



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Pathogenicity-associated protein domains: The fiercely-conserved evolutionary signatures



Seema Patel*

Bioinformatics and Medical Informatics Research Center, San Diego State University, San Diego 92182, USA

ARTICLE INFO

Keywords:

Protein domains
Pathogenesis
Phylogenetic conservation
Diversification
Shuffling
Immune manipulation

ABSTRACT

Proteins have highly conserved domains that determine their functionality. Out of the thousands of domains discovered so far across all living forms, some of the predominant clinically-relevant domains include IENR1, HNHC, HELICc, Pro-kuma_activ, Tryp_Spc, Lactamase_B, PbH1, ChtBD3, CBM49, acidPPc, G3P_acyltransf, RPOL8c, KbaA, HAMP, HisKA, Hr1, Dak2, APC2, Citrate_ly_lig, DALR, VKc, YARHG, WR1, PWI, ZnF.BED, TUDOR, MHC_II_beta, Integrin_B_tail, Excalibur, DISIN, Cadherin, ACTIN, PROF, Robl_LC7, MIT, Kelch, GAS2, B41, Cyclin_C, Connexin_CCC, OmpH, Bac_rhodopsin, AAA, Knot1, NH, Galanin, IB, Elicitin, ACTH, Cache_2, CHASE, AgrB, PRP, IGR, and Antimicrobial21. These domains are distributed in nucleases/helicases, proteases, esterases, lipases, glycosylase, GTPases, phosphatases, methyltransferases, acyltransferase, acetyltransferase, polymerase, kinase, ligase, synthetase, oxidoreductase, protease inhibitors, nucleic acid binding proteins, adhesion and immunity-related proteins, cytoskeletal component-manipulating proteins, lipid biosynthesis and metabolism proteins, membrane-associated proteins, hormone-like and signaling proteins, etc. These domains are ubiquitous stretches or folds of the proteins in pathogens and allergens. Pathogenesis alleviation efforts can benefit enormously if the characteristics of these domains are known. Hence, this review catalogs and discusses the role of such pivotal domains, suggesting hypotheses for better understanding of pathogenesis at molecular level.

1. Introduction

Proteins are prone to stress-driven modifications in their primary sequences (Marchler-Bauer et al., 2014). Due to the accumulated heterogeneity, homology analysis tools like Basic Local Alignment Search Tool (BLAST) assign low identity scores to even closely-related proteins (Pearson, 2013). Also, the in vitro conditions prevent elaboration of some pathogenically-critical proteins, which misguide interpretations. Such experimentally-absent proteins are termed as 'hypothetical proteins' and dismissed as unimportant from potential drug target candidate list. To overcome these lacunae, an innovative way of analysis is of paramount importance. Proteins have motifs which are strictly conserved, which can reveal their phylogenetic links, diversification paths and functions (Marchler-Bauer et al., 2014). These conserved sites or folds are domains (Marchler-Bauer et al., 2014). Protein databases vary in the number and notations of the domains. SMART (Simple Modular Architecture Research Tool) database catalogs more than a thousand protein domains (Ponting et al., 1999). These

domains belong to different categories of protein. In silico analysis of several viral, bacterial and allergen (from animal and plant origin) proteins have identified frequently-occurring domains. A majority of these pathogenesis-mediating domains are shared between the pathogens as well as allergens. It is important to understand these crucial domains that facilitate the establishment of pathogenesis, as they are potential druggable targets. This review discusses some of such protein domains that manipulate human components and lead to morbidity and lethality.

2. Domains and their functions

The frequently-appearing domains can be clustered into different categories for ease of understanding, though the boundaries are not crisp and often overlapping. Many proteins are known to be modular which contain domains belonging to more than one category of the six possible enzyme classes such as hydrolases, transferases, lyases, oxidoreductases, isomerases and ligases (Cai and Chou, 2005). Also, the

Abbreviations: SMART, Simple Modular Architecture Research Tool; HNHC, homing endonucleases; IENR1, intron-encoded endonuclease repeat; HTH, helix-turn-helix; PAMPs, pathogen associated molecular patterns; CARDs, caspase activation and recruitment domains; UDG, uracil DNA glycosylase; ChtBD, chitin-binding domain; CBM, carbohydrate binding module; CTD, C-terminal domain

* Bioinformatics and Medical Informatics Research Center, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182, USA.

E-mail address: seemabiotech83@gmail.com.

<http://dx.doi.org/10.1016/j.genrep.2017.04.004>

Received 18 December 2016; Received in revised form 29 March 2017; Accepted 7 April 2017

Available online 08 April 2017

2452-0144/ © 2017 Elsevier Inc. All rights reserved.

critical domain-harboring proteins and peptides are protease inhibitors, nucleic acid binding proteins, adhesion and immunity-related proteins, cytoskeletal protein manipulating proteins, lipid biosynthesis and metabolism proteins, membrane-associated proteins, hormone-like and signaling proteins, etc. Relevance of the oft-appearing domains has been discussed below.

2.1. Hydrolase

2.1.1. Nuclease/helicase

HNHc (histidine (His) – asparagine (Asn) – histidine (His)) is a domain in homing endonucleases (the DNA and RNA-targeting enzymes), inteins and introns. Homing nucleases are vital for recombination, genome rearrangement, and virulence (Mehta et al., 2004). This domain of roughly 50 aa has conserved Asn and His residues (Veluchamy et al., 2009). IENR1 (intron-encoded endonuclease repeat) domain in C terminal of HNH family nucleases is made of HTH (helix-turn-helix) sub-domain and globular $\beta\beta\alpha\alpha\beta$ fold (as in type II KH domains), that binds to DNA (Landthaler and Shub, 2003; Oddone et al., 2007). YaeQ domain is a variation of the PD-(D/E)XK motifs in nucleases (and phosphodiesterases), occurring generally in hypothetical proteins (Guzzo et al., 2007). This domain shows homology to transcription elongation protein RfaH and exhibits compensatory activity (Wong et al., 1998). YqgF is a ribonuclease domain, present in ribosomal and RNA-associated proteins (Jin and Pawson, 2012). YqgF proteins are substitutes of RuvC, the nuclease, resolving Holliday junctions during recombination (Bennett et al., 1993). Spt6p (a critical chromatin control gene of *Saccharomyces cerevisiae*) has a domain homologous to YqgF (Ponting, 2002). HELICc (helicase domain near the C terminus) is part of RNA helicases (such as RIG-I (retinoic acid-inducible gene I), MDA5 (melanoma differentiation-associated gene 5), and LGP2 (laboratory of genetics and physiology 2)) (Zou et al., 2009). This domain is involved in viral PAMPs (pathogen associated molecular patterns) recognition (Bhat et al., 2015). This domain co-occurs with other critical domains like DEXD/H (proteins instrumental in fidelity control of splicing process) and CARDs (caspase activation and recruitment domains) (Liu and Cheng, 2015; Zou et al., 2009).

2.1.2. Protease

Pro-kuma_activ (named after pro-kumamolisin, an extracellular proteinase) is a α and β sandwich folded propeptide present in proteases (like trypsin, M28, pyrolysins), which when cleaved, the enzyme becomes active (Comellas-Bigler et al., 2004; Muszewska et al., 2011). Tryp_SPc (Trypsin-like serine protease) domains are present in the serine protease zymogens, which undergo partial cleavage for activation of the protease. Serine proteases are offense and defense proteins of all living organisms -virus, bacteria, plants, invertebrates to human, occurring as surface proteins, secreted molecules, plant latex, digestive enzymes, venom in glands etc. (Gasteiger et al., 2003; Tripathi and Sowdhamini, 2008). Encompassing trypsin, chymotrypsin, collagenase, elastase, thrombin, and subtilisin, among others, it performs a plethora of functions like digestion, blood coagulation, immune regulation, protein metabolism, and apoptosis (Di Cera, 2009; Gohara and Di Cera, 2011). If sensitized and activated, this enzyme leads to inflammation, neural diseases and cancers (Jirásková-Vaníčková et al., 2011; Patel, 2017a, 2017b).

2.1.3. Esterase

Lactamase_B is a domain in metal-dependent hydrolases, which include the proteins like β -lactamases, thiolesterases, glyoxalase II family, glutathione hydrolase, and competence proteins etc. (Bradford, 2001). Most of the Lactamase_B-containing proteins bind two zinc ions as cofactor and resist β -lactam antibiotics (Kong et al., 2010). A majority of this domain-containing proteins in bacteria are hypothetical proteins (van Tonder et al., 2014). Lack of the right conditions in vitro might be causal of the lack of experimental evidence

of Lactamase_B-containing proteins.

2.1.4. Lipase

DDHD (named after the four conserved residues) domain has conserved Asp and His residues, modification of which leads to loss of phospholipase and membrane trafficking activity (Inoue et al., 2012). DDHD domain-containing phospholipase A1 family of proteins are required for organelle biogenesis and brain functioning (Inoue et al., 2012; Yamashita et al., 2010). Mutation in this motif has been associated with hereditary spastic paraplegia, a neural disease of slowly progressive weakness (Gonzalez et al., 2013). The acidPPc (acid phosphatase) domain is present in phosphatidate phosphatase, a critical enzyme that acts on phosphate monoesters, liberating diacylglycerol and inorganic phosphate (Carman and Han, 2006). This domain has been detected in HCV and virulent strains of dengue virus. Perturbation of this enzyme in human has been associated with diseases like prostate cancer and osteoporosis, among other pathologies (Araujo and Vihko, 2013).

2.1.5. Glycosylase

UDG (Uracil DNA glycosylase) domain occurs in proteins of uracil DNA glycosylase superfamily. This enzyme removes any uracil (generated by deamination of cytosine) from DNA, averting mutations and aberrations in information pathways (Lucas-Lledó et al., 2011). A study on Vaccinia virus reveals that the uracil DNA glycosylase is crucial for virus DNA replication (De Silva and Moss, 2003). Also, in silico analysis discovered the UDG domain in hepatitis C virus (HCV). Pbh1 (parallel beta-helix repeats) are motifs present in many carbohydrate-lysing enzymes such as pectate lyases, and rhamnogalacturonases (Heffron et al., 1998). These domains are present in SHCBP1 (centralspindlin complex, made of motor protein MKLP1 and GTPase-activating protein MgcRacGAP), involved in cytokinesis initiation (Asano et al., 2014; Pavicic-Kaltenbrunner et al., 2007). These domains are present in polyductin proteins, defect in which causes autosomal recessive polycystic kidney disease (PKD) (Onuchic et al., 2002). These repeats are abundant in exopolysaccharide allergens of *Platanus acerifolia* (London planetree) and pectate lyase 1 of *Juniperus ashei* (Ashejuniper). These leucine-rich repeats (LRR) are present in highly N-glycosylated proteins and are involved in carbohydrate moiety recognition and/or modification (Heffron et al., 1998). The pkhd1 (polycystic kidney and hepatic disease 1) gene product polyductin, associated with kidney disease (Igarashi, 2002; Menezes and Onuchic, 2006) and congenital hepatic fibrosis (Gunay-Aygun et al., 2010) contains these repeats. ChtBD3 (Chitin-binding domain type 3), a chitin-binding domain has been associated with host pathogenesis (Tran et al., 2011). ChtBD3 is present in some Ebola virus strains (such as some isolates of Mayinga-76 outbreak and isolate A0A0F71MH5 from Libria-14 outbreak) as well as dengue virus serotype 3 strains (Messina et al., 2014). A number of pathogenic bacteria, including *Vibrio cholerae* elaborate an enzyme chitin oligosaccharide deacetylase which contains a ChtBD3 domain. Immense role of this domain in virulence is well-substantiated. CBM49 is a carbohydrate binding module (CBM), found at the C terminal of cellulases (Guillén et al., 2010; Shoseyov et al., 2006). The binding of CBM domains to complex glycans has been linked to pathogenesis. Some dengue virus serotype 2 isolates such as P14337, and Q9WDA6 contain a CBM49, whereas isolate Q9WDA6 contains a CBM25 (a starch binding domain found in bacterial amylases).

2.1.6. Guanosine triphosphate

RAB is a domain in Rab subfamily of small guanosine triphosphate (GTPases) (Diekmann et al., 2011). These proteins have wide and tissue-specific distribution, which play part in vesicle trafficking across membranes to their destined targets. These GTPases interact with numerous other components like sorting adaptors, tethering factors, kinases, phosphatases etc. for proper vesicular transport, defect in which can lead to immunodeficiencies, inflammations, neural pathol-

ogies and cancers (Stenmark, 2009). RUN is an N-terminal domain present in proteins crosstalking with Ras-like GTPase (especially in Rap and Rab family members), thus plays role in signaling pathways (Callebaut et al., 2001; Terawaki et al., 2015). The proteins harboring this domain regulate cytoskeletal organization, autophagy, endocytosis, and endosomal maturation; the functions clearly indicating role in pathogenesis. Further, this domain is often associated with DUF4206 domain (Callebaut et al., 2001; Patel and Côté, 2013). DUFs (domains of unknown function), as their name suggest are heavily-modified domains with poor annotations (Goodacre et al., 2014). Tubulin is a domain in tubulin proteins belonging to GTPase family, playing role in polymer formation (Prigozhina et al., 2001). Tubulin proteins harbor immense heterogeneity at their C-terminal end (Redeker et al., 1992). Bacteria have a tubulin homolog, known as FtsZ (filamentous temperature-sensitive protein Z) proteins, that plays role in cell division. FtsZ protein is drafted to the membrane by the actin-related protein FtsA, and together both the proteins form Z ring, initiating bacterial cytokinesis (Loose and Mitchison, 2014). EFh (EF-hand) are Ca^{2+} binding α helical domains of Miro GTPases, the Ca^{2+} sensors maintaining mitochondrial homeostasis (Suzuki et al., 2014). Trematode tegument proteins have this domain, which is characterized to show Ig (immunoglobulin)-binding properties (Wu et al., 2015). ARF (ADP-ribosylation factor) domains are present in GTPases (like Ras) and homologues. This domain is involved in post-Golgi vesicular transport (Boman et al., 2002). A tyrosine kinase Pyk2 regulates Arf1 gene activity through the protein ASAP1 (Arf GTPase-activating protein) (Inoue et al., 2008; Kruljac-Letunic et al., 2003).

2.2. Transferases

PreSET are N-terminal part of cysteine-rich Zn^{2+} -binding SET (Su (var)3–9, Enhancer-of-zeste, Trithorax) domains in histone lysine methyltransferases (HMTase) (Binda et al., 2010; Dillon et al., 2005). PreSET domain has been detected in plant pollen allergens (such as Lig v and Bet v). G3P_{acyl}transf domain is present in glycerol-3-phosphate acyltransferase, a rate limiting enzyme for triacylglycerol biosynthesis (Wendel et al., 2009). This enzyme is required for immune response of the host as observed in Cocksackievirus infection to mice (Karlsson et al., 2009). Some aggressive viral pathogens like dengue serotype 2 isolates lack it, though other serotypes harbor it. CAT (chloramphenicol acetyltransferase) is a trimeric domain that exists in chloramphenicol acetyltransferase, a bacterial enzyme that can metabolize antibiotic chloramphenicol, leading to drug resistance (Biswas et al., 2012; Yao et al., 1999). CTD (C-terminal domain) of RNA polymerase II plays role in pre-mRNA processing, including splicing via phosphorylation (at Tyr residue) (Millhouse and Manley, 2005). On the other hand, transcription termination can occur by dephosphorylation of the Tyr residue at the CTD (Schreieck et al., 2014). CTD crosstalks with the complex transcript elongation factor SPT4/SPT5 to regulate transcription (Dürr et al., 2014). RPOL8c is a subunit of RNA polymerase I, II and III, with role in transcription (Cramer et al., 2001) and microRNA gene regulation (Wang et al., 2010).

2.3. Kinases

KbaA is a key domain in the protein for KinB-signaling pathway activation in sporulation. KinA kinase regulates sporulation initiation in *Bacillus subtilis*, by controlling phosphate supply to the phosphorelay system (Dartois et al., 1996). Going by the literature, this domain in dengue virus might be involved in signaling pathways as well. Though most of these viruses contain this domain, it is missing in a serotype isolate P27909. TPK_{B1} binding domain of thiamine pyrophosphokinase binds to vitamin B1 as the enzyme transfers pyrophosphate group from ATP to vitamin B1, in order to form the coenzyme thiamin pyrophosphate (Baker et al., 2001). This coenzyme is required for functionality of cytosolic transketolase and mitochondrial enzymes for

oxidative decarboxylation of pyruvate, α -ketoglutarate or branched chain amino acids (Mayr et al., 2011). TyrKc are catalytic domain of Tyr-specific kinase subfamily, a group of cell surface receptors. This domain often co-occurs with FN3 (fibronectin type-III), IG (immunoglobulin) and Igc2 (immunoglobulin C-2 type) domains (Bernsel and Von Heijne, 2005). UBA (ubiquitin associated) domain is present in C terminus of proteins like p62, BMSC-UbP, HHR23A, Rad23, SNF1-like kinases, and plays role in inter- and intramolecular communications (Chang et al., 2006; Raasi et al., 2004). These proteins bind to ubiquitin which mediates proteasome complex degradation and optimal protein level retention in cells (Su and Lau, 2009). HAMP (Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases) are approximately 55 aa-long domains present in the proteins coded by transducing genes (Kishii et al., 2007). This domain containing helices and coiled-coil regions, often undergo conformational changes, relaying signals for chemotaxis, pathogenesis, and biofilm formation (Airoola et al., 2013, 2010; Hulko et al., 2006; Matamouros et al., 2015). HWE_{HK} domain is present in HWE type histidine kinases, known to mediate environmental signaling (Galperin, 2005; Karniol and Vierstra, 2004; Lavín et al., 2007). HisKA is a crucial sensor kinase in pathogenic bacteria, including plant pathogen *Pseudomonas syringae* (Willett and Kirby, 2012). It has been detected in pathogenic virus Ebola as well. Hr1 (homology region 1) domain is N-terminal part of Rho effector, or Serine/threonine C-related kinase proteins (PKN/PRK) that occur in multiple isozyme forms. PKN1 (Protein Kinase N1) isoforms abound in neural cells, playing role in cytoskeletal organization and neuronal differentiation. Neuro-pathologies like amyotrophic lateral sclerosis (ALS) and Alzheimer's disease arise due to malfunction of PKN1. Hr1 domain interacts with the small GTPases Rho and Rac, regulating actin dynamics (Flynn et al., 1998; Thauerer et al., 2014; Watson et al., 2016). Dak2 (di- Mg^{2+} ATP binding) domain is found in dihydroxyacetone kinases family, which helps bacteria to imbibe host fatty acids into their membrane phospholipids, via phosphotransferase activity (Kinch et al., 2005; Parsons et al., 2014). YARHG is a 70 amino acid-long extracellular domain, which gets its name from the corresponding conserved motif in the protein sequence. This domain is detected in peptidases and kinase proteins, and predicted to bind bacterial cell wall or its adjacent components as outer membrane lipid or lipopolysaccharide (Coggill et al., 2013; Coggill and Bateman, 2012).

2.4. Ligase and synthetases

APC2 (cullin homology protein), the anaphase promoting complex or cyclosome is part of a ubiquitin ligase that regulates phase transition of mitosis (Puliappadamba et al., 2011; Zhang et al., 2010; Zhou et al., 2011). Some dengue virus isolates (such as P14337 and P29990) contain this domain. Citrate_{ly}lig is the C-terminal domain in the cytosolic enzyme citrate lyase ligase that catalyzes citrate fermentation into acetyl-CoA, and oxaloacetate, coupled with ATP hydrolysis. However, apart from lipid biosynthesis, this domain has been discovered to bind DNA as well (Meyer et al., 1997) and to play role in tumor growth, following acetylation of lysine residues (Lin et al., 2013; Zaidi et al., 2012). DALR is an anticodon binding domain of tRNA synthetase (arganyl/cysteinyll), made of α helices. In human, this domain-containing protein DALRD3 interacts with protein WDR6 (WD Repeat Domain 6) and C3orf60 (chromosome 3 open reading frame 60), involved in autophagy and protein assembly, respectively (Grinchuk et al., 2010; Schyth et al., 2015). DALR₁ domain detected in pollen might have role in manipulating gene expression. DALR₂ domain is found in cysteinyl-tRNA-synthetases that link amino acid to its cognate transfer RNA (Tveit et al., 2014). VKc is the catalytic subunit of vitamin K epoxide reductase. This enzyme processes blood coagulation factors to vitamin K (Oldenburg et al., 2006).

2.5. Protease inhibitors

WR1 is domain in Worm-specific repeat type 1 proteins. This cysteine-rich domain is detected in nematode *Caenorhabditis elegans* (Marchler-Bauer et al., 2014); however, many pathogenic viruses possess this domain or homologues. This domain often co-occurs with KU (BPTI/Kunitz family of serine protease inhibitors) domains.

2.6. Nucleic acid binding proteins

PWI (proline-tryptophan-isoleucine) domains are present in pre-mRNA processing components, the spliceosome, and known to bind RNA as well as DNA (Szymczyzna et al., 2003). PWI-like domains are present in N-terminal of helicases (e.g. Brr2) (Absmeier et al., 2015). Zinc fingers are motifs known to bind DNA, which can be of many types such as BED (named after the *Drosophila* proteins BEAF and DREF), UBR1 (Ubiquitin Protein Ligase E3 Component N-Recognin 1), UBP (ubiquitin-binding domain), U1, LIM (named after the LIN-11, ISL-1 and MEC-3 proteins in *Caenorhabditis elegans*), TTF (transcription termination factor), DBF, CHCC, CDGSH (Cys-Asp-Gly-Ser-His), ZZ, PMZ (plant mutator transposase), and C4 (Gupta et al., 2012). ZnF_BED is a zinc finger domain in chromatin-boundary-element-binding proteins and transposases, required for terminal inverted repeat (TIR) and sub-terminal repeat binding, facilitating their autonomous transposition (Smith et al., 2012). ZnF_A20 domain in N terminus of ZNF216 protein is an inhibitor of cell death-like zinc finger. This domain crosstalks with IKKgamma, RIP, and TRAF6 proteins is involved in ubiquitin mediated IL-1-induced NF-kappaB activation, apoptosis and proteasomal degradation (Huang et al., 2004; Searle et al., 2012). ZnF_NFX (nuclear transcription factor, X-box binding-like 1) is a zinc finger domain in several proteins, including blast resistance Pi54 protein in rice plant (Gupta et al., 2012). Homologues of all these zinc finger motifs have been detected in pathogenic viruses like HCV, HIV, and dengue. ZM (ZASP (Z band alternatively spliced PDZ-containing protein) -like motif) is about 26 aa-long pattern in an α -actinin-binding protein ZASP, and homologues (Lin et al., 2014). ZM domain plays a role in cytoskeletal protein-protein interactions and provides structural integrity to sarcomeres (Klaavuniemi et al., 2004). As a number of proteins involved in ion channel interactions, cytoplasmic and nuclear signaling, enzymatic reactions and cytoskeletal organization bind to Z-line, mutation in ZASP leads to muscular diseases (Martinelli et al., 2014). Mutation and aberrant isoforms in ZASP can lead to myofibrillar myopathy, cardiomyopathy etc. In human, the ZASP binds to mechanosensing protein Ankrd2 (Ankyrin Repeat Domain 2) and the tumor suppressor protein p53. TUDOR domain, a 60 aa-long motif is present in RNA-binding proteins and is involved in RNA metabolism and interactions. Several copies have been detected in arthropods like *Drosophila* (23 instances), and their epigenetic role in modification of chromatin, and gene expression has come forth (Altschul et al., 1997). Binding of this domains to methyl-arginine/ lysine residues, ligand, microRNPs, small RNAs and PIWI (named after P-element Induced Wimpy testis in *Drosophila*) proteins has surfaced. Literature reveals their presence in fungi, protozoa, plants and metazoans (Ying and Chen, 2012), but in silico analyses are revealing their presence in viruses as well. TUDOR domain-containing protein 1, 4 and 5 are antigens expressed on testis cells and are hallmarks of cancer (Yoon et al., 2011).

2.7. Adhesion and immunity-related proteins

Human cell membrane manipulation property was evidenced from domains like MHC_II_beta, and Integrin_B_tail etc. MHC_II_beta (Class II histocompatibility antigen beta) domain is part of the MHC II glycoproteins expressed on antigen-presenting cells (APC) like macrophages, dendritic cells and B lymphocytes. These components are critical as they display fragmented antigens for recognition by helper T cells and successive immune response (Vyas et al., 2008). Integrin_B_tail (Integ-

rin beta subunit cytoplasmic) domain is involved in cell adhesion (Bodeau et al., 2001). Flo11 domain made mostly of β sheets occurs at the N-terminal of Flo11 protein (a flocculin family adhesion protein) as found in yeast (*Saccharomyces cerevisiae*). This protein mediates hyphal formation, invasive growth and plays role in inter-cellular communications (Goossens and Willaert, 2012; Kraushaar et al., 2015). Excalibur (extracellular calcium-binding region) are domains of bacterial surface proteins, showing similarity with Ca^{2+} -binding loop of calmodulin-like EFh domains (Rigden et al., 2003). SVWC (Single domain von Willibrand factor type C) is a group of adhesin proteins. These cysteine-rich proteins play role in immunity and diseases. Bone morphogenetic protein (BMP) is regulated by proteins with VWC domain such as chordin, CHL2 (chordin-like 2), and CV2 (crossveinless 2) (Fujisawa et al., 2009; Zhang et al., 2007). TNFR (Tumor necrosis factor receptor/ nerve growth factor receptor) are repeat-rich extracellular domains, with role in growth factor and cytokine binding. TNF- α (tumor necrosis factor-alpha) is a cytokine mediating diverse inflammatory conditions. The pathological mechanism involves binding of TNF- α to TNFR (Deng, 2007). TNFR-1 acts as a death receptor on ligand-mediated activation, leading to apoptosis (Park et al., 2014). TSP1 (thrombospondin) domain is characterized to regulate cell interactions in vertebrates. Thrombospondins are glycoproteins with calcium-dependent anti-angiogenic property (Iruela-Arispe et al., 2004; Lawler and Lawler, 2012). SCPU (Spore Coat Protein U) domain is found in a bacterial protein family including spore coat proteins, adhesive pili proteins and biofilm-forming proteins (Chin et al., 2015). *Mycococcus xanthus* mcu gene cluster (a CU (chaperone/usher) gene cluster) plays role in spore coat formation (Cao et al., 2015). SCP (sperm coating protein) is a member in the large family SCP/Tpx-1/Ag5/PR-1/Sc7, known to contain extracellular domains. This domain, spanning 120–170 aa and capable of acquiring α - β - α sandwich conformation has been identified in nematode secretome, insect allergen, and semen. During pathogenesis, the expression of genes coding for this protein is upregulated, playing role in immune exacerbation and chronic condition (Chalmers et al., 2008). DISIN (disintegrins) domains inhibit ligand-receptor association. The disintegrin proteins and metalloproteases, are together termed as ADAMs (a disintegrin and metalloprotease), which mediate cellular adhesion and recognition of sequences (Huang et al., 2003). An ADAM with thrombospondin type 1 repeats-13 (ADAMTS13) inhibits platelet aggregation and arterial thrombosis by cleavage of VWF (Xiao et al., 2011). Canary grass (*Phalaris canariensis*) pollen Pha a 1 DISIN is likely to induce pathogenesis via interference with adhesion of integrins. Amb_V domain is found in Amb V pollen allergen in ragweed (*Ambrosia* sp.). A C-terminal helix is the key T cell epitope, leading to immune reactions, though free sulphhydryl groups play role too (Canis et al., 2012). The presence of a similar domain in HCV indicates strong conservation of this domain. C4 is the C-terminal domain in type 4 procollagens, distributed in skin. This domain with tandem repeat renders the triple-helix collagen protein kinked and sheet-like. Mutation in this protein leads to autoimmune diseases like Goodpasture's syndrome (kidney and lungs inflammation) and Alport syndrome (kidney disease) (Abreu-Velez and Howard, 2012). Cadherin proteins mediate calcium dependent cell-cell adhesion and CNS (central nervous system) synapse control. Cadherin_pro domain occurs in N-terminal of cadherins. This prodomain lacks cadherin-cadherin interaction ability, but cleavage of its prosequence in the endoplasmic reticulum (ER) and Golgi apparatus can activate adhesive nature of the cadherin, conferring ability to control synapses (Koch et al., 2004; Latefi et al., 2009; Reinés et al., 2012). CCP (complement control protein) domain containing SUSHI repeats (60 aa long and cysteine-rich) was identified in pathogenic viruses, which played role in complement activation by them. Literature review implies these CCP domains in arthropods like mosquitoes (including *Aedes* sp.) and fruit flies (*Drosophila* sp.), acting as human complement analog and eradicating bacteria (Xiao et al., 2014). CHAD (conserved histidine alpha-helical domain) is an α -helical domain with conserved His residues, which chelates metals. It interacts

with CYTH domain present in adenylyl cyclase and the mammalian thiamine triphosphatases (Iyer and Aravind, 2002). Cell adhesion necessitates binding of integrins with their ligands, which can be influenced by multiple domains. B_lectin is domain present in mannose-specific proteins. Apart from mannose, it recognizes *N*-acetylglucosamine, which can activate classical complement pathway (Muto et al., 2001).

2.8. Cytoskeletal protein binding

ACTIN domain is characteristic of ACTIN subfamily in ACTIN/mreB/sugarkinase/Hsp70 superfamily, clustered together by their common ATPase domain. Cortactin is an actin (F-actin- and Arp2/3 complex)-binding protein, regulating cytoskeleton dynamics and cortical actin-assembly (Shvetsov et al., 2009). PROF domain binds to actin monomers, membrane polyphosphoinositides and poly-L-proline (Michaelsen-Preusse et al., 2016). Robl_LC7 (Roadblock/LC7 family) domains regulate dynein, a motor protein, mediating several other adaptive functions. Mgl is a type of Robl_LC7, gene for which co-occur with gene encoding small GTPases (such as Ras superfamily involved in transduction pathways) (Miertzschke et al., 2011; Wuichet and Søgaard-Andersen, 2015). Also, Robl_LC7 domains group with PROF domain, under profilin-like clan. MIT, involved in microtubule manipulation is present in virulent strains (Zaire and Sudan) of Ebola, while missing in avirulent strain (Reston). Kelch is a conserved domain with β -propeller topology. This repeat-rich domain is widely present across organisms, from virus (Wang et al., 2014; Wilton et al., 2008), plants to humans, and it mediates protein-protein interactions. The kelch-like (KLHL) gene family is spread across multiple chromosomes in human, and several of their coded proteins bind to the E3 ligase cullin 3, playing role in ubiquitination, signaling (such as NF- κ B pathway inhibition), gene expression, actin binding and involved in several diseases (Dhanoa et al., 2013). GAS2 (Growth-arrest-specific protein 2) domains manipulate actin microfilaments, bind to microtubules and lead to cell division arrest (Zhang et al., 2011). Bet v protein essentially contains a GAS2 domain. B41, a plasma membrane-binding domain appears to be a critical domain for pathogenesis. It clearly indicates the role of this domain in attaching to host membrane. A conserved neuronal protein GRP1-associated scaffolding protein (GASP) has a B41 domain (as part of a FERM domain), implicated in binding to membrane as well as cytoskeletal elements like actin (MacNeil and Pohajdak, 2009). DPBB (double-psi beta-barrel) domains are N terminal motifs present in lipoproteins like expansins (such as Phl p) (Kerff et al., 2008).

2.9. Lipid biosynthesis and metabolism

TLC, the acronym of TRAM, LAG1 and CLN8 homology is a domain in membrane proteins, and it has been linked to ceramide synthesis, lipid regulation and neural processes (Winter and Ponting, 2002). Dengue virus polyprotein has this domain, and as the virus manipulates human neural system, it seems rational. AAI (alpha-amylase inhibitor) domain is a tetra-helix fold, which forms a part of LTP (lipid transfer protein) proteins (Zottich et al., 2011). Polyketide synthases are a large group of multifunctional enzymes responsible for elaboration of myriad secondary metabolites, the polyketides, including antibiotics (Anand et al., 2010; Ansari et al., 2004). The diverse array of polyketides is formed by molecular assembly, characterized by the successive addition of chain extension units. This group of enzyme occurs in bacteria, fungi and plants. Apart from the acyl carrier protein (ACP), acyltransferase (AT), and a ketosynthase domains, a variety of β -carbon processing domains (such as ER, KR) occur in these long, modular proteins (Cane, 2010), some of which have been discussed here. PKS_ER are domains in enoylreductase in polyketide synthase enzymes (Gu et al., 2009). PKS_KR is a ketoreductase that reduces keto group to a hydroxy group. Also, studies have found the epimerase activity of PKS_KR which lies in

the conserved Tyr or Ser, flanked by either Tyr or the triad of Leu, Asp, Asp residues (Bonnett et al., 2013; Xie et al., 2016). PKS_TE are domains in thioesterases, catalyzing non-ribosomal synthesis of cyclic peptide antibiotics (Heathcote et al., 2001). PLP (proteolipid protein) is a transmembrane myelin protein or lipophilin, playing role in stabilization of myelin sheaths and axonal survival. Mutant form of this protein causes neuropathies like Pelizaeus-Merzbacher disease and spastic paraplegia type 2 (Arvanitis et al., 2002; Garbern et al., 1997; Miller et al., 2009). COLIPASE is a domain in small pancreatic protein with five conserved disulfide bonds, playing role in lipid metabolism (Berton et al., 2007). In human, colipase-dependent pancreatic triglyceride lipase digests fat into fatty acids and monoacylglycerols (Johnson et al., 2013). COLIPASE is part of flavivirus polyprotein propeptide (119–204 amino acid).

2.10. Transcriptional regulators and stress-related proteins

UreE_C is the C-terminal domain in an accessory protein UreE, which hydrolyses urea into ammonia and carbamic acid. A study reports that *Klebsiella aerogenes* urease catalytic site binds to nickel ions and interacts with accessory proteins, including UreE for activation (Merloni et al., 2014; Song et al., 2001). Skp1 (S-phase kinase-associated protein 1) is a component of the kinase complex and it binds to F-box containing proteins like Cdc4, Skp2, and cyclin F. These adapter proteins act as transcription elongation factor and carry out proteasomal degradation of target proteins (in the form of Skp1-Cul1-F-box protein (SCF), the E3 ubiquitin ligase) by ubiquitylation (Chandra Dantu et al., 2016; Yumimoto et al., 2013). FoP_duplication domain, the acronym of C-terminal duplication domain of Friend of PRMT1, is a target of arginine methyltransferases in humans (van Dijk et al., 2010). Fop is associated with chromatin and is an activator of estrogen receptor target genes (van Dijk et al., 2010). Most isolates of dengue virus have this domain in their polyprotein. H4 is the domain of Histone H4 protein, which is likely to be involved in manipulating human interferon-beta (IFN- β) genes, as literature on hyperacetylation of H3 and H4, inactivating interferon gene expression exists (Parekh and Maniatis, 1999). A yeast study has reported the role of H4 in nucleosome assembly during replication (Shibahara et al., 2000). This domain is conserved in pathogenic viruses like dengue. HALZ is a homeobox associated leucine zipper domain present in transcription factors. The homeodomain binds to DNA while the leucine zipper carries out protein-protein interactions. With prolific growth regulatory role, this domain has been widely studied in plants (Elhiti and Stasolla, 2009). HTH_MARR domain is a helix-turn-helix motif occurring in MarR-family transcriptional regulators, thus facilitating in multiple antibiotic resistance. Also, this DNA binding domain is frequently found in hypothetical proteins, as in *Staphylococcus aureus* (Mohan and Venugopal, 2012), so likely to be present in other bacterial hypothetical proteins as well. Cyclin_C are domains in Cyclin family of proteins, critical for cell cycle progression. The proteins and cyclin-dependent kinase (Cdk) enzymes work in sync to induce phosphorylation of RNA polymerase II. The C-type cyclin along with Cdk8 (recruited by the multi-protein complex Mediator) also responds to stress and carries out transcription regulation and modulation of gene expression by recruiting the fission tools comprising mitochondrial fragmentation and programmed cell death (Strich and Cooper, 2014). Connexin_CCC is a cysteine-rich domain in gap junction channel protein connexin (Riquelme et al., 2013). This protein has many subtypes. The C-terminal of connexin43 regulates assembly, gating, and binding to regulatory proteins by undergoing phosphorylation by kinases (Shin et al., 2001). BAG domain in heat shock protein regulator plays role as co-chaperone of Hsp70 chaperones for proper protein folding with quality control and degradation pathways (Bracher and Verghese, 2015). Role of this domain in regulating the heat shock protein quality check pathways can be correlated to the pathogenesis of the isolates harboring it. BTP (in bromodomain transcription factors and PHD domain (a small

protein domain) containing proteins) are domains of histone-like transcription factors (chromatin-associated proteins, histone acetyltransferases) (Koutelou et al., 2010). This domain recognizes lysine in histones and acetylates them, following which chromatin configuration and gene expression change, leading to viral replication, cancer, and inflammation (Sanchez and Zhou, 2009). BTAD (bacterial transcriptional activator domain) is present in Actinobacteria (Huang et al., 2015); BRLZ (basic region leucine zipper also known as bZIP) is a domain in DNA-binding transcription regulators. This domain performs myriad critical gene regulation tasks by undergoing certain degree of flexibility in their configuration (Miller, 2009). Some well-characterized members of this group of proteins include CREB (cAMP response element binding protein) (Thiel et al., 2005), and MafK (MAF BZIP Transcription Factor K) (Töröcsik et al., 2002). WHy domain occurs in water stress and hypersensitive response proteins, playing role in adaptation to stress, including cold temperature, and desiccation (Ciccarelli and Bork, 2005; Jaspard and Hunault, 2014). This domain is found in LEA (late embryogenesis abundant) proteins, detected in bacteria and archaea, plants, nematodes, and typically induced by exposure to stress conditions (Anderson et al., 2015).

2.11. Membrane-associated proteins

LANC_like domains are present in membrane-associated Lanthionine synthetase C-like protein (LanC) (Chen and Ellis, 2008). The proteins LANCL1 (P40 seven-transmembrane-domain protein) and LANCL2 (testes-specific adriamycin sensitivity protein) are produced profusely in the brain and testes, for their immune defense role (Landlinger et al., 2006; Mayer et al., 2001). Lanthionines (macrocyclic thioether) are present in lantibiotics, a type of antimicrobial peptides, elaborated by some bacterial strains. OmpH (outer membrane protein H) domain has been found critical for pathogenesis. *Pasteurella multocida*, a Gram-negative bacterium causes fatal diseases in animals (porcine atrophic rhinitis, bovine diseases), birds (avian fowl cholera) and sometimes in human, of which OmpH is a major surface antigen (Lee et al., 2007; Okay et al., 2012). Also, other enteric pathogenic bacteria like *Salmonella typhimurium*, *Escherichia coli*, *Yersinia enterocolitica* have the ompH genes, which can be borne in chromosome or plasmids. OmpH shares homology with α helix of the HLA-B27 (human leukocyte antigen subtype), which has been suspected to play role in inflammatory arthritis (Singh and Karrar, 2014). Presence of this critical domain in pollen of grasses suggests similar pathogenic mechanism. Some other pathogenicity-causing motifs in several pollens were readily identified, such as OmpH in KBG41 (Kentucky bluegrass) allergen. Bac_rhodopsin (Bacteriorhodopsin-like protein for sensory function) domain is present in G protein-coupled receptors, a type of photoreceptors (Palfi et al., 2010). Cg6151-P domain is part of a conserved membrane protein of about 190–200 aa long, but little functional characterization (Yao et al., 2009). 7TM_GPCR_Srsx domain occurs in serpentine type seven-transmembrane G-protein-coupled receptor class chemoreceptor Srsx (a Srg superfamily member) (Nagarathnam et al., 2012). This domain has been detected in pathogenic viruses, including dengue virus. AAA are domains in the metalloproteases ATPases, the membrane-tethered enzymes with wide array of functions, including degradation of misfolded proteins, membrane quality control, membrane fusion, DNA replication etc. (Krzywdka et al., 2002). The AAA domain is located in the middle of the protein and it often co-occurs with C-terminal Zn^{2+} -dependent protease (Scharfenberg et al., 2015). Viral non-structural protein NS4A contains this domain as well.

2.12. Hormone-like and signaling proteins

Knot1 is domain in knottins, a broad array of proteins including plant lectins, antimicrobial peptides (e.g. a plant cyclotide kalata B1), plant proteinase/amylose inhibitors, plant γ -thionins and arthropod

defensins. These proteins possess a multitude of functions including inhibitory, cytotoxic, antiviral or hormone-like activity (Gracy et al., 2008). The cysteine-rich domain derives its name from its knot-like topology. Three sulfide bridges in the knot interconnecting each other provide high stability to the protein against temperature, pH and chemicals (Herzig and King, 2015). This protein has shown potential to block insect voltage-gated calcium channels (Herzig and King, 2015). NH (Neurohypophysial hormone) is a domain in proteins of vasopressin/oxytocin gene family. The neurophysin protein with this domain serves as receptor for peptide hormone oxytocin, regulated by phosphatidylinositol-calcium second messenger system (de Bree and Burbach, 1998; Elphick and Rowe, 2009; Van Kesteren et al., 1995). Galanin is a 29 aa-long neuropeptide that controls growth hormone, insulin, somatostatin, adrenal secretion, smooth muscle activity etc. By its endocrine regulation, it intervenes in pain, inflammation, memory, learning, mood swings, feeding, and sexual activities (Kask et al., 1996). Role of this peptide in neural diseases, angiogenesis, cancer, obesity and diabetes has come forth as well (Poritsanos et al., 2009; Stevenson et al., 2012). Pathogenesis via the stimulation of phospholipase C (GAL₂) has been recognized (Lang et al., 2015). IB (Insulin growth factor-binding) domain-containing proteins are growth factors, which bind to receptors for their functions (Siegfried et al., 1992). Elicitin is a group of plant necrotic proteins, exuded by phytopathogenic fungi and oomycetes like *Phytophthora*, *Pythium*, *Hyaloperonospora*, *Albugo* etc. (Uhlíková et al., 2016). This PAMP domain is sulfide-rich (about 6 in number) and it possesses versatile functionality, including the manipulation of host signaling pathways. Some of this domain's sulfur residues have been identified as glycosylation sites. Pythiosis-causing *Pythium insidiosum* elaborates elicitin, which might be mediating the human pathogenesis. The elicitin is a sterol-carrying protein which might be sequestering human cholesterol (Lerkuthirat et al., 2015). The β isoform of elicitin has higher plasma membrane affinity than that of α isoform. The genes inf2A and inf2B were identified to induce the elicitin activity (Huitema et al., 2005). Interestingly, pathogenic bacteria like *Mycobacterium tuberculosis* has three inf genes as well. CT or CTCK (C-terminal cystine knot-like) domains are present in growth factors such as TGF β (transforming growth factor-beta), NGF (nerve growth factor), PDGF (platelet-derived growth factor) and GCH (human chorionic gonadotropin). The knot formed of six cysteines is conserved in the CT domain, though the proteins harboring them can assume multi-meric forms, mediating an array of functions like cell growth, embryonic development, organogenesis, intercellular communication, differentiation, tissue repair and remodeling etc. This domain occurs in VWF, the glycoprotein involved in cell adhesion, homeostasis (Zhou and Springer, 2014), and mucins (Iyer and Acharya, 2011). ACTH_domain, which is present in corticotrophins has been linked to virus immune evasion. As per literature, SARS (Severe Acute Respiratory Syndrome), and influenza virus manipulate host corticosteroid stress response to circumvent the immune response, by expressing protein homologous to host ACTH (adrenocorticotropic hormone) (Wheatland, 2004). As host immunity produces antibodies against the viral ACTH, the antibody binds to host ACTH as well, leading to adrenal gland injuries, hampering corticosteroid secretion (Wheatland, 2004). Also, adrenal deficiency, and the dearth of ACTH in HIV patients has been reported (Shashidhar and Shashikala, 2012). ACTH_domain has been detected only in dengue serotype 2 isolates (P14337, P29990, Q9WDA6 etc.). This ACTH-based molecular mimicry mechanism might be linked to the higher virulence of this serotype. Amelin (Ameloblastin precursor) are a group of proteins, found in mammalian enamel matrix. This ameloblastin amelin plays role in tooth crystal formation, as growth factor, though it has been detected to occur in extracellular matrix during embryogenesis and has been discovered to play role in bone repair (Tamburstuen et al., 2011). Ameloblastin binds to calcium and is sensitive to matrix proteases like enamelysin and kallikrein. A study has found that ameloblastin can regulate the genes related to immune responses, by expression of cytokines and induction of STAT

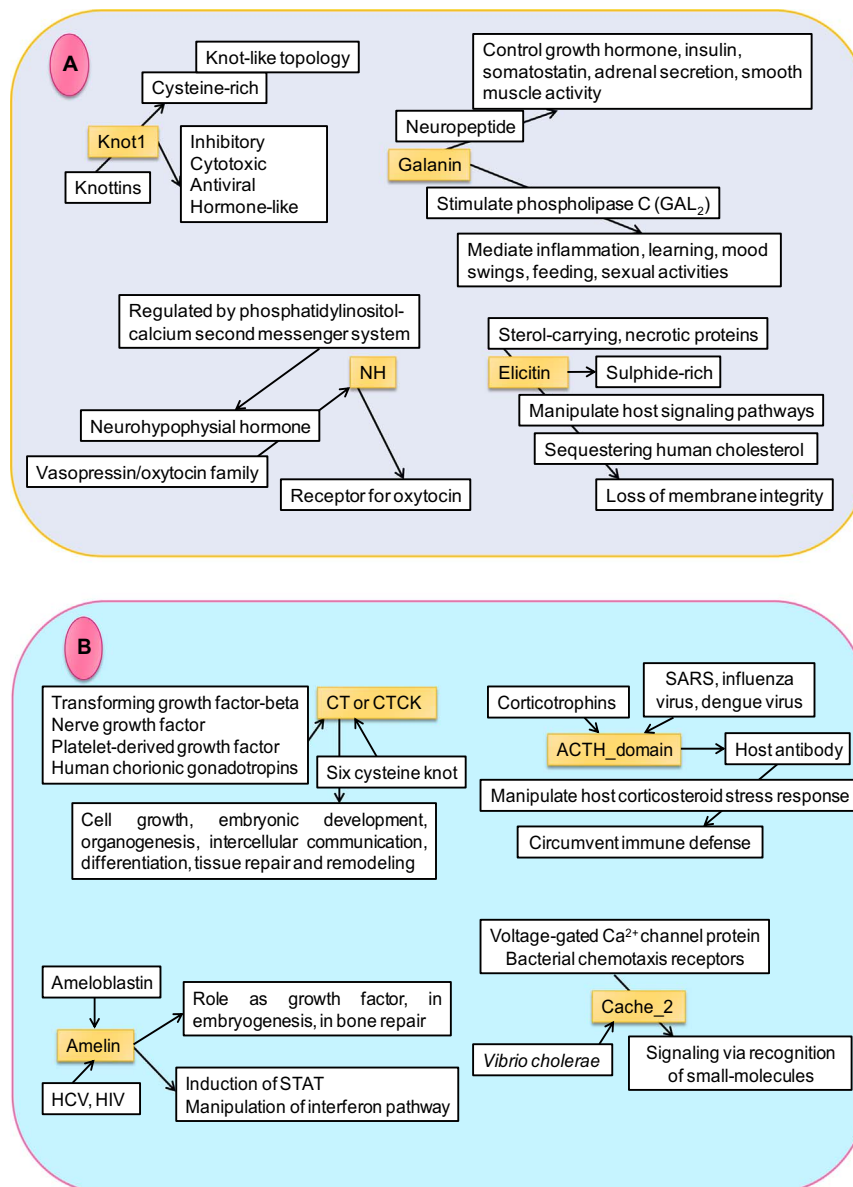


Fig. 1. Pathogenic domains with hormone manipulation properties (A) and (B).

(signal transducer and activator of transcription) in the interferon pathway (Tambursten et al., 2010). Interestingly, analysis has showed that HCV and HIV have amelin domain in their glycoproteins. DEP (named after the proteins Dishevelled, Egl-10, and Pleckstrin P) domain is present in G-protein signaling regulatory proteins. This globular domain made of three-helix bundle, a β -hairpin and two other β -strands modulates signal transduction by manipulating GTPase activity (Capelluto et al., 2014; Wong et al., 2000). DDRGK is a domain occurring widely in plant and vertebrate proteins, and it is named after the corresponding amino acid motif. Studies reveal its role in multiple cell signaling pathways, including NF-kappaB signaling (Wu et al., 2010). Cache_2 is an extracellular domain involved in signaling via recognition of small-molecules. Proteins forming voltage-gated Ca²⁺ channels and bacterial chemotaxis receptors possess this domain. This domain has been well-studied in *Vibrio cholerae* (Upadhyay et al., 2016). CHASE (cyclase/histidine kinase-associated sensing extracellular) is a conserved extracellular sensory domain that helps in perception of environmental changes. As the name indicates, this domain is present in signal transducing systems like histidine kinases, adenylate cyclases, diguanylate cyclases, serine/threonine protein kinases, phosphodies-

terases and methyl-accepting chemotaxis proteins. CHASE domains can be of many types based on functions, out of which CHASE2, 3, 6 are well-studied. CHASE2 are part of serine/threonine kinases, which is followed by transmembrane helices (Mascher et al., 2006; Zhulin et al., 2003). Adequate numbers of studies report their presence in bacteria (Cyanobacteria etc.) signal sensing proteins, however their presence in viruses and pollen is rather new discovery. AgrB (Accessory gene regulator B) family proteins include AgrB from *Staphylococcus aureus* and FsrB from *Enterococcus faecalis*, both regulating expression of virulence genes (Robinson et al., 2005). These are quorum-sensing apparatus in the bacteria, coordinating bacterial communication (Hsieh et al., 2008). Also, these signaling genes have been recently discovered in bacteriophage genomes (Hargreaves et al., 2014). This domain is assumed to perform regulatory role in the dengue virus. Fig. 1 illustrates the pathogenic mechanisms of these domains.

2.13. Other proteins

PRP (prion protein) domain occurs in prion proteins, known to cause neural diseases among animals, such as scrapie, bovine spongi-

Table 1
Protein classes, subclasses, the protein domains falling under them and the functions of the domains.

Protein class	Protein subclass	Protein domains	Functions	
Hydrolase	Nuclease/helicase	HNHC	For recombination, genome rearrangement, and virulence	
		IENR1	Binds to DNA	
		YaeQ	Shows homology to transcription elongation protein	
		YqgFc	Role in recombination	
		HELICc	Recognize viral PAMPs (pathogen associated molecular patterns)	
	Protease	Pro-kuma_activ	Protein hydrolysis	
		Tryp_SPc	Protein hydrolysis	
	Esterase	Lactamase_B	Ester hydrolysis	
			Hydrolyse β -lactam antibiotics	
	Lipase	DDHD	Phospholipase and membrane trafficking activity	
		acidPPc	Present in phosphatidate phosphatase	
	Glycosylase	UDG	Required for virus DNA replication	
		PbH1	Present in pectate lyases, and rhamnogalacturonases	
	Guanosine triphosphate			Involved in cytokinesis initiation
			ChtBD3	In virulence factors of viruses
			CBM	In virulence factors of viruses, bacteria
			RAB	Role in vesicle trafficking across membranes
			RUN	Role in signaling pathways
				Role in regulation of cytoskeletal organization, autophagy, endocytosis, and endosomal maturation
			Tubulin	Role in polymer formation
FtsZ			Role in cell division	
EFh			Part of Ca^{2+} sensors maintaining mitochondrial homeostasis	
ARF			Involved in post-Golgi vesicular transport	
Transferases	–	PreSET	Detected in plant pollen allergens	
		G3P_acyltransf	Required for triacylglycerol biosynthesis	
			Required for immune response of the host	
		CAT	Metabolize antibiotic chloramphenicol	
		CTD	Role in pre-mRNA processing	
		RPOLSc	Role in transcription	
Kinases	–	KbaA	Role in sporulation	
		TPK_B1_binding	Required for the functions of transketolase and mitochondrial enzymes	
		TyrKc	Present in cell surface receptors	
		UBA	Plays role in inter- and intramolecular communications	
		HAMP	Relay signals for chemotaxis, pathogenesis, and biofilm formation	
		HWE_HK	Mediate environmental signaling	
		HisKA	A crucial sensor kinase in pathogens	
		Hr1	Part of Rho effector or PKN enzyme	
			Role in cytoskeletal organization and neuronal differentiation	
			Helps bacteria to imbibe host fatty acids into their membrane phospholipids	
Ligase and synthetases	–	YARHG	Binds to bacterial cell wall or its adjacent components as outer membrane lipid or lipopolysaccharide	
		APC2	Regulates phase transition of mitosis	
		Citrate_ly_lig	Role in lipid biosynthesis	
			Binds to DNA	
Protease inhibitors	–		Role in tumor growth	
		DALR	Involved in autophagy and protein assembly	
Nucleic acid binding proteins	–	VKc	Role in blood coagulation	
		WR1	Inhibits certain proteins	
		PWI	Binds to RNA as well as DNA	
		ZnF_BED	Present in transposases	
		ZnF_A20	Role in apoptosis and proteasomal degradation	
		ZM	Provides structural integrity to sarcomeres	
Adhesion and immunity-related proteins	–	TUDOR	Involved in RNA metabolism and interactions	
			Modulates gene expression	
			Biomarkers of cancer	
		MHC_II_beta	Evoke immune response	
		Integrin_B_tail	Involved in cell adhesion	
		Flo11	Mediates hyphal formation, invasive growth and plays role in inter-cellular communication	
		Excalibur	In bacterial surface proteins	
		SVWC	Role in immunity and diseases	
		TNFR	Role in growth factor and cytokine binding	
		TSP1	Regulate cell interactions in vertebrates	
		SCPU	Present in spore coat protein, adhesive pili proteins and biofilm-forming proteins	
		SCP	Present in nematode secretome, insect allergen, and semen	
			Play role in immune exacerbation and chronic condition	
		DISIN	Inhibit ligand-receptor association	
		Amb_V	Trigger immune reactions	
		C4	Component of collagens	
Cadherin_pro	Control synapses			
CCP	Role in complement activation by pathogens			
CHAD	Chelates metals			

(continued on next page)

Table 1 (continued)

Protein class	Protein subclass	Protein domains	Functions		
Cytoskeletal protein binding	–	B_lectin	Can activate classical complement pathway		
		ACTIN	Regulate cytoskeleton dynamics and cortical actin-assembly		
		PROF	Regulate cytoskeleton dynamics by binding to actin monomers		
		Rob1_LC7	Regulates dynein, a motor protein functions		
		MIT	Manipulate microtubules		
		Kelch	Mediates protein-protein interactions		
		GAS2	Manipulates actin microfilaments, bind to microtubules and lead to cell division arrest		
		B41	Attaches to host membrane and controls actin		
		DPBB	High homology to expansins and GH45 enzymes; Part of plant allergens		
		Lipid biosynthesis and metabolism	–	TLC	Role in ceramide synthesis, lipid regulation and neural processes
AAI	Forms a part of LTP (lipid transfer protein) protein, a plant allergen				
PKS_KR	Polyketide synthase enzymes				
PKS_TE					
PLP	Role in stabilization of myelin sheaths and axonal survival				
COLIPASE	Role in lipid metabolism				
UreE_C	Role in urea hydrolysis				
Skp1	Role as transcription elongation factor				
	Role in proteasomal degradation				
	Activation of estrogen receptor target genes				
Transcriptional regulators and stress-related proteins	–	H4	Manipulation of human IFN- β genes		
		HALZ	Role in protein-protein interactions		
		HTH_MARR	Part of transcriptional regulators		
			Confers multiple antibiotic resistance		
		Cyclin_C	Important for cell cycle progression		
		Connexin_CCC	Regulates assembly, gating, and binding to regulatory proteins		
		BAG	Acts as co-chaperones		
		BTP	Part of histone-like transcription factors		
		BTAD	Role as transcription regulator		
		BRLZ	Role as transcription regulator		
		WHy	Plays role in adaptation to stress, including cold temperature, and desiccation		
		Membrane-associated proteins	–	LANC_like	Have immune defense role
				OmpH	Highly pathogenic (from bacteria to plant pollens)
				Bac_rhodopsin	Present in photoreceptors
				7TM_GPCR_Srsx	Present in pathogenic viruses
				AAA	Part of ATPases; Play role in degradation of misfolded proteins, membrane quality control, membrane fusion, DNA replication
					Have inhibitory, cytotoxic, antiviral or hormone-like activity
		Hormone-like and signaling proteins	–	Knot1	Part of vasopressin/oxytocin
				NH	Controls growth hormone, insulin, somatostatin, adrenal secretion, smooth muscle activity
				Galanin	Part of growth factors
IB	Sequesters human cholesterol; manipulates host signaling pathways				
Elicitin	Part of growth factors				
CT or CTCK	Present in corticotrophins				
ACTH_domain	Part of growth factors; regulates immune responses				
Amelin	Modulates signal transduction				
DEP	Plays role in NF-kappaB signaling				
DDRCK	Plays role in perception of environmental changes				
CHASE	Regulates the expression of virulence genes; Controls quorum-sensing				
AgrB	Malfunaction can cause neural diseases				
PRP	–				
IGR	Exerts antimicrobial, and antifungal properties				
Antimicrobial21					
Other proteins	–				

form encephalopathy (BSE), kuru and Creutzfeldt-Jakob disease. The disease progresses when cellular α -helix-rich prion protein converts into β -sheet-rich amyloid fibril-forming form (Krammer et al., 2008; Kupfer et al., 2009). IGR domain is found in fungal and plant proteins; however, its annotation is too sparse and its function is unknown. Antimicrobial21 is a plant peptide, with two disulfide bonds which gives the peptide an α -helical hairpin fold topology. This peptide is antimicrobial, and antifungal, which binds to fungal conidia, penetrates and pools in the cytoplasm, leading to fungal death (Gautam et al., 2012; Nolde et al., 2011).

Several DUFs (Domain of unknown functions), though poorly-annotated frequently occur in pathogenesis-related proteins. DUF1237 occur in Ebola virus isolate from Zaire strain (isolate Q6V1Q2). This domain overlaps with B41 domain, also adjacent domains occurring before this DUF are exactly same (IENR1, DEP, LamG, Lipid_DES, YqgFc) in another Zaire isolate (A0A0G2Y8I7) which indicates the domain DUF1237 might be just a modified form of B41 domain. DUF1338 in DENV-1 (P17763) and DENV-4 (Q2YHF0 and Q5UCB8)

has zinc-binding function and it is a part of putative metal hydrolase (Marchler-Bauer et al., 2014). DUF1866 in Ebola virus isolate from Reston strain (isolate Q8JXP5 and Q91DD4) is likely to be either Cyt-b5 (cytochrome b5-like heme/steroid binding domain) or CASc (caspase, interleukin-1 beta converting enzyme homologues) domain. DUF862 in Ebola virus isolate from Reston strain (isolate Q8JXP5) lies just above HisKA domain, which might be a Telomerase_RBD domain. DUF4208 in Ebola virus Sudan isolate (Q5XX01) could be PLCyc, RUN, or Cyclin_C domains.

Also, a myriad other catalytically-active motifs present in protozoans and metazoans were detected in the virus polyproteins, the motifs/ domain being Amb_V_allergen, AWS, BTAD, CBD_II, CHAD, CHASE2, ClpB_D2-small, CT, CTD, CxxC_CXXC_SSSS, Dak2, DALR_1, DHDPs, DISIN, DPBB_1, DUF1907, EB_dh, EFh, Excalibur, FA58C, FCD, Galanin, GAS2, GCK, Gp_dh_N, HAMP, HELICc, HWE_HK, Kelch, Knot1, KR, Lig_chan-Glu_bd, LU, MA, MBD, NH, OmpH, PKS_TE, PROF, PWI, RAB, RL11, Rob1_LC7, Romo1, SAA, ShKT, Skp1, SMC_hinge, SRP54, TNFR, YaeQ, and ZnF_C3H1. SMART database can be referred

to for further information on these protein domains (Ponting et al., 1999). Even if some of the domains are non-functional, the findings indicate homology and phylogenetic conservation among organisms. Also, positional shuffling of the domains affirms mosaic nature of the virus nucleic acid, which has been already proven in some DNA virus (Iyer et al., 2002).

Table 1 lists all the pathogenically-critical protein domains outlined above.

3. Discussion

The diverse repertoires of domains have originated by stepwise or drastic reshuffling, depending on the stressors encountered. Several domains co-occur in one protein and crosstalk for critical functions. A huge number of above-discussed domains manipulate host actin protein, hormones and neurons. Its striking that despite being at different levels of evolutionary hierarchy, organisms have significant number of domains shared. The identified domain number is vast and ever-increasing, also they are being frequently re-annotated. Yet, the domains characterized here constitute the core of the pathogenesis mechanisms exploited by most pathogens and allergens.

Many proteins are intrinsically unstructured (such as stress encountering proteins), yet they have highly conserved, structured domains. These domains are clues to the phylogenetic origin, evolutionary trajectories and permutation paths leading to the origin of other protein domains. Many of these critical domains occur in hypothetical proteins of pathogenic bacteria, which are normally ignored while searching for drug targets. It can be hypothesized that the hypothetical proteins with any of these domains are likely to be virulence factors and are eligible to be targeted. Patel has analyzed and reviewed extensively in this area, that has shed light on the conserved domains and their obligatory role in pathogenesis (Patel, 2016a, 2016b, 2016c; Patel, 2017a, 2017b; Patel and Patel, 2016). The current work will be an interesting addition in this direction.

4. Conclusion

To conclude, the most-conserved domains in pathogens and allergens are generally VWC, YARHG, WH1, RICTOR_M, Pro-kuma_activ, IENR1, B41, Y1_Tnp, HOX, HOLI, PLCyC, Hr1, H4, GGDEF, LPD_N, CHASE2, Galanin, Dak2, DALR_1, HAMP, PWI, EFh, Excalibur, CT, Pbh1, HELICc, Kelch, Robl_LC7, YaeQ, PreSET, Bet_v_1, GAS2, CHAD, Integrin_B_tail, MHC_II_beta, DISIN, etc. Using these domains as clues, virulence agents and inflammation mediators can be identified. The genes responsible for coding these protein sequences deserves attention.

Compliance with ethical standards

The author declares that there is no competing interest. This work does not involve human participants or animal models. There are no coauthors, so consent is not required.

References

Abreu-Velez, A.M., Howard, M.S., 2012. Collagen IV in normal skin and in pathological processes. *N. Am. J. Med. Sci.* 4, 1–8. <http://dx.doi.org/10.4103/1947-2714.92892>.

Absmeier, E., Rosenberger, L., Apelt, L., Becke, C., Santos, K.F., Stelzl, U., Wahl, M.C., 2015. A noncanonical PWI domain in the N-terminal helicase-associated region of the spliceosomal Brr2 protein. *Acta Crystallogr. D. Biol. Crystallogr.* 71, 762–771. <http://dx.doi.org/10.1107/S1399004715001005>.

Airola, M.V., Sukomon, N., Samanta, D., Borbat, P.P., Freed, J.H., Watts, K.J., Crane, B.R., 2013. HAMP domain conformers that propagate opposite signals in bacterial chemoreceptors. *PLoS Biol.* 11, e1001479. <http://dx.doi.org/10.1371/journal.pbio.1001479>.

Airola, M.V., Watts, K.J., Bilwes, A.M., Crane, B.R., 2010. Structure of concatenated HAMP domains provides a mechanism for signal transduction. *Structure* 18, 436–448. <http://dx.doi.org/10.1016/j.str.2010.01.013>.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search

programs. *Nucleic Acids Res.* 25, 3389–3402.

Anand, S., Prasad, M.V.R., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., Mohanty, D., 2010. SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res.* 38, W487–W496. <http://dx.doi.org/10.1093/nar/gkq340>.

Anderson, D., Ferreras, E., Trindade, M., Cowan, D., 2015. A novel bacterial Water Hypersensitivity-like protein shows in vivo protection against cold and freeze damage. *FEMS Microbiol. Lett.* 362 fmv110. <http://dx.doi.org/10.1093/femsle/fmv110>.

Ansari, M.Z., Yadav, G., Gokhale, R.S., Mohanty, D., 2004. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* 32, W405–W413. <http://dx.doi.org/10.1093/nar/gkh359>.

Araujo, C.L., Vihko, P.T., 2013. Structure of acid phosphatases. *Methods Mol. Biol.* 1053, 155–166. http://dx.doi.org/10.1007/978-1-62703-562-0_11.

Arvanitis, D.N., Yang, W., Boggs, J.M., 2002. Myelin proteolipid protein, basic protein, the small isoform of myelin-associated glycoprotein, and p42MAPK are associated in the Triton X-100 extract of central nervous system myelin. *J. Neurosci. Res.* 70, 8–23. <http://dx.doi.org/10.1002/jnr.10383>.

Asano, E., Hasegawa, H., Hyodo, T., Ito, S., Maeda, M., Chen, D., Takahashi, M., Hamaguchi, M., Senga, T., 2014. SHCBP1 is required for midbody organization and cytokinesis completion. *Cell Cycle* 13, 2744–2751. <http://dx.doi.org/10.4161/15384101.2015.945840>.

Baker, L.J., Dorocke, J.A., Harris, R.A., Timm, D.E., 2001. The crystal structure of yeast thiamin pyrophosphokinase. *Structure* 9, 539–546.

Bennett, R.J., Dunderdale, H.J., West, S.C., 1993. Resolution of holliday junctions by RuvC resolvase: cleavage specificity and DNA distortion. *Cell* 74, 1021–1031. [http://dx.doi.org/10.1016/0092-8674\(93\)90724-5](http://dx.doi.org/10.1016/0092-8674(93)90724-5).

Bernsel, A., Von Heijne, G., 2005. Improved membrane protein topology prediction by domain assignments. *Protein Sci.* 14, 1723–1728. <http://dx.doi.org/10.1110/ps.051395305>.

Berton, A., Sebban-Kreuzer, C., Crenon, I., 2007. Role of the structural domains in the functional properties of pancreatic lipase-related protein 2. *FEBS J.* 274, 6011–6023. <http://dx.doi.org/10.1111/j.1742-4658.2007.06123.x>.

Bhat, A., Paria, A., Deepika, A., Sreedharan, K., Makesh, M., Bedekar, M.K., Purushothaman, C.S., Rajendran, K.V., 2015. Molecular cloning, characterisation and expression analysis of melanoma differentiation associated gene 5 (MDA5) of green chromiide, *Ectropus suratensis*. *Gene* 557, 172–181. <http://dx.doi.org/10.1016/j.gene.2014.12.028>.

Binda, O., LeRoy, G., Bua, D.J., Garcia, B.A., Gozani, O., Richard, S., 2010. Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: regulation of lysine methyltransferases by physical interaction with their substrates. *Epigenetics* 5, 767–775. <http://dx.doi.org/10.4161/epi.5.8.13278>.

Biswas, T., Houghton, J.L., Garneau-Tsodikova, S., Tsodikov, O.V., 2012. The structural basis for substrate versatility of chloramphenicol acetyltransferase CATI. *Protein Sci.* 21, 520–530. <http://dx.doi.org/10.1002/pro.2036>.

Bodeau, A.L., Berrier, A.L., Mastrangelo, A.M., Martinez, R., LaFlamme, S.E., 2001. A functional comparison of mutations in integrin beta cytoplasmic domains: effects on the regulation of tyrosine phosphorylation, cell spreading, cell attachment and beta1 integrin conformation. *J. Cell Sci.* 114, 2795–2807.

Boman, A.L., Salo, P.D., Hauglund, M.J., Strand, N.L., Rensink, S.J., Zhndankina, O., 2002. ADP-ribosylation factor (ARF) interaction is not sufficient for yeast GGA protein function or localization. *Mol. Biol. Cell* 13, 3078–3095. <http://dx.doi.org/10.1091/mbc.E02-02-0078>.

Bonnett, S.A., Whicher, J.R., Papireddy, K., Florova, G., Smith, J.L., Reynolds, K.A., 2013. Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. *Chem. Biol.* 20, 772–783. <http://dx.doi.org/10.1016/j.chembiol.2013.04.014>.

Bracher, A., Verghese, J., 2015. GrpE, Hsp110/Grp170, HspBP1/Sil1 and BAG domain proteins: nucleotide exchange factors for Hsp70 molecular chaperones. *Subcell. Biochem.* 78, 1–33. http://dx.doi.org/10.1007/978-3-319-11731-7_1.

Bradford, P.A., 2001. Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. *Clin. Microbiol. Rev.* 14, 933–951. table of contents. <http://dx.doi.org/10.1128/CMR.14.4.933-951.2001>.

Cai, Y.-D., Chou, K.-C., 2005. Predicting enzyme subclass by functional domain composition and pseudo amino acid composition. *J. Proteome Res.* 4, 967–971. <http://dx.doi.org/10.1021/pr0500399>.

Callebaut, I., de Gunzburg, J., Goud, B., Mornon, J.P., 2001. RUN domains: a new family of domains involved in Ras-like GTPase signaling. *Trends Biochem. Sci.* 26, 79–83.

Cane, D.E., 2010. Programming of erythromycin biosynthesis by a modular polyketide synthase. *J. Biol. Chem.* 285, 27517–27523. <http://dx.doi.org/10.1074/jbc.R110.144618>.

Canis, M., Becker, S., Gröger, M., Kramer, M.F., 2012. IgE reactivity patterns in patients with allergic rhinoconjunctivitis to ragweed and mugwort pollens. *Am. J. Rhinol. Allergy* 26, 31–35. <http://dx.doi.org/10.2500/ajra.2012.26.3698>.

Cao, S., Wu, M., Xu, S., Yan, X., Mao, X., 2015. Identification of a putative flavin adenine dinucleotide-binding monooxygenase as a regulator for *Mycococcus xanthus* development. *J. Bacteriol.* 197, 1185–1196. <http://dx.doi.org/10.1128/JB.02555-14>.

Capelluto, D.G.S., Zhao, X., Lucas, A., Lemkul, J.A., Xiao, S., Fu, X., Sun, F., Bevan, D.R., Finkielstein, C.V., 2014. Biophysical and molecular-dynamics studies of phosphatidic acid binding by the Dvl-2 DEP domain. *Biophys. J.* 106, 1101–1111. <http://dx.doi.org/10.1016/j.bpj.2014.01.032>.

Carman, G.M., Han, G.-S., 2006. Roles of phosphatidate phosphatase enzymes in lipid metabolism. *Trends Biochem. Sci.* 31, 694–699. <http://dx.doi.org/10.1016/j.tibs.2006.10.003>.

Chalmers, I.W., McArdle, A.J., Coulson, R.M., Wagner, M.A., Schmid, R., Hirai, H.,

- Hoffmann, K.F., 2008. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. *BMC Genomics* 9, 89. <http://dx.doi.org/10.1186/1471-2164-9-89>.
- Chandra Dantu, S., Nathubhai Kachariya, N., Kumar, A., 2016. Molecular dynamics simulations elucidate the mode of protein recognition by Skp1 and the F-box domain in the SCF complex. *Proteins* 84, 159–171. <http://dx.doi.org/10.1002/prot.24963>.
- Chang, Y.-G., Song, A.-X., Gao, Y.-G., Shi, Y.-H., Lin, X.-J., Cao, X.-T., Lin, D.-H., Hu, H.-Y., 2006. Solution structure of the ubiquitin-associated domain of human BMSC-UbP and its complex with ubiquitin. *Protein Sci.* 15, 1248–1259. <http://dx.doi.org/10.1110/ps.051995006>.
- Chen, J.-G., Ellis, B.E., 2008. GCR2 is a new member of the eukaryotic lanthionine synthetase component C-like protein family. *Plant Signal. Behav.* 3, 307–310.
- Chin, C.-Y., Hara, Y., Ghazali, A.-K., Yap, S.-J., Kong, C., Wong, Y.-C., Rozali, N., Koh, S.-F., Hoh, C.-C., Puthucherry, S.D., Nathan, S., 2015. Global transcriptional analysis of *Burkholderia pseudomallei* high and low biofilm producers reveals insights into biofilm production and virulence. *BMC Genomics* 16, 471. <http://dx.doi.org/10.1186/s12864-015-1692-0>.
- Ciccarelli, F.D., Bork, P., 2005. The WHY domain mediates the response to desiccation in plants and bacteria. *Bioinformatics* 21, 1304–1307. <http://dx.doi.org/10.1093/bioinformatics/bti221>.
- Coggill, P., Bateman, A., 2012. The YARHG domain: an extracellular domain in search of a function. *PLoS One* 7, e35575. <http://dx.doi.org/10.1371/journal.pone.0035575>.
- Coggill, P., Eberhardt, R.Y., Finn, R.D., Chang, Y., Jaroszewski, L., Godzik, A., Das, D., Xu, Q., Axelrod, H.L., Aravind, L., Murzin, A.G., Bateman, A., 2013. Two Pfam protein families characterized by a crystal structure of protein lpg2210 from *Legionella pneumophila*. *BMC Bioinformatics* 14, 265. <http://dx.doi.org/10.1186/1471-2105-14-265>.
- Comellas-Bigler, M., Maskos, K., Huber, R., Oyama, H., Oda, K., Bode, W., 2004. 1.2 Å crystal structure of the serine carboxyl proteinase pro-kumamolisin; structure of an intact pro-subtilase. *Structure* 12, 1313–1323. <http://dx.doi.org/10.1016/j.str.2004.04.013>.
- Cramer, P., Bushnell, D.A., Kornberg, R.D., 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863–1876. <http://dx.doi.org/10.1126/science.1059493>.
- Dartois, V., Djavakhishvili, T., Hoch, J.A., 1996. Identification of a membrane protein involved in activation of the KinB pathway to sporulation in *Bacillus subtilis*. *J. Bacteriol.* 178, 1178–1186.
- de Bree, F.M., Burbach, J.P., 1998. Structure-function relationships of the vasopressin prohormone domains. *Cell. Mol. Neurobiol.* 18, 173–191.
- De Silva, F.S., Moss, B., 2003. *Vaccinia virus* uracil DNA glycosylase has an essential role in DNA synthesis that is independent of its glycosylase activity: catalytic site mutations reduce virulence but not virus replication in cultured cells. *J. Virol.* 77, 159–166.
- Deng, G.-M., 2007. Tumor necrosis factor receptor pre-ligand assembly domain is an important therapeutic target in inflammatory arthritis. *BioDrugs* 21, 23–29.
- Dhanoo, B.S., Cogliati, T., Satish, A.G., Bruford, E.A., Friedman, J.S., 2013. Update on the Kelch-like (KLHL) gene family. *Hum. Genomics* 7, 13. <http://dx.doi.org/10.1186/1479-7364-7-13>.
- Di Cera, E., 2009. Serine proteases. *IUBMB Life* 61, 510–515. <http://dx.doi.org/10.1002/iub.186>.
- Diekmann, Y., Seixas, E., Gouw, M., Tavares-Cadete, F., Seabra, M.C., Pereira-Leal, J.B., 2011. Thousands of rab GTPases for the cell biologist. *PLoS Comput. Biol.* 7, e1002217. <http://dx.doi.org/10.1371/journal.pcbi.1002217>.
- Dillon, S.C., Zhang, X., Trievel, R.C., Cheng, X., 2005. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* 6, 227. <http://dx.doi.org/10.1186/gb-2005-6-8-227>.
- Düring, J., Lolas, I.B., Sørensen, B.B., Schubert, V., Houben, A., Melzer, M., Deutzmann, R., Grasser, M., Grasser, K.D., 2014. The transcript elongation factor SPT4/SPT5 is involved in auxin-related gene expression in *Arabidopsis*. *Nucleic Acids Res.* 42, 4332–4347. <http://dx.doi.org/10.1093/nar/gku096>.
- Elhiti, M., Stasolla, C., 2009. Structure and function of homodomain-leucine zipper (HD-Zip) proteins. *Plant Signal. Behav.* 4, 86–88.
- Elphick, M.R., Rowe, M.L., 2009. NGFFamide and echinotocin: structurally unrelated myoactive neuropeptides derived from neurophysin-containing precursors in sea urchins. *J. Exp. Biol.* 212, 1067–1077. <http://dx.doi.org/10.1242/jeb.027599>.
- Flynn, P., Mellor, H., Palmer, R., Panayotou, G., Parker, P.J., 1998. Multiple interactions of PRK1 with RhoA. Functional assignment of the Hrl1 repeat motif. *J. Biol. Chem.* 273, 2698–2705.
- Fujisawa, T., Huang, Y., Sebald, W., Zhang, J.-L., 2009. The binding of von Willebrand factor type C domains of Chordin family proteins to BMP-2 and Tsg is mediated by their SD1 subdomain. *Biochem. Biophys. Res. Commun.* 385, 215–219. <http://dx.doi.org/10.1016/j.bbrc.2009.05.041>.
- Galperin, M.Y., 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol.* 5, 35. <http://dx.doi.org/10.1186/1471-2180-5-35>.
- Garbern, J.Y., Cambi, F., Tang, X.M., Sima, A.A., Vallat, J.M., Bosch, E.P., Lewis, R., Shi, M., Sohi, J., Kraft, G., Chen, K.L., Joshi, I., Leonard, D.G., Johnson, W., Raskind, W., Dlouhy, S.R., Pratt, V., Hodes, M.E., Bird, T., Kamholz, J., 1997. Proteolipid protein is necessary in peripheral as well as central myelin. *Neuron* 19, 205–218.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788.
- Gautam, A., Singh, H., Tyagi, A., Chaudhary, K., Kumar, R., Kapoor, P., Raghava, G.P.S., 2012. CPPsite: a curated database of cell penetrating peptides. *Database (Oxford)* 2012, bas015. <http://dx.doi.org/10.1093/database/bas015>.
- Gohara, D.W., Di Cera, E., 2011. Allostery in trypsin-like proteases suggests new therapeutic strategies. *Trends Biotechnol.* 29, 577–585. <http://dx.doi.org/10.1016/j.tibtech.2011.06.001>.
- Gonzalez, M., Nampoothiri, S., Kornblum, C., Oteyza, A.C., Walter, J., Konidari, I., Hulme, W., Spezziani, F., Schöls, L., Züchner, S., Schüle, R., 2013. Mutations in phospholipase DDHD2 cause autosomal recessive hereditary spastic paraplegia (SPG54). *Eur. J. Hum. Genet.* 21, 1214–1218. <http://dx.doi.org/10.1038/ejhg.2013.29>.
- Goodacre, N.F., Gerloff, D.L., Uetz, P., 2014. Protein domains of unknown function are essential in bacteria. *MBio* 5 e00744-13. <http://dx.doi.org/10.1128/mBio.00744-13>.
- Goossens, K.V.Y., Willaert, R.G., 2012. The N-terminal domain of the Flo11 protein from *Saccharomyces cerevisiae* is an adhesin without mannose-binding activity. *FEMS Yeast Res.* 12, 78–87. <http://dx.doi.org/10.1111/j.1567-1364.2011.00766.x>.
- Gracy, J., Le-Nguyen, D., Gelly, J.-C., Kaas, Q., Heitz, A., Chiche, L., 2008. KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. *Nucleic Acids Res.* 36, D314–D319. <http://dx.doi.org/10.1093/nar/gkm939>.
- Grinchuk, O.V., Jenjaroenpun, P., Orlov, Y.L., Zhou, J., Kuznetsov, V.A., 2010. Integrative analysis of the human cis-antisense gene pairs, miRNAs and their transcription regulation patterns. *Nucleic Acids Res.* 38, 534–547. <http://dx.doi.org/10.1093/nar/gkp954>.
- Gu, L., Wang, B., Kulkarni, A., Geders, T.W., Grindberg, R.V., Gerwick, L., Håkansson, K., Wipf, P., Smith, J.L., Gerwick, W.H., Sherman, D.H., 2009. Metamorphic enzyme assembly in polyketide diversification. *Nature* 459, 731–735. <http://dx.doi.org/10.1038/nature07870>.
- Guillén, D., Sánchez, S., Rodríguez-Sanoja, R., 2010. Carbohydrate-binding domains: multiplicity of biological roles. *Appl. Microbiol. Biotechnol.* 85, 1241–1249. <http://dx.doi.org/10.1007/s00253-009-2331-y>.
- Gunay-Aygun, M., Tuchman, M., Font-Montgomery, E., Lukose, L., Edwards, H., Garcia, A., Ausavarat, S., Ziegler, S.G., Piwnicka-Worms, K., Bryant, J., Bernardini, I., Fischer, R., Huizing, M., Guay-Woodford, L., Gahl, W.A., 2010. PKHD1 sequence variations in 78 children and adults with autosomal recessive polycystic kidney disease and congenital hepatic fibrosis. *Mol. Genet. Metab.* 99, 160–173. <http://dx.doi.org/10.1016/j.jmgm.2009.10.010>.
- Gupta, S.K., Rai, A.K., Kanwar, S.S., Sharma, T.R., 2012. Comparative analysis of zinc finger proteins involved in plant disease resistance. *PLoS One* 7, e42578. <http://dx.doi.org/10.1371/journal.pone.0042578>.
- Guzzo, C.R., Nagem, R.A.P., Barbosa, J.A.R.G., Farah, C.S., 2007. Structure of Xanthomonas axonopodis pv. citri YaeQ reveals a new compact protein fold built around a variation of the PD-(D/E)XK nuclease motif. *Proteins* 69, 644–651. <http://dx.doi.org/10.1002/prot.21556>.
- Hargreaves, K.R., Kropinski, A.M., Clokie, M.R.J., 2014. What does the talking?: quorum sensing signalling genes discovered in a bacteriophage genome. *PLoS One* 9, e85131. <http://dx.doi.org/10.1371/journal.pone.0085131>.
- Heathcote, M.L., Staunton, J., Leadlay, P.F., 2001. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* 8, 207–220.
- Heffron, S., Moe, G.R., Sieber, V., Mengaud, J., Cossart, P., Vitali, J., Journak, F., 1998. Sequence profile of the parallel beta helix in the peptidase superfamily. *J. Struct. Biol.* 122, 223–235. <http://dx.doi.org/10.1006/jsbi.1998.3978>.
- Herzig, V., King, G.F., 2015. The cystine knot is responsible for the exceptional stability of the insecticidal spider toxin ω-Hexatoxin-Hv1a. *Toxins (Basel)* 7, 4366–4380. <http://dx.doi.org/10.3390/toxins7104366>.
- Hsieh, H.-Y., Tseng, C.W., Stewart, G.C., 2008. Regulation of Rot expression in *Staphylococcus aureus*. *J. Bacteriol.* 190, 546–554. <http://dx.doi.org/10.1128/JB.00536-07>.
- Huang, J., Teng, L., Li, L., Liu, T., Li, L., Chen, D., Xu, L.-G., Zhai, Z., Shu, H.-B., 2004. ZNF216 is an A20-like and IkappaB kinase gamma-interacting inhibitor of NFkappaB activation. *J. Biol. Chem.* 279, 16847–16853. <http://dx.doi.org/10.1074/jbc.M309491200>.
- Huang, X., Huang, P., Robinson, M.K., Stern, M.J., Jin, Y., 2003. UNC-71, a disintegrin and metalloprotease (ADAM) protein, regulates motor axon guidance and sex myoblast migration in *C. elegans*. *Development* 130, 3147–3161.
- Huang, X., Pinto, D., Fritz, G., Mascher, T., 2015. Environmental sensing in actinobacteria: a comprehensive survey on the signaling capacity of this phylum. *J. Bacteriol.* 197, 2517–2535. <http://dx.doi.org/10.1128/JB.00176-15>.
- Huitema, E., Vleeshouwers, V.G.A.A., Cakir, C., Kamoun, S., Govers, F., 2005. Differences in intensity and specificity of hypersensitive response induction in *Nicotiana* spp. by INF1, INF2A, and INF2B of *Phytophthora infestans*. *Mol. Plant-Microbe Interact.* 18, 183–193. <http://dx.doi.org/10.1094/MPMI-18-0183>.
- Hulko, M., Berndt, F., Gruber, M., Linder, J.U., Truffault, V., Schultz, A., Martin, J., Schultz, J.E., Lupas, A.N., Coles, M., 2006. The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell* 126, 929–940. <http://dx.doi.org/10.1016/j.cell.2006.06.058>.
- Igarashi, P., 2002. Genetics and pathogenesis of polycystic kidney disease. *J. Am. Soc. Nephrol.* 13, 2384–2398. <http://dx.doi.org/10.1097/01.ASN.0000028643.17901.42>.
- Inoue, H., Baba, T., Sato, S., Ohtsuki, R., Takemori, A., Watanabe, T., Tagaya, M., Tani, K., 2012. Roles of SAM and DDHD domains in mammalian intracellular phospholipase A1 KIAA0725p. *Biochim. Biophys. Acta* 1823, 930–939. <http://dx.doi.org/10.1016/j.bbamer.2012.02.002>.
- Inoue, H., Ha, V.L., Prekeris, R., Randazzo, P.A., 2008. Arf GTPase-activating protein ASAP1 interacts with Rab11 effector FIP3 and regulates pericentrosomal localization of transferrin receptor-positive recycling endosome. *Mol. Biol. Cell* 19, 4224–4237. <http://dx.doi.org/10.1091/mbc.E08-03-0290>.
- Iruela-Arispe, M.L., Luque, A., Lee, N., 2004. Thrombospondin modules and angiogenesis. *Int. J. Biochem. Cell Biol.* 36, 1070–1078. <http://dx.doi.org/10.1016/j.biocel.2004.01.025>.

- Iyer, L.M., Aravind, L., 2002. The catalytic domains of thiamine triphosphatase and CyaB-like adenyllyl cyclase define a novel superfamily of domains that bind organic phosphates. *BMC Genomics* 3, 33.
- Iyer, L.M., Koonin, E.V., Aravind, L., 2002. Extensive domain shuffling in transcription regulators of DNA viruses and implications for the origin of fungal APSES transcription factors. *Genome Biol.* 3 RESEARCH0012.
- Iyer, S., Acharya, K.R., 2011. Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. *FEBS J.* 278, 4304–4322. <http://dx.doi.org/10.1111/j.1742-4658.2011.08350.x>.
- Jaspard, E., Hunault, G., 2014. Comparison of amino acids physico-chemical properties and usage of late embryogenesis abundant proteins, hydrophilins and WHY domain. *PLoS One* 9, e109570. <http://dx.doi.org/10.1371/journal.pone.0109570>.
- Jin, J., Pawson, T., 2012. Modular evolution of phosphorylation-based signalling systems. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 367, 2540–2555. <http://dx.doi.org/10.1098/rstb.2012.0106>.
- Jirásková-Vaničková, J., Ettrich, R., Vorlová, B., Hoffman, H.E., Lepšík, M., Jansa, P., Konvalinka, J., 2011. Inhibition of human serine racemase, an emerging target for medicinal chemistry. *Curr. Drug Targets* 12, 1037–1055.
- Johnson, K., Ross, L., Miller, R., Xiao, X., Lowe, M.E., 2013. Pancreatic lipase-related protein 2 digests fats in human milk and formula in concert with gastric lipase and carboxyl ester lipase. *Pediatr. Res.* 74, 127–132. <http://dx.doi.org/10.1038/pr.2013.90>.
- Karlsson, E.A., Wang, S., Shi, Q., Coleman, R.A., Beck, M.A., 2009. Glycerol-3-phosphate acyltransferase 1 is essential for the immune response to infection with coxsackievirus B3 in mice. *J. Nutr.* 139, 779–783. <http://dx.doi.org/10.3945/jn.108.101683>.
- Karniol, B., Vierstra, R.D., 2004. The HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signaling. *J. Bacteriol.* 186, 445–453.
- Kask, K., Berthold, M., Kahl, U., Nordvall, G., Bartfai, T., 1996. Delineation of the peptide binding site of the human galanin receptor. *EMBO J.* 15, 236–244.
- Kerff, F., Amoroso, A., Herman, R., Sauvage, E., Petrella, S., Filée, P., Charlier, P., Joris, B., Tabuchi, A., Nikolaidis, N., Cosgrove, D.J., 2008. Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16876–16881. <http://dx.doi.org/10.1073/pnas.0809382105>.
- Kinch, L.N., Cheek, S., Grishin, N.V., 2005. EDD, a novel phosphotransferase domain common to mannose transporter EIIA, dihydroxyacetone kinase, and DegV. *Protein Sci.* 14, 360–367. <http://dx.doi.org/10.1110/ps.041114805>.
- Kishii, R., Falzon, L., Yoshida, T., Kobayashi, H., Inouye, M., 2007. Structural and functional studies of the HAMP domain of EnvZ, an osmosensing transmembrane histidine kinase in *Escherichia coli*. *J. Biol. Chem.* 282, 26401–26408. <http://dx.doi.org/10.1074/jbc.M701342200>.
- Klaavaniemi, T., Kelloniemi, A., Yläne, J., 2004. The ZASP-like motif in actinin-associated LIM protein is required for interaction with the alpha-actinin rod and for targeting to the muscle Z-line. *J. Biol. Chem.* 279, 26402–26410. <http://dx.doi.org/10.1074/jbc.M401871200>.
- Koch, A.W., Farooq, A., Shan, W., Zeng, L., Colman, D.R., Zhou, M.-M., 2004. Structure of the neural (N)-cadherin prodomain reveals a cadherin extracellular domain-like fold without adhesive characteristics. *Structure* 12, 793–805. <http://dx.doi.org/10.1016/j.str.2004.02.034>.
- Kong, K.-F., Schneper, L., Mathee, K., 2010. Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. *APMIS* 118, 1–36. <http://dx.doi.org/10.1111/j.1600-0463.2009.02563.x>.
- Koutelou, E., Hirsch, C.L., Dent, S.Y.R., 2010. Multiple faces of the SAGA complex. *Curr. Opin. Cell Biol.* 22, 374–382. <http://dx.doi.org/10.1016/jceb.2010.03.005>.
- Krammer, C., Kremmer, E., Schätzl, H.M., 2008. Dynamic interactions of Sup35p and PrP prion protein domains modulate aggregate nucleation and seeding. *Prion* 2, 99–106.
- Kraushaar, T., Brückner, S., Veelders, M., Rhinow, D., Schreiner, F., Birke, R., Pagenstecher, A., Mösche, H.-U., Essen, L.-O., 2015. Interactions by the fungal Flo11 adhesin depend on a fibronectin type III-like adhesin domain girdled by aromatic bands. *Structure* 23, 1005–1017. <http://dx.doi.org/10.1016/j.str.2015.03.021>.
- Kruljac-Leticina, A., Moelleken, J., Kallian, A., Wieland, F., Blaukat, A., 2003. The tyrosine kinase Pyk2 regulates Arf1 activity by phosphorylation and inhibition of the Arf-GTPase-activating protein ASAP1. *J. Biol. Chem.* 278, 29560–29570. <http://dx.doi.org/10.1074/jbc.M302278200>.
- Krzywyda, S., Brzozowski, A.M., Verma, C., Karata, K., Ogura, T., Wilkinson, A.J., 2002. The crystal structure of the AAA domain of the ATP-dependent protease FtsH of *Escherichia coli* at 1.5 Å resolution. *Structure* 10, 1073–1083.
- Kupfer, L., Hinrichs, W., Groschup, M.H., 2009. Prion protein misfolding. *Curr. Mol. Med.* 9, 826–835.
- Landlinger, C., Salzer, U., Prohaska, R., 2006. Myristoylation of human LanC-like protein 2 (LANCL2) is essential for the interaction with the plasma membrane and the increase in cellular sensitivity to adriamycin. *Biochim. Biophys. Acta* 1758, 1759–1767. <http://dx.doi.org/10.1016/j.bbame.2006.07.018>.
- Landthaler, M., Shub, D.A., 2003. The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille. *Nucleic Acids Res.* 31, 3071–3077.
- Lang, R., Gundlach, A.L., Holmes, F.E., Hobson, S.A., Wynick, D., Hökfelt, T., Kofler, B., 2015. Physiology, signaling, and pharmacology of galanin peptides and receptors: three decades of emerging diversity. *Pharmacol. Rev.* 67, 118–175. <http://dx.doi.org/10.1124/pr.112.006536>.
- Latefi, N.S., Pedraza, L., Schohl, A., Li, Z., Ruthazer, E.S., 2009. N-cadherin prodomain cleavage regulates synapse formation in vivo. *Dev. Neurobiol.* 69, 518–529. <http://dx.doi.org/10.1002/dneu.20718>.
- Lavín, J.L., Kiil, K., Resano, O., Ussery, D.W., Oguiza, J.A., 2007. Comparative genomic analysis of two-component regulatory proteins in *Pseudomonas syringae*. *BMC Genomics* 8, 397. <http://dx.doi.org/10.1186/1471-2164-8-397>.
- Lawler, P.R., Lawler, J., 2012. Molecular basis for the regulation of angiogenesis by thrombospondin-1 and -2. *Cold Spring Harb. Perspect. Med.* 2, a006627. <http://dx.doi.org/10.1101/cshperspect.a006627>.
- Lee, J., Kim, Y.B., Kwon, M., 2007. Outer membrane protein H for protective immunity against *Pasteurella multocida*. *J. Microbiol.* 45, 179–184.
- Lerksuthirat, T., Lohnoo, T., Inkomlue, R., Rujirawat, T., Yingyong, W., Khositnithikul, R., Phaonakrop, N., Roytrakul, S., Sullivan, T.D., Krajaejun, T., 2015. The elicitor-like glycoprotein, ELI025, is secreted by the pathogenic oomycete *Pythium insidiosum* and evades host antibody responses. *PLoS One* 10, e0118547. <http://dx.doi.org/10.1371/journal.pone.0118547>.
- Lin, R., Tao, R., Gao, X., Li, T., Zhou, X., Guan, K.-L., Xiong, Y., Lei, Q.-Y., 2013. Acetylation stabilizes ATP-citrate lyase to promote lipid biosynthesis and tumor growth. *Mol. Cell* 51, 506–518. <http://dx.doi.org/10.1016/j.molcel.2013.07.002>.
- Lin, X., Ruiz, J., Bajraktari, L., Ohman, R., Banerjee, S., Gribble, K., Kaufman, J.D., Wingfield, P.T., Griggs, R.C., Fischbeck, K.H., Mankodi, A., 2014. Z-disc-associated, alternatively spliced, PDZ motif-containing protein (ZASP) mutations in the actin-binding domain cause disruption of skeletal muscle actin filaments in myofibrillar myopathy. *J. Biol. Chem.* 289, 13615–13626. <http://dx.doi.org/10.1074/jbc.M114.550418>.
- Liu, Y.-C., Cheng, S.-C., 2015. Functional roles of DExD/H-box RNA helicases in pre-mRNA splicing. *J. Biomed. Sci.* 22, 54. <http://dx.doi.org/10.1186/s12929-015-0161-z>.
- Loose, M., Mitchison, T.J., 2014. The bacterial cell division proteins FtsA and FtsZ self-organize into dynamic cytoskeletal patterns. *Nat. Cell Biol.* 16, 38–46. <http://dx.doi.org/10.1038/ncb2885>.
- Lucas-Lledó, J.I., Maddamsetti, R., Lynch, M., 2011. Phylogenomic analysis of the uracil-DNA glycosylase superfamily. *Mol. Biol. Evol.* 28, 1307–1317. <http://dx.doi.org/10.1093/molbev/msq318>.
- MacNeil, A.J., Pohajdak, B., 2009. Getting a GRASP on CASP: properties and role of the cytohesin-associated scaffolding protein in immunity. *Immunol. Cell Biol.* 87, 72–80. <http://dx.doi.org/10.1038/icb.2008.71>.
- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2014. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. <http://dx.doi.org/10.1093/nar/gku1221>.
- Martinelli, V.C., Kyle, W.B., Kojic, S., Vitulo, N., Li, Z., Belgrano, A., Mairui, P., Banks, L., Vatta, M., Valle, G., Faulkner, G., 2014. ZASP interacts with the mechanosensing protein Ankrd2 and p53 in the signalling network of striated muscle. *PLoS One* 9, e92259. <http://dx.doi.org/10.1371/journal.pone.0092259>.
- Mascher, T., Helmann, J.D., Unden, G., 2006. Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol. Mol. Biol. Rev.* 70, 910–938. <http://dx.doi.org/10.1128/MMBR.00020-06>.
- Matamouros, S., Hager, K.R., Miller, S.I., 2015. HAMP domain rotation and tilting movements associated with signal transduction in the PhoQ sensor kinase. *MBio* 6, e00616-15. <http://dx.doi.org/10.1128/mBio.00616-15>.
- Mayer, H., Bauer, H., Prohaska, R., 2001. Organization and chromosomal localization of the human and mouse genes coding for LanC-like protein 1 (LANCL1). *Cytogenet. Cell Genet.* 93, 100–104. doi: 56958.
- Mayr, J.A., Freisinger, P., Schlachter, K., Rolinski, B., Zimmermann, F.A., Scheffner, T., Haack, T.B., Koch, J., Ahting, U., Prokisch, H., Sperl, W., 2011. Thiamine pyrophosphokinase deficiency in encephalopathic children with defects in the pyruvate oxidation pathway. *Am. J. Hum. Genet.* 89, 806–812. <http://dx.doi.org/10.1016/j.ajhg.2011.11.007>.
- Mehta, P., Katta, K., Krishnaswamy, S., 2004. HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily. *Protein Sci.* 13, 295–300. <http://dx.doi.org/10.1110/ps.03115604>.
- Menezes, L.F., Onuchic, L.F., 2006. Molