9	2.64E–04	

Where very similar motifs are returned for a protein group, only a representative motif is shown.

Information content (Edwards et al., 2007), “^” start of the protein, “\$” end of the protein, “.” any amino acid, {} defines the range of a repeat in the motif sequence, [] defines which amino acids can occur at a given motif position.

Italic font is used when Probability (Sig-value) is higher than the 0.05 confidence level.

Shigella flexneri glucosyl transferases (Gtr) that are integral membrane proteins embedded within the cytoplasmic membrane. These glucosyl transferases contribute to the altering of the structure of the bacterial surface lipopolysaccharide (LPS) O-antigen along with O-acetyltransferase (Lehane et al., 2005; Ramiscal et al., 2010). The KK motif has been shown to be essential for the activity of Gtrs. However, Ramiscal et al. showed that the KK motif in a recently identified GtrIc is not critical for its activity (Ramiscal et al., 2010). We hypothesize that the KK\$ motif instances identified here in diverse proteins may play an adhesive role similar to the plasminogen binding instances in α -enolase. We note that plants have a KK\$ variant (Gidda et al., 2009) of a known eukaryotic cytoplasmically exposed endoplasmic reticulum (ER) localization motif KKxx\$ found in mammals, yeast and plants (Nilsson et al., 1989; Jackson et al., 1990; Contreras et al., 2004). It is therefore conceivable that the bacterial KK\$ motif could in some proteins direct invading proteins to certain parts of the eukaryotic host cell. However, we do not think this is very plausible, since the enrichment of KK\$ motifs spans many known bacterial lipoproteins (Table 3) which seem unlikely to migrate to this host cell location.

NOVEL MOTIFS

The most significant novel motif (p -value 0.0003) discovered is a C-terminal G[AG].\$ motif in the type IV secretion system dataset. The full list of unrelated proteins containing the G[AG].\$ motif is represented in Table 4. The MEME regular expression pattern of the motif in these proteins is described in Figure 1. Four of the nine unrelated proteins containing this motif appear to be identified equivalents of the type IV secretion system components in the well-studied *Agrobacterium tumefaciens*: VirB4, VirB8, VirB11, and VirB7 [TrwH has 59% identity with VirB7 family (Patey et al., 2006)]. VirB4 and VirB11 are known energetic components of the type IV secretion system in *A. tumefaciens*. Both of these proteins are membrane associated NTPases on the inner membrane (Tegtmeier et al., 2011). VirB8 on the other hand, is an essential inner membrane component of type IV secretion systems that is believed to form a homodimer and has been shown to be of importance for complex stability in *A. tumefaciens* (Sivanesan and Baron, 2011). The VirB7 is an outer membrane lipoprotein that localizes exocellularly and associates with the type IV secretion system pilus. Both VirB7 lipid modification and

disulfide cross-linking have been shown to be important for pilus assembly (Sagulenko et al., 2001). The *Helicobacter pylori* protein Cag7 that is among the proteins containing the C-terminal G[AG].\$ motif has previously been proposed to be a transmembrane protein that is associated with the pilus (Rohde et al., 2003; Tegtmeier et al., 2011). At least five of the nine unrelated proteins containing the G[AG].\$ motif seem to be associated with the bacterial membranes and it is thus possible that this motif would be involved in the targeting and/or attachment of these proteins into the bacterial membranes. However, since the motif has been specifically identified within type IV secretion proteins, it is more likely that the motif facilitates interaction with a component of the type IV secretion system itself. We inspected the distribution of the motif across effector proteins (Figure 2) and noted that there are typically one or none per species, suggesting that the motif is not itself enriched strongly among effector proteins themselves.

Other novel motifs discovered are summarized in Table 3 and in Figure 1. Their significance is in the range between that for the nominal significance level ($p < 0.05$) and the Bonferroni adjusted significance level ($p < 0.004$). While it is likely that a number of these motifs are genuine, a few may be false positives. The LP.G.Y motif found in the adherence dataset superficially resembles a Gram-positive bacteria cell wall anchoring LP.TG motif. Cleavage between the Thr and Gly by sortase or a related enzyme leads to covalent anchoring of the new C-terminal Thr to the cell wall (Navarre and Schneewind, 1994; Gaspar et al., 2005). Cell wall-anchored surface proteins of Gram-positive pathogens play important roles during the establishment of many infectious diseases. While it could be hypothesized that the LP.G.Y motif is similarly involved in the anchoring of bacterial proteins to the cell surface, there are two lines of evidence that argue against this. Firstly, there is no enrichment for T or similar amino acids between P and G in the instances of the motif returned (Figure 1). Secondly, this motif is present both in Gram-positive and Gram-negative bacterial proteins in our study. Accordingly, we consider LP.G.Y a potential novel motif involved in bacterial adhesion through an unidentified mechanism.

REPEATED MOTIFS

While SLiMfinder looks for motifs which recur one or more times in a number of independent proteins, it is of biological

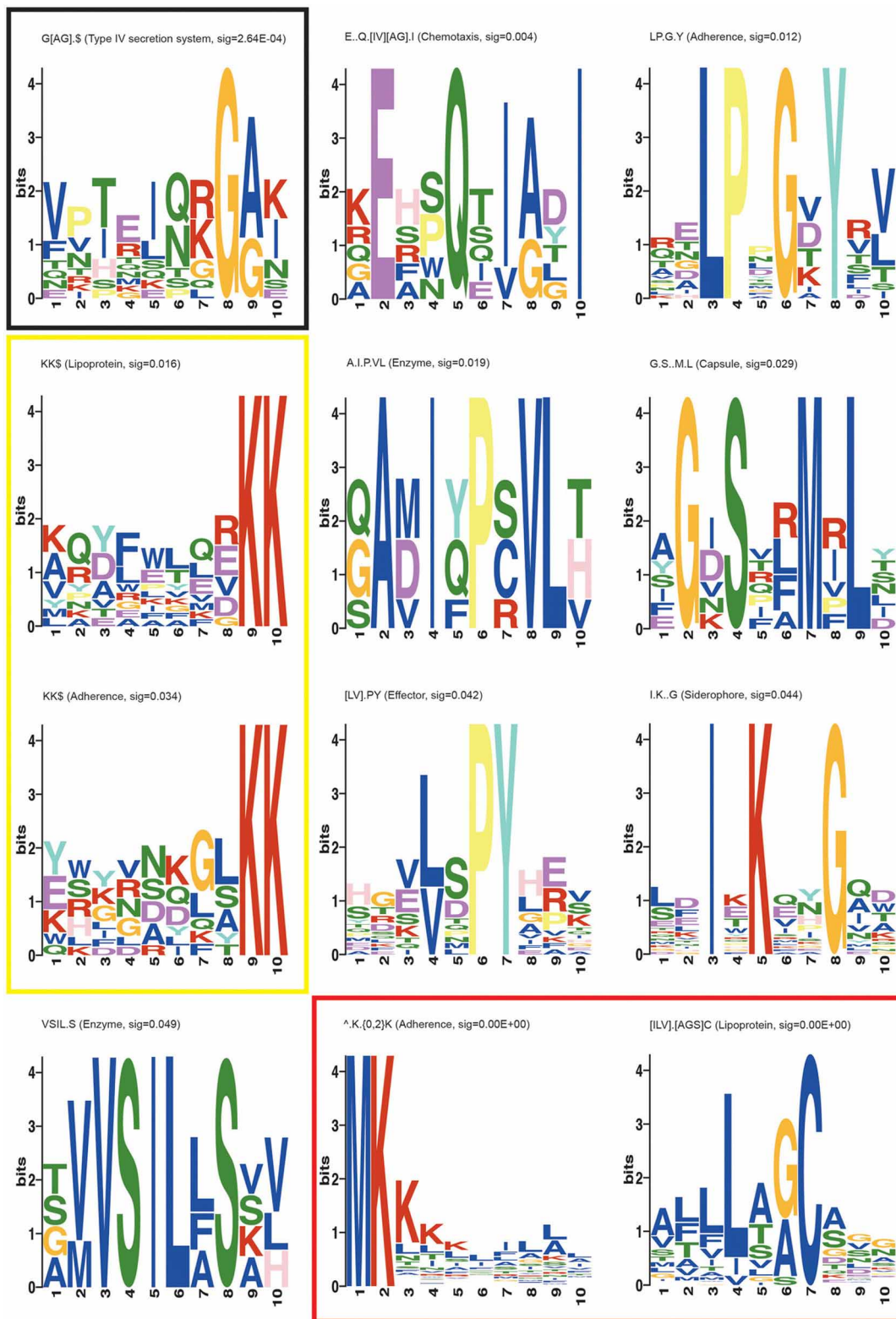


FIGURE 1 | MEME suite motif logos of the novel and known motifs returned in the SLiMfinder analysis. Each position in the motif is represented as a stack of letters. The total height of the stack is the “information content” of that position in the motif in bits. The height of the individual letters in a stack is the probability

of the letter at that position multiplied by the total information content of the stack. Black box: the most significant novel motif G[AG].\$, Yellow box: KKS motifs found in Adherence and Lipoprotein datasets, Red box: Known bacterial motifs $\hat{.}K.(0,2)K$ and [ILV].[AGS]C.

Table 4 | List of proteins containing G[AG].\$ and KK\$ motifs.

Pattern	Protein group	P-value (Sig)	Match	No. unrelated proteins	Description
G[AG].\$	Type IV secretion	2.64E-04	GGN	9	virB11 protein homolog 9992 YP_034060 49476019 8040 [<i>Bartonella henselae</i> str. Houston-1]
			GAK		virB4 10558 NP_863348 32469876 8343 VirB4 [<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> 81-176]
			GAK		virB8 10560 NP_863298 32469826 8344 VirB8 [<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> 81-176]
			GAE		cag pathogenicity island protein (cag11) 10866 NP_207327 15645157 8497 [<i>Helicobacter pylori</i> 26695]
			GGK		trwF protein 9938 YP_034270 49476229 8013 [<i>Bartonella henselae</i> str. Houston-1]
			GAI		trwH2 hypothetical protein BH15720 9944 YP_034268 49476227 8016 [<i>Bartonella henselae</i> str. Houston-1]
			GAS		cag pathogenicity island protein (cag25) 10894 NP_207342 15645172 8511 [<i>Helicobacter pylori</i> 26695]
			GGN		Putative type IV secretion system protein 41299 NP_790379 NP_790379.1 18355 [<i>Pseudomonas syringae</i> pv. tomato str. DC3000]
			GAI		trwH1 hypothetical protein BH15690 9942 YP_034265 49476224 8015 [<i>Bartonella henselae</i> str. Houston-1]
			GGN		cag pathogenicity island protein (cag7) 10904 NP_207323 15645153 8516 [<i>Helicobacter pylori</i> 26695]
KK\$	Adherence	0.034	KK	11	hmw2C putative accessory processing protein [<i>Haemophilus influenzae</i>] 2847 AAA20526 482843 2554
			KK		kpsC polysaccharide modification protein [<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> NCTC 11168] 10454 NP_282555 15792732 8291
			KK		ica operon transcriptional regulator [<i>Staphylococcus aureus</i> subsp. <i>aureus</i> MW2] 3137 NP_647402 21284314 2699
			KK		pavA adherence and virulence protein A [<i>Streptococcus agalactiae</i> 2603V/R] 9719 NP_688199 22537348 7896
			KK		Type 4 fimbrial biogenesis protein PilO [<i>Pseudomonas aeruginosa</i> PAO1] 8738 NP_253729 15600235 7426
			KK		Type 4 fimbrial biogenesis protein PilN [<i>Pseudomonas aeruginosa</i> PAO1] 8740 NP_253730 15600236 7427
			KK		Putative collagen binding protein [<i>Streptococcus pyogenes</i> MGAS315] SpyM3_0098 9709 NP_663902 21909634 7891
			KK		oapA opacity associated protein [<i>Haemophilus influenzae</i> Rd KW20] 2865 NP_438494 16272282 2563
			KK		neuC1 putative N-acetylglucosamine-6-phosphate 2-epimerase/N-acetylglucosamine-6-phosphatase [<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> NCTC 11168] 10349 NP_282290 15792467 8218
			KK		waaE D,D-heptose 1-phosphate adenosyltransferase/7-phosphate kinase [<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> NCTC 11168] 10396 NP_282297 15792474 8262
			KK		hmw1C putative accessory processing protein [<i>Haemophilus influenzae</i>] 2841 AAA20529 475773 2551
			KK		cytotoxin [<i>Escherichia coli</i> O157:H7] 11176 AAC70163 3822209 8652
			KK\$		Lipoprotein
KK	ylpB/yscJ needle complex inner membrane lipoprotein [<i>Yersinia pestis</i> CO92] 20866 NP_395193 16082747 13354				
KK	Yop proteins translocation lipoprotein J OS = <i>Yersinia enterocolitica</i> GN = yscJ PE = 2 SV = 1 18600 AltName: Full = Lipoprotein ylpB; Flags: Precursor; Q01251 11895				

(Continued)

Table 4 | Continued

Pattern	Protein group	P-value (Sig)	Match	No. unrelated proteins	Description
			KK		Lipoprotein [<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18] 44580 NP_455466 NP_455466.1 19996
			KK		LPP20 lipoprotein OS = <i>Helicobacter pylori</i> GN = lpp20 PE = 1 SV = 1 17433 P0A0V0 10728
			KK		Outer membrane factor of efflux pump [<i>Escherichia coli</i> str. K-12 substr. MG1655] 58055 NP_418504 NP_418504 24053
			KK		Lipoprotein, putative [<i>Enterococcus faecalis</i> V583] 9889 NP_816134 29376980 7982
			KK		Iron transport lipoprotein SirF [<i>Staphylococcus aureus</i> subsp. <i>aureus</i> Mu50] 8986 NP_371657 15924123 7550
			KK		Export protein prsA cytoplasmic membrane protein, protein folding 3480 GBAA2336 prsA-3 GBAA2336 2870
			KK		LPP20 lipoprotein OS = <i>Helicobacter pylori</i> J99 GN = lpp20 PE = 3 SV = 1 17432 P0A0V1
			KK		Yop proteins translocation lipoprotein J OS = <i>Yersinia pseudotuberculosis</i> GN = yscJ PE = 2 SV = 1 18863 AltName: Lipoprotein ylpB P69973 12158
			KK		YaeC family lipoprotein [<i>Enterococcus faecalis</i> V583] [<i>Enterococcus faecalis</i> V583] 39473 NP_815743 NP_815743.1 17442
			KK		Major outer membrane lipoprotein OS = <i>Yersinia pestis</i> GN = lpp PE = 3 SV = 1 24369 Q8ZDZ6 15171
			KK		Putative lipoprotein [<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. Ty2] 44956 NP_805720 NP_805720.1 20184

interest when those motifs are themselves repeated within the proteins, for example, representing multiple adhesion sites. Accordingly, we investigated the frequency of repeats of the identified motifs. Duplicated motifs were found with between two and four copies in proteins. The lipoprotein lipid anchoring motif was found repeated three times in the protein HrpB3 of *Xanthomonas euvesicatoria* (instances LAGC, LALC, and LSAC). Among these motif instances, LAGC and LSAC are known lipid anchoring motifs (Klein et al., 2005; Konkel et al., 2010). The third instance may represent a true positive anchoring motif, a degenerate motif that is no longer functional or a false positive sequence that fulfills some other functional role in the protein. However, it is clear that the repetition of this well-known motif is in some cases biologically important for function. Thus, for novel motifs, repetition within as well as between proteins may be a potential further indication of important function. An example would be the threefold repetition of the “LP.G.Y” motif in the surface-anchored fimbrial subunit protein SpaG of *Corynebacterium diphtheriae*. This motif has a known structure in the collagen binding domain of *Staphylococcus aureus* (PDB entry 1D2P) (Deivanayagam et al., 2000). Collagen is itself a repetitive structure, occurring in many dense repeats in the host extracellular matrix. The repetition of this bacterial motif in this particular protein may indicate its potential role in making multiple contacts with collagen. However, other instances of the motif detected by SLiMfinder only occurred once in each protein, suggesting that a single copy may be sufficient.

DISTRIBUTION OF SHORT LINEAR MOTIFS ACROSS EFFECTOR PROTEINS OF DIFFERENT SPECIES

We visualized the cross-species distribution of the SLiMfinder identified novel motifs (see **Table 3B**) among the annotated effector proteins of other species. The species were chosen to include those present in the MvirDB database that contributed motifs to the discovery, in order to display a varied set of species that could be visualized with ease. It is likely that they also exist in other organisms, although distinguishing true and false positives is not possible computationally. The visualization is normalized to correct for the fact that some species have very few proteins and that some motifs have very few instances. The total number of UPCs are indicated in brackets before each bacterial species as well as the total incidences of a motif in UPCs across all bacterial species indicated before each motif regular expression. The novel SLiMfinder identified effector protein motifs $\wedge..I.\{0,1\}N$, $[LV].PY$ and $\wedge..I[ST]$ are found among the effector proteins of many species, but are absent in those of many other species, including those with a reasonable number of annotated effector proteins (**Figure 2**).

We also looked at the distribution of known motifs (see **Tables 1, 3A**) across species (**Figure 3**). While some effector motifs (see second section of **Table 1**) show a wide phylogenetic distribution, others are restricted to only a few species, such as the G.LR...T motif involved in Rho GAP function. The nuclear localization signals (at the bottom of **Figure 3**) show a relatively restricted distribution. The WEK[IM]..FF late endocytic compartment localization motif is restricted to the genus

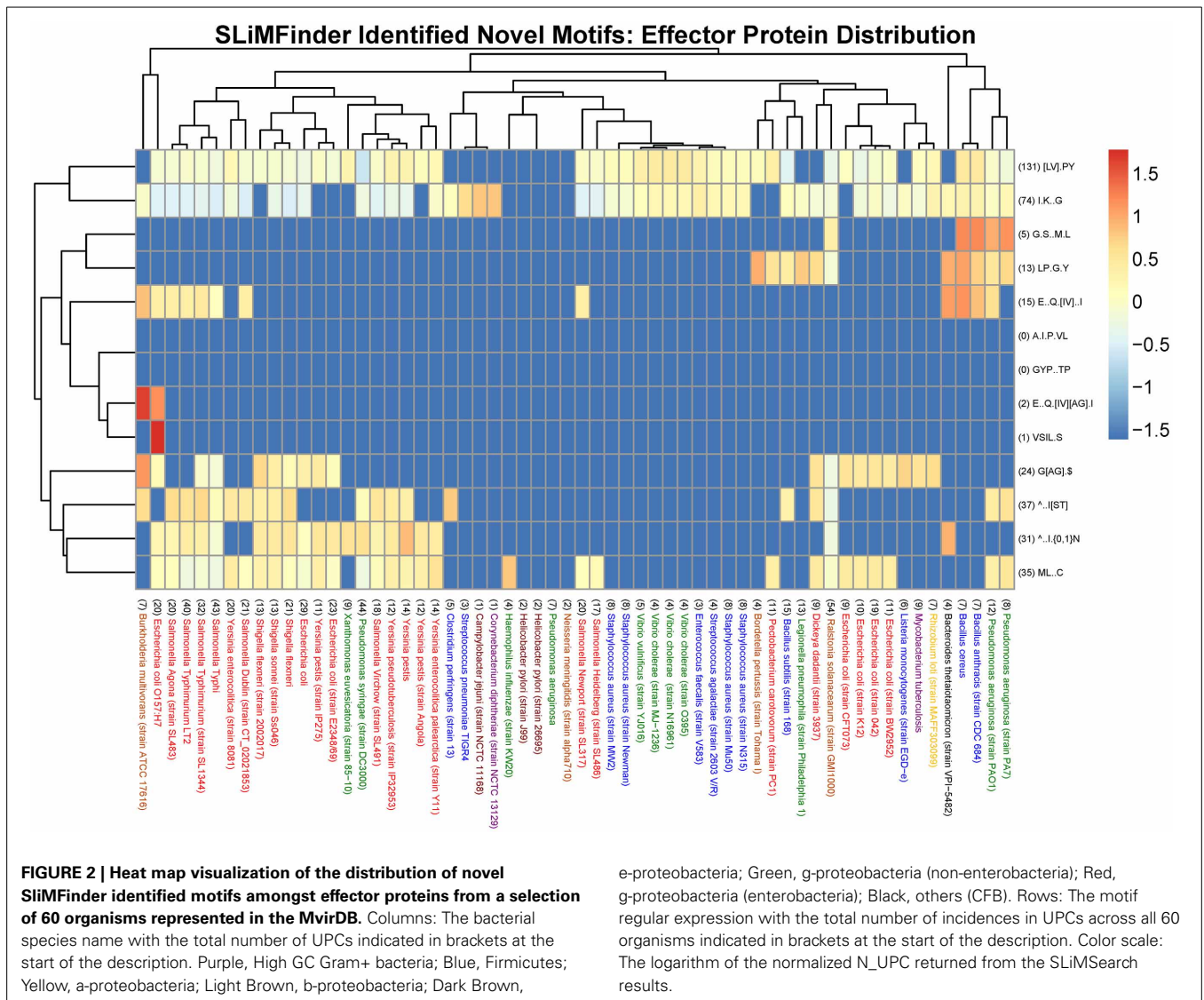


FIGURE 2 | Heat map visualization of the distribution of novel SLiMfinder identified motifs amongst effector proteins from a selection of 60 organisms represented in the MvirDB. Columns: The bacterial species name with the total number of UPCs indicated in brackets at the start of the description. Purple, High GC Gram+ bacteria; Blue, Firmicutes; Yellow, a-proteobacteria; Light Brown, b-proteobacteria; Dark Brown, e-proteobacteria; Green, g-proteobacteria (non-enterobacteria); Red, g-proteobacteria (enterobacteria); Black, others (CFB). Rows: The motif regular expression with the total number of incidences in UPCs across all 60 organisms indicated in brackets at the start of the description. Color scale: The logarithm of the normalized N_UPC returned from the SLiMSearch results.

Salmonella. While the ubiquitin ligase motifs L...TC and C.D are found in more than 71 and 199 instances respectively across the dataset whereas a number of species lack one or both of these motifs. The two SH3 binding motifs [RKY]..P.P and P.P.[KR] also show a restricted distribution. Similarly, the two PDZ binding motifs,...[ST].[ACVILF]\$ and ...[VLIFY].[ACVILF]\$ show a restricted distribution. Bacterial effector proteins may under certain circumstances be under negative selection to avoid motifs that bind to common domains in the host such as PDZ and SH3 domains.

It can be seen that strains of a species often have very similar motif distributions (Figures 2–4). There is a weak but not convincing trend (Figure 3) for the known motif distribution among effector proteins of the Firmicutes (Blue) to group together, relative the gamma-proteobacteria (Red and Green). While the Group 2 Bacillus species, anthracis and cereus, cluster together (Figure 3), many sets of closely related species (Figure 4) do not show particularly close relationships in terms of motif distribution. This may result from two factors: firstly, motifs are

highly dynamic during evolution, and secondly, factors that play a role in pathogenicity also evolve very fast. It is also difficult to compare rare vs. common motifs, since rare ones may be missed simply because of variation among proteins in the definition of effector proteins, while common motifs may be dominated by false positives that obscure the biologically relevant signals.

DISCUSSION

We believe that SLiMs are one potential class of new antimicrobial substances for the development of antimicrobial peptides and drugs. While they may lack the potency of antimicrobial peptides that damage the bacterial membrane, they may have other benefits. In particular, those that mimic peptide components of uniquely prokaryotic motifs are likely to have less off-target effects. The value of developing such therapeutic approaches depends on the range of species likely to be affected by the peptide therapeutic. While targeting eukaryotic peptides mimicked by prokaryote effector proteins provides a potential line of



FIGURE 3 | Heat map visualization of the distribution of known virulence motifs amongst effector proteins from a selection of 60 organisms represented in the MvirDB. Columns: The bacterial species name with the total number of UPCs indicated in brackets at the start of the description. Purple, High GC Gram+ bacteria; Blue, Firmicutes; Yellow, a-proteobacteria; Light Brown, b-proteobacteria; Dark Brown,

e-proteobacteria; Green, g-proteobacteria (non-enterobacteria); Red, g-proteobacteria (enterobacteria); Black, others (CFB). Rows: The motif regular expression with the total number of incidences in UPCs across all 60 organisms indicated in brackets at the start of the description. Color scale: The logarithm of the normalized N_UPC returned from the SLiMSearch results.

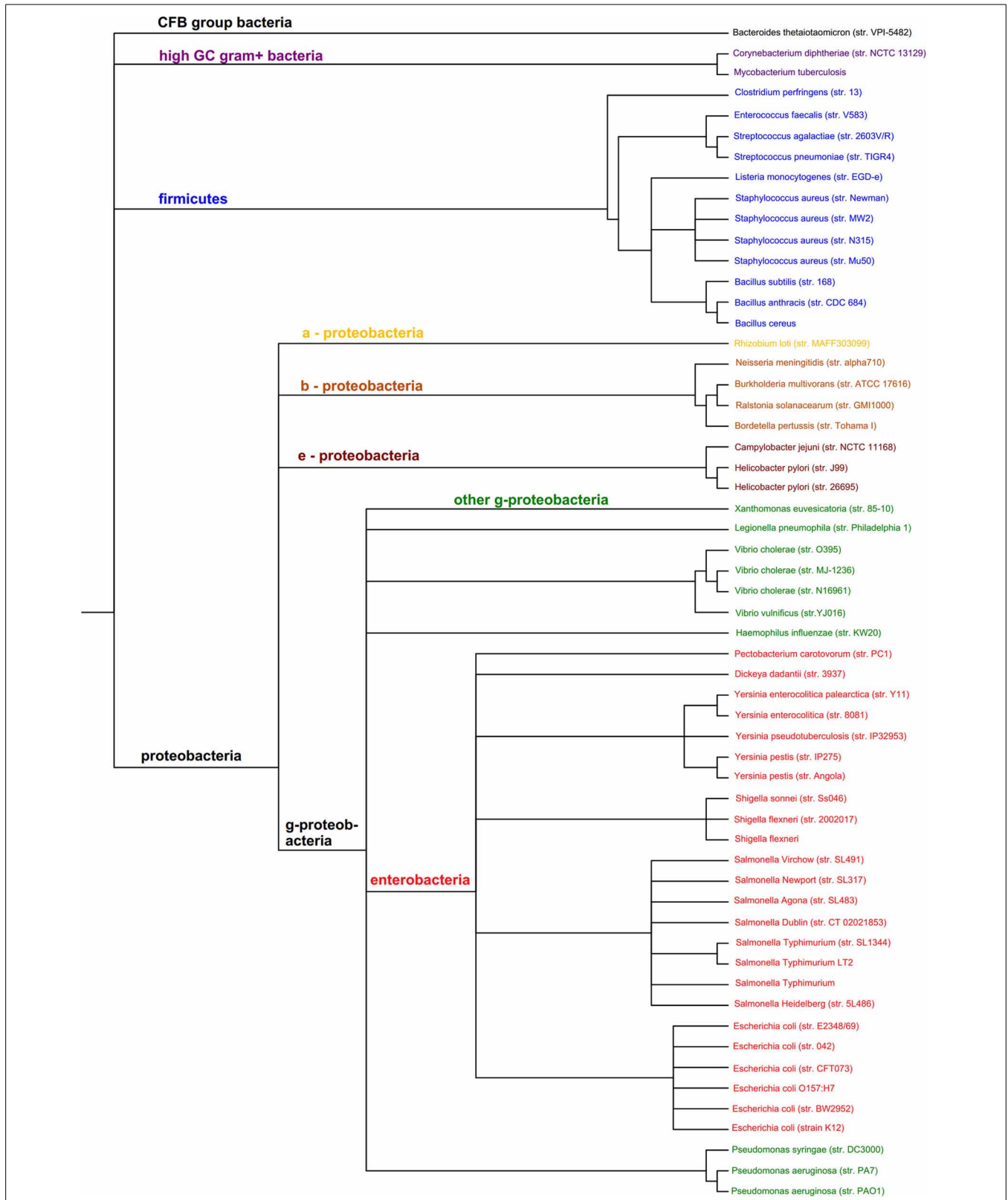


FIGURE 4 | Phylogenetic tree of the 60 organisms used to assess the distribution of prokaryotic protein motifs in Figures 2, 3. Purple, High GC Gram+ bacteria; Blue, Firmicutes; Yellow, a-proteobacteria; Light Brown,

b-proteobacteria; Dark Brown, e-proteobacteria; Green, g-proteobacteria (non-enterobacteria); Red, g-proteobacteria (enterobacteria); Black, others (CFB).

attack, the evolutionary plasticity of such motifs in both bacteria (**Figure 3**) and in hosts (Nedeva and Russell, 2005) suggest that bacteria can rapidly evolve alternative effector strategies to replace one targeted host component with another. Nevertheless, where such drugs are developed for other indications in treating non-infectious disease, they may also have an impact on bacterial pathogenesis and would certainly be worth investigating. This problem of evolutionary evasion by pathogens is also relevant, however, to many adhesion motifs. In order for peptide therapeutics to be more robust in the face of rapid evolution of pathogen resistance, they may need to target fundamental components of bacterial biology. Targeting aspects of the central machinery of bacterial Type IV secretion systems may be a good compromise between targeting a component that is central to pathogenicity, while not affecting the biology of advantageous bacteria in the host. In this respect, the G[AG].\$ motif identified in this study is a potential candidate worthy of further investigation. Some clues as to the function of this motif may be provided by the pattern of evolution. Presumably this motif has evolved in multiple components of the Type IV secretion system because of a selection pressure for these proteins to interact with some common factor. Identifying the common interaction partners of these proteins may help in pinpointing its potential functional role. In targeting such pathogenicity systems, the benefit of focusing on recurrent motifs is that they may be small enough interaction surfaces to be feasibly targeted by peptidomimetics, and important enough that it is difficult for the bacterial system to evolve resistance (Baron and Coombes, 2007; Paschos et al., 2011).

The shortlist of predicted motifs that we have generated provides a resource for researchers interested in the mechanisms of action of virulence factor proteins across a diverse range of bacterial species. The limitations of the list are well-illustrated by the fact that the motif discovery failed to rediscover the many mimicked eukaryotic motifs. This reflects not only the fact that some motifs have not evolved multiple times in unrelated proteins, but also the limitations in the datasets provided to the SLiMfinder approach. Ideally, datasets should have less than a 100 proteins which have clearly identified similar functions. The challenge is to group proteins according to function efficiently, since the annotation of protein function is highly variable, and frequently relies on computational predictions arising from homology rather than from direct experimentation. The bigger challenge is how to test and manipulate these motifs to provide insights into the mechanisms of action and to determine potential means of interrupting pathogenic processes. While mutagenesis studies can identify the key features of motif function, targeting of a motif may also be progressed by experimental use of bioactive peptides. However, identification of more potent peptidomimetic compounds that resemble such motifs will ideally need 3D models of the peptide regions in complex with their target interactors.

What, then, is the contribution that computational screening of novel motifs may play in the discovery of novel antimicrobial peptides? Firstly, it clearly will not identify all known motifs, since patterns of recurrent evolution or of strong sequence conservation are not seen for all antimicrobial peptides. Computational screens will also have some “false positives” in two senses: firstly,

statistical false positives where the motif arose simply by chance; and secondly, biological false positives where the motif that functions effectively within its biological context of a larger protein and that protein's complexes, but it will not function as a stand-alone synthetic peptide. This could reflect a lack of strong affinity for its targets or it could reflect an inability to be delivered to the appropriate context in the first place. Nevertheless, computational screens have the advantage that they can be performed on high throughput sequencing of organisms about which little else is known and for which biological screening by mutagenesis is painstaking or impossible. The advantage of computational prioritization is that it identifies a subset of peptides which are enriched for biologically active peptides. Clearly, the strategy we adopted here is only detecting a small fraction of known motifs, in part because of the stringent correction for statistical mismatches that could be false positives, but also because many motifs do not recur in known unrelated proteins that fall into the same functional class. Discovery for bioactive peptides could follow other strategies, including searches for evolutionary conservation (Davey et al., 2012a). However, pathogenicity factors frequently evolve rapidly, and so conservation may not be an effective signal. Bioactivity predictors based on biophysical properties within the peptide sequences are an alternative strategy (Dosztanyi et al., 2009; Thomas et al., 2010; Mooney et al., 2012, 2013). These have the disadvantage that there is no straightforward statistical approach available to determine likely false discovery rates, but are very valuable in prioritizing a list of peptides for further experimental characterization. Other computational approaches focus more on particular classes of antimicrobial peptides with a strong therapeutic potential, including ribosomal and non-ribosomal cyclic peptides (Prieto et al., 2012; Kedarisetti et al., 2014). While their computational screening methods have the benefit that they focus more strongly on peptides in classes of known therapeutic benefit, we believe that the computational screening approach we identified here complements their approaches, and widens the diversity of peptides for experimental investigation and validation.

ACKNOWLEDGMENTS

This work was funded by the Wellcome Trust Computational Infection Biology PhD Programme (supporting Heini Ruhanen and Daniel Hurley), by Science Foundation Ireland (grant no. 08-IN.1-B1864) and by the Irish Research Council Graduate Education Research Programme in Bioinformatics and Computational Biology (supporting Kevin T. O'Brien).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00004/abstract>

Table S1 | Protein sequence identifiers for each functional category downloaded from the MvirDB.

REFERENCES

- Alto, N. M., Shao, F., Lazar, C. S., Brost, R. L., Chua, G., Mattoo, S., et al. (2006). Identification of a bacterial type III effector family with G protein mimicry functions. *Cell* 124, 133–145. doi: 10.1016/j.cell.2005.10.031

- Alto, N. M., Weflen, A. W., Rardin, M. J., Yasar, D., Lazar, C. S., Tonikian, R., et al. (2007). The type III effector EspF coordinates membrane trafficking by the spatiotemporal activation of two eukaryotic signaling pathways. *J. Cell Biol.* 178, 1265–1278. doi: 10.1083/jcb.200705021
- Atmakuri, K., Cascales, E., and Christie, P. J. (2004). Energetic components VirD4, VirB11 and VirB4 mediate early DNA transfer reactions required for bacterial type IV secretion. *Mol. Microbiol.* 54, 1199–1211. doi: 10.1111/j.1365-2958.2004.04345.x
- Babu, M. M., Priya, M. L., Selvan, A. T., Madera, M., Gough, J., Aravind, L., et al. (2006). A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J. Bacteriol.* 188, 2761–2773. doi: 10.1128/JB.188.8.2761-2773.2006
- Bagos, P. G., Tsigros, K. D., Liakopoulos, T. D., and Hamodrakas, S. J. (2008). Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *J. Proteome Res.* 7, 5082–5093. doi: 10.1021/pr800162c
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Banerjee, A., Van Sorge, N. M., Sheen, T. R., Uchiyama, S., Mitchell, T. J., and Doran, K. S. (2010). Activation of brain endothelium by pneumococcal neuraminidase NanA promotes bacterial internalization. *Cell. Microbiol.* 12, 1576–1588. doi: 10.1111/j.1462-5822.2010.01490.x
- Baron, C., and Coombes, B. (2007). Targeting bacterial secretion systems: benefits of disarmament in the microcosm. *Infect. Disord. Drug Targets* 7, 19–27. doi: 10.2174/187152607780090685
- Bergmann, S., Wild, D., Diekmann, O., Frank, R., Bracht, D., Chhatwal, G. S., et al. (2003). Identification of a novel plasmin(ogen)-binding motif in surface displayed alpha-enolase of *Streptococcus pneumoniae*. *Mol. Microbiol.* 49, 411–423. doi: 10.1046/j.1365-2958.2003.03557.x
- Black, D. S., and Bliska, J. B. (2000). The RhoGAP activity of the *Yersinia pseudotuberculosis* cytotoxin YopE is required for antiphagocytic function and virulence. *Mol. Microbiol.* 37, 515–527. doi: 10.1046/j.1365-2958.2000.02021.x
- Boucrot, E., Beuzon, C. R., Holden, D. W., Gorvel, J. P., and Meresse, S. (2003). *Salmonella typhimurium* SifA effector protein requires its membrane-anchoring C-terminal hexapeptide for its biological function. *J. Biol. Chem.* 278, 14196–14202. doi: 10.1074/jbc.M207901200
- Braun, V., and Rehn, K. (1969). Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the *E. coli* cell wall. The specific effect of trypsin on the membrane structure. *Eur. J. Biochem.* 10, 426–438. doi: 10.1111/j.1432-1033.1969.tb00707.x
- Brown, N. F., Szeto, J., Jiang, X., Coombes, B. K., Finlay, B. B., and Brumell, J. H. (2006). Mutational analysis of *Salmonella* translocated effector members SifA and SopD2 reveals domains implicated in translocation, subcellular localization and function. *Microbiology* 152, 2323–2343. doi: 10.1099/mic.0.28995-0
- Buchaklian, A. H., and Klug, C. S. (2006). Characterization of the LSGGQ and H motifs from the *Escherichia coli* lipid A transporter MsbA. *Biochemistry* 45, 12539–12546. doi: 10.1021/bi060830a
- Burts, M. L., Williams, W. A., Debord, K., and Missiakas, D. M. (2005). EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1169–1174. doi: 10.1073/pnas.0405620102
- Conradi, J., Huber, S., Gaus, K., Mertink, F., Royo Gracia, S., Strijowski, U., et al. (2012a). Cyclic RGD peptides interfere with binding of the *Helicobacter pylori* protein CagL to integrins alphaVbeta3 and alpha5beta1. *Amino Acids* 43, 219–232. doi: 10.1007/s00726-011-1066-0
- Conradi, J., Tegtmeyer, N., Wozna, M., Wissbrock, M., Michalek, C., Gagell, C., et al. (2012b). An RGD helper sequence in CagL of *Helicobacter pylori* assists in interactions with integrins and injection of CagA. *Front. Cell. Infect. Microbiol.* 2:70. doi: 10.3389/fcimb.2012.00070
- Contreras, I., Ortiz-Zapater, E., and Aniento, F. (2004). Sorting signals in the cytosolic tail of membrane proteins involved in the interaction with plant ARF1 and coatomer. *Plant J.* 38, 685–698. doi: 10.1111/j.1365-313X.2004.02075.x
- Cornelis, G. R., and Van Gijsegem, F. (2000). Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* 54, 735–774. doi: 10.1146/annurev.micro.54.1.735
- Craik, D. J., Fairlie, D. P., Liras, S., and Price, D. (2013). The future of peptide-based drugs. *Chem. Biol. Drug Des.* 81, 136–147. doi: 10.1111/cbdd.12055
- Davey, N. E., Cowan, J. L., Shields, D. C., Gibson, T. J., Coldwell, M. J., and Edwards, R. J. (2012a). SLiMPrints: conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. *Nucleic Acids Res.* 40, 10628–10641. doi: 10.1093/nar/gks854
- Davey, N. E., Van Roey, K., Weatheritt, R. J., Toedt, G., Uyar, B., Altenberg, B., et al. (2012b). Attributes of short linear motifs. *Mol. Biosyst.* 8, 268–281. doi: 10.1039/c1mb05231d
- Davey, N. E., Haslam, N. J., Shields, D. C., and Edwards, R. J. (2010). SLiMFinder: a web server to find novel, significantly over-represented, short protein motifs. *Nucleic Acids Res.* 38, W534–W539. doi: 10.1093/nar/gkq440
- Davey, N. E., Haslam, N. J., Shields, D. C., and Edwards, R. J. (2011a). SLiMSearch 2.0: biological context for short linear motifs in proteins. *Nucleic Acids Res.* 39, W56–W60. doi: 10.1093/nar/gkr402
- Davey, N. E., Trave, G., and Gibson, T. J. (2011b). How viruses hijack cell regulation. *Trends Biochem. Sci.* 36, 159–169. doi: 10.1016/j.tibs.2010.10.002
- Dean, P. (2011). Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol. Rev.* 35, 1100–1125. doi: 10.1111/j.1574-6976.2011.00271.x
- Deivanayagam, C. C., Rich, R. L., Carson, M., Owens, R. T., Danthuluri, S., Bice, T., et al. (2000). Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8, 67–78. doi: 10.1016/S0969-2126(00)00081-2
- Derbise, A., Song, Y. P., Parikh, S., Fischetti, V. A., and Pancholi, V. (2004). Role of the C-terminal lysine residues of streptococcal surface enolase in Glu- and Lys-plasminogen-binding activities of group A streptococci. *Infect. Immun.* 72, 94–105. doi: 10.1128/IAI.72.1.94-105.2004
- Deslandes, L., Olivier, J., Peeters, N., Feng, D. X., Khounlotham, M., Boucher, C., et al. (2003). Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8024–8029. doi: 10.1073/pnas.1230660100
- Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N. P., et al. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* 13, 6580–6603. doi: 10.2741/3175
- Dinkel, H., Michael, S., Weatheritt, R. J., Davey, N. E., Van Roey, K., Altenberg, B., et al. (2012). ELM—the database of eukaryotic linear motifs. *Nucleic Acids Res.* 40, D242–D251. doi: 10.1093/nar/gkr1064
- Dosztanyi, Z., Meszaros, B., and Simon, I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 25, 2745–2746. doi: 10.1093/bioinformatics/btp518
- Downen, R. H., Engel, J. L., Shao, F., Ecker, J. R., and Dixon, J. E. (2009). A family of bacterial cysteine protease type III effectors utilizes acylation-dependent and -independent strategies to localize to plasma membranes. *J. Biol. Chem.* 284, 15867–15879. doi: 10.1074/jbc.M900519200
- Edwards, R. J., Davey, N. E., and Shields, D. C. (2007). SLiMFinder: a probabilistic method for identifying over-represented, convergently evolved, short linear motifs in proteins. *PLoS ONE* 2:e967. doi: 10.1371/journal.pone.0000967
- Edwards, R. J., Davey, N. E., and Shields, D. C. (2008). CompariMotif: quick and easy comparisons of sequence motifs. *Bioinformatics* 24, 1307–1309. doi: 10.1093/bioinformatics/btn105
- Fuxreiter, M., Tompa, P., and Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* 23, 950–956. doi: 10.1093/bioinformatics/btm035
- Garau, G., Lemaire, D., Vernet, T., Dideberg, O., and Di Guilmi, A. M. (2005). Crystal structure of phosphorylcholine esterase domain of the virulence factor choline-binding protein e from *Streptococcus pneumoniae*: new structural features among the metallo-beta-lactamase superfamily. *J. Biol. Chem.* 280, 28591–28600. doi: 10.1074/jbc.M502744200
- Garmory, H. S., and Titball, R. W. (2004). ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. *Infect. Immun.* 72, 6757–6763. doi: 10.1128/IAI.72.12.6757-6763.2004
- Gaspar, A. H., Marraffini, L. A., Glass, E. M., Debord, K. L., Ton-That, H., and Schneewind, O. (2005). *Bacillus anthracis* sortase A (SrtA) anchors LPXTG motif-containing surface proteins to the cell wall envelope. *J. Bacteriol.* 187, 4646–4655. doi: 10.1128/JB.187.13.4646-4655.2005
- Gidda, S. K., Shockey, J. M., Rothstein, S. J., Dyer, J. M., and Mullen, R. T. (2009). *Arabidopsis thaliana* GPAT8 and GPAT9 are localized to the ER and possess distinct ER retrieval signals: functional divergence of the dilysine ER retrieval motif in plant cells. *Plant Physiol. Biochem.* 47, 867–879. doi: 10.1016/j.plaphy.2009.05.008

- Ham, J. H., Majerczak, D. R., Arroyo-Rodriguez, A. S., Mackey, D. M., and Coplin, D. L. (2006). WtsE, an AvrE-family effector protein from *Pantoea stewartii* subsp. *stewartii*, causes disease-associated cell death in corn and requires a chaperone protein for stability. *Mol. Plant Microbe Interact.* 19, 1092–1102. doi: 10.1094/MPMI-19-1092
- Ham, J. H., Majerczak, D. R., Nomura, K., Mecey, C., Uribe, F., He, S. Y., et al. (2009). Multiple activities of the plant pathogen type III effector proteins WtsE and AvrE require WxxxE motifs. *Mol. Plant Microbe Interact.* 22, 703–712. doi: 10.1094/MPMI-22-6-0703
- Hamiaux, C., Van Eerde, A., Parsot, C., Broos, J., and Dijkstra, B. W. (2006). Structural mimicry for vinculin activation by IpaA, a virulence factor of *Shigella flexneri*. *EMBO Rep.* 7, 794–799. doi: 10.1038/sj.embor.7400753
- Hammerschmidt, S., Tillig, M. P., Wolff, S., Vaerman, J. P., and Chhatwal, G. S. (2000). Species-specific binding of human secretory component to SpsA protein of *Streptococcus pneumoniae* via a hexapeptide motif. *Mol. Microbiol.* 36, 726–736. doi: 10.1046/j.1365-2958.2000.01897.x
- Harris, T. O., Shelver, D. W., Bohnsack, J. F., and Rubens, C. E. (2003). A novel streptococcal surface protease promotes virulence, resistance to opsonophagocytosis, and cleavage of human fibrinogen. *J. Clin. Invest.* 111, 61–70. doi: 10.1172/JCI200316270
- Hayashi, T., Morohashi, H., and Hatakeyama, M. (2013). Bacterial EPIYA effectors—where do they come from? What are they? Where are they going? *Cell. Microbiol.* 15, 377–385. doi: 10.1111/cmi.12040
- Hicks, S. W., Charron, G., Hang, H. C., and Galan, J. E. (2011). Subcellular targeting of *Salmonella* virulence proteins by host-mediated S-palmitoylation. *Cell Host Microbe* 10, 9–20. doi: 10.1016/j.chom.2011.06.003
- Hicks, S. W., and Galan, J. E. (2013). Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. *Nat. Rev. Microbiol.* 11, 316–326. doi: 10.1038/nrmicro3009
- Higashi, H., Yokoyama, K., Fujii, Y., Ren, S., Yuasa, H., Saadat, I., et al. (2005). EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. *J. Biol. Chem.* 280, 23130–23137. doi: 10.1074/jbc.M503583200
- Itzek, A., Gillen, C. M., Fulde, M., Friedrichs, C., Rodloff, A. C., Chhatwal, G. S., et al. (2010). Contribution of plasminogen activation towards the pathogenic potential of oral streptococci. *PLoS ONE* 5:e13826. doi: 10.1371/journal.pone.0013826
- Jackson, L. K., Nawabi, P., Hentea, C., Roark, E. A., and Haldar, K. (2008). The *Salmonella* virulence protein SifA is a G protein antagonist. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14141–14146. doi: 10.1073/pnas.0801872105
- Jackson, M. R., Nilsson, T., and Peterson, P. A. (1990). Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *EMBO J.* 9, 3153–3162.
- Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662. doi: 10.1110/ps.0303703
- Kadaveru, K., Vyas, J., and Schiller, M. R. (2008). Viral infection and human disease—insights from minimotifs. *Front. Biosci.* 13, 6455–6471. doi: 10.2741/3166
- Kedariseti, P., Mizianty, M. J., Kaas, Q., Craik, D. J., and Kurgan, L. (2014). Prediction and characterization of cyclic proteins from sequences in three domains of life. *Biochim. Biophys. Acta.* 1844, 181–190. doi: 10.1016/j.bbapap.2013.05.002
- Klein, C., Garcia-Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., et al. (2005). The membrane proteome of *Halobacterium salinarum*. *Proteomics* 5, 180–197. doi: 10.1002/pmic.200400943
- Konkel, M. E., Larson, C. L., and Flanagan, R. C. (2010). *Campylobacter jejuni* FlpA binds fibronectin and is required for maximal host cell adherence. *J. Bacteriol.* 192, 68–76. doi: 10.1128/JB.00969-09
- Kovacs-Simon, A., Titball, R. W., and Michell, S. L. (2011). Lipoproteins of bacterial pathogens. *Infect. Immun.* 79, 548–561. doi: 10.1128/IAI.00682-10
- Lee, S. F., Kelly, M., McAlister, A., Luck, S. N., Garcia, E. L., Hall, R. A., et al. (2008). A C-terminal class I PDZ binding motif of EspI/NleA modulates the virulence of attaching and effacing *Escherichia coli* and *Citrobacter rodentium*. *Cell. Microbiol.* 10, 499–513. doi: 10.1111/j.1462-5822.2007.01065.x
- Lehane, A. M., Korres, H., and Verma, N. K. (2005). Bacteriophage-encoded glucosyltransferase GtrII of *Shigella flexneri*: membrane topology and identification of critical residues. *Biochem. J.* 389, 137–143. doi: 10.1042/BJ20050102
- Lety, M. A., Frehel, C., Dubail, I., Beretti, J. L., Kayal, S., Berche, P., et al. (2001). Identification of a PEST-like motif in listeriolysin O required for phagosomal escape and for virulence in *Listeria monocytogenes*. *Mol. Microbiol.* 39, 1124–1139. doi: 10.1111/j.1365-2958.2001.02281.x
- Liverman, A. D., Cheng, H. C., Trosky, J. E., Leung, D. W., Yarbrough, M. L., Burdette, D. L., et al. (2007). Arp2/3-independent assembly of actin by *Vibrio* type III effector VopL. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17117–17122. doi: 10.1073/pnas.0703196104
- Martinez, E., Schroeder, G. N., Berger, C. N., Lee, S. F., Robinson, K. S., Badea, L., et al. (2010). Binding to Na⁺/H⁺ exchanger regulatory factor 2 (NHERF2) affects trafficking and function of the enteropathogenic *Escherichia coli* type III secretion system effectors Map, EspI and NleH. *Cell. Microbiol.* 12, 1718–1731. doi: 10.1111/j.1462-5822.2010.01503.x
- Mooney, C., Haslam, N. J., Holton, T. A., Pollastri, G., and Shields, D. C. (2013). PeptideLocator: prediction of bioactive peptides in protein sequences. *Bioinformatics* 29, 1120–1126. doi: 10.1093/bioinformatics/btt103
- Mooney, C., Pollastri, G., Shields, D. C., and Haslam, N. J. (2012). Prediction of short linear protein binding regions. *J. Mol. Biol.* 415, 193–204. doi: 10.1016/j.jmb.2011.10.025
- Navarre, W. W., and Schneewind, O. (1994). Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Mol. Microbiol.* 14, 115–121. doi: 10.1111/j.1365-2958.1994.tb01271.x
- Neduva, V., and Russell, R. B. (2005). Linear motifs: evolutionary interaction switches. *FEBS Lett.* 579, 3342–3345. doi: 10.1016/j.febslet.2005.04.005
- Nilsson, T., Jackson, M., and Peterson, P. A. (1989). Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic reticulum. *Cell* 58, 707–718. doi: 10.1016/0092-8674(89)90105-0
- Nogueira, S. V., Smith, A. A., Qin, J. H., and Pal, U. (2012). A surface enolase participates in *Borrelia burgdorferi*-plasminogen interaction and contributes to pathogen survival within feeding ticks. *Infect. Immun.* 80, 82–90. doi: 10.1128/IAI.05671-11
- Paschos, A., Den Hartigh, A., Smith, M. A., Atluri, V. L., Sivanesan, D., Tsolis, R. M., et al. (2011). An *in vivo* high-throughput screening approach targeting the type IV secretion system component VirB8 identified inhibitors of *Brucella abortus* 2308 proliferation. *Infect. Immun.* 79, 1033–1043. doi: 10.1128/IAI.00993-10
- Patey, G., Qi, Z., Bourg, G., Baron, C., and O'Callaghan, D. (2006). Swapping of periplasmic domains between *Brucella suis* VirB8 and a pSB102 VirB8 homologue allows heterologous complementation. *Infect. Immun.* 74, 4945–4949. doi: 10.1128/IAI.00584-06
- Prieto, C., Garcia-Estrada, C., Lorenzana, D., and Martin, J. F. (2012). NRPSp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics* 28, 426–427. doi: 10.1093/bioinformatics/btr659
- Rabin, S. D., and Hauser, A. R. (2005). Functional regions of the *Pseudomonas aeruginosa* cytotoxin ExoU. *Infect. Immun.* 73, 573–582. doi: 10.1128/IAI.73.1.573-582.2005
- Ramiscal, R. R., Tang, S. S., Korres, H., and Verma, N. K. (2010). Structural and functional divergence of the newly identified GtrIc from its Gtr family of conserved *Shigella flexneri* serotype-converting glucosyltransferases. *Mol. Membr. Biol.* 27, 114–122. doi: 10.3109/09687680903552250
- Robert-Seilantiantz, A., Shan, L., Zhou, J. M., and Tang, X. (2006). The *Pseudomonas syringae* pv. tomato DC3000 type III effector HopF2 has a putative myristoylation site required for its avirulence and virulence functions. *Mol. Plant Microbe Interact.* 19, 130–138. doi: 10.1094/MPMI-19-0130
- Rohde, J. R., Breikreutz, A., Chenal, A., Sansonetti, P. J., and Parsot, C. (2007). Type III secretion effectors of the IpaH family are E3 ubiquitin ligases. *Cell Host Microbe* 1, 77–83. doi: 10.1016/j.chom.2007.02.002
- Rohde, M., Puls, J., Buhrdorf, R., Fischer, W., and Haas, R. (2003). A novel sheathed surface organelle of the *Helicobacter pylori* cag type IV secretion system. *Mol. Microbiol.* 49, 219–234. doi: 10.1046/j.1365-2958.2003.03549.x
- Sabet, C., Lecuit, M., Cabanes, D., Cossart, P., and Bierne, H. (2005). LPXTG protein InIj, a newly identified internalin involved in *Listeria monocytogenes* virulence. *Infect. Immun.* 73, 6912–6922. doi: 10.1128/IAI.73.10.6912-6922.2005
- Sagulenko, V., Sagulenko, E., Jakubowski, S., Spudich, E., and Christie, P. J. (2001). VirB7 lipoprotein is exocellular and associates with the *Agrobacterium tumefaciens* T pilus. *J. Bacteriol.* 183, 3642–3651. doi: 10.1128/JB.183.12.3642-3651.2001
- Sato, K., Mori, H., Yoshida, M., and Mizushima, S. (1996). Characterization of a potential catalytic residue, Asp-133, in the high affinity ATP-binding site of

- Escherichia coli* SecA, translocation ATPase. *J. Biol. Chem.* 271, 17439–17444. doi: 10.1074/jbc.271.29.17439
- Schlumberger, M. C., Friebel, A., Buchwald, G., Scheffzek, K., Wittinghofer, A., and Hardt, W. D. (2003). Amino acids of the bacterial toxin SopE involved in G nucleotide exchange on Cdc42. *J. Biol. Chem.* 278, 27149–27159. doi: 10.1074/jbc.M302475200
- Shan, L., Thara, V. K., Martin, G. B., Zhou, J. M., and Tang, X. (2000). The pseudomonas AvrPto protein is differentially recognized by tomato and tobacco and is localized to the plant plasma membrane. *Plant Cell* 12, 2323–2338. doi: 10.1105/tpc.12.12.2323
- Sivanesan, D., and Baron, C. (2011). The dimer interface of *Agrobacterium tumefaciens* VirB8 is important for type IV secretion system function, stability, and association of VirB2 with the core complex. *J. Bacteriol.* 193, 2097–2106. doi: 10.1128/JB.00907-10
- Srikanth, C. V., Wall, D. M., Maldonado-Contreras, A., Shi, H. N., Zhou, D., Demma, Z., et al. (2010). Salmonella pathogenesis and processing of secreted effectors by caspase-3. *Science* 330, 390–393. doi: 10.1126/science.1194598
- Stirling, F. R., Cuzick, A., Kelly, S. M., Oxley, D., and Evans, T. J. (2006). Eukaryotic localization, activation and ubiquitylation of a bacterial type III secreted toxin. *Cell. Microbiol.* 8, 1294–1309. doi: 10.1111/j.1462-5822.2006.00710.x
- Sun, J., Maresso, A. W., Kim, J. J., and Barbieri, J. T. (2004). How bacterial ADP-ribosylating toxins recognize substrates. *Nat. Struct. Mol. Biol.* 11, 868–876. doi: 10.1038/nsmb818
- Sundaramoorthy, R., Fyfe, P. K., and Hunter, W. N. (2008). Structure of *Staphylococcus aureus* EsxA suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein. *J. Mol. Biol.* 383, 603–614. doi: 10.1016/j.jmb.2008.08.047
- Suzuki, M., Mimuro, H., Kiga, K., Fukumatsu, M., Ishijima, N., Morikawa, H., et al. (2009). *Helicobacter pylori* CagA phosphorylation-independent function in epithelial proliferation and inflammation. *Cell Host Microbe* 5, 23–34. doi: 10.1016/j.chom.2008.11.010
- Szurek, B., Rossier, O., Hause, G., and Bonas, U. (2002). Type III-dependent translocation of the *Xanthomonas* AvrBs3 protein into the plant cell. *Mol. Microbiol.* 46, 13–23. doi: 10.1046/j.1365-2958.2002.03139.x
- Tahir, Y. E., Kuusela, P., and Skurnik, M. (2000). Functional mapping of the *Yersinia enterocolitica* adhesin YadA. Identification Of eight NSVAIG - S motifs in the amino-terminal half of the protein involved in collagen binding. *Mol. Microbiol.* 37, 192–206. doi: 10.1046/j.1365-2958.2000.01992.x
- Tegtmeier, N., Hartig, R., Delahay, R. M., Rohde, M., Brandt, S., Conradi, J., et al. (2010). A small fibronectin-mimicking protein from bacteria induces cell spreading and focal adhesion formation. *J. Biol. Chem.* 285, 23515–23526. doi: 10.1074/jbc.M109.096214
- Tegtmeier, N., Wessler, S., and Backert, S. (2011). Role of the cag-pathogenicity island encoded type IV secretion system in *Helicobacter pylori* pathogenesis. *FEBS J.* 278, 1190–1202. doi: 10.1111/j.1742-4658.2011.08035.x
- Thomas, S., Karnik, S., Barai, R. S., Jayaraman, V. K., and Idicula-Thomas, S. (2010). CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Res.* 38, D774–D780. doi: 10.1093/nar/gkp1021
- Tzfira, T., Vaidya, M., and Citovsky, V. (2004). Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. *Nature* 431, 87–92. doi: 10.1038/nature02857
- Wurtele, M., Wolf, E., Pederson, K. J., Buchwald, G., Ahmadian, M. R., Barbieri, J. T., et al. (2001). How the *Pseudomonas aeruginosa* ExoS toxin downregulates Rac. *Nat. Struct. Biol.* 8, 23–26. doi: 10.1038/83007
- Zhang, L., Zhang, C., Ojcius, D. M., Sun, D., Zhao, J., Lin, X., et al. (2012). The mammalian cell entry (Mce) protein of pathogenic *Leptospira* species is responsible for RGD motif-dependent infection of cells and animals. *Mol. Microbiol.* 83, 1006–1023. doi: 10.1111/j.1365-2958.2012.07985.x
- Zhang, Y., and Barbieri, J. T. (2005). A leucine-rich motif targets *Pseudomonas aeruginosa* ExoS within mammalian cells. *Infect. Immun.* 73, 7938–7945. doi: 10.1128/IAI.73.12.7938-7945.2005
- Zhang, Y., Higashide, W. M., McCormick, B. A., Chen, J., and Zhou, D. (2006). The inflammation-associated *Salmonella* SopA is a HECT-like E3 ubiquitin ligase. *Mol. Microbiol.* 62, 786–793. doi: 10.1111/j.1365-2958.2006.05407.x
- Zhou, C. E., Smith, J., Lam, M., Zemla, A., Dyer, M. D., and Slezak, T. (2007). MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res.* 35, D391–D394. doi: 10.1093/nar/gkl791
- Zhu, Y., Li, H., Long, C., Hu, L., Xu, H., Liu, L., et al. (2007). Structural insights into the enzymatic mechanism of the pathogenic MAPK phosphothreonine lyase. *Mol. Cell* 28, 899–913. doi: 10.1016/j.molcel.2007.11.011
- Zimmermann, L., Peterhans, E., and Frey, J. (2010). RGD motif of lipoprotein T, involved in adhesion of *Mycoplasma conjunctivae* to lamb synovial tissue cells. *J. Bacteriol.* 192, 3773–3779. doi: 10.1128/JB.00253-10

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 October 2013; accepted: 05 January 2014; published online: 21 January 2014.

Citation: Ruhanen H, Hurley D, Ghosh A, O'Brien KT, Johnston CR and Shields DC (2014) Potential of known and short prokaryotic protein motifs as a basis for novel peptide-based antibacterial therapeutics: a computational survey. *Front. Microbiol.* 5:4. doi: 10.3389/fmicb.2014.00004

This article was submitted to *Antimicrobials, Resistance and Chemotherapy*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Ruhanen, Hurley, Ghosh, O'Brien, Johnston and Shields. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.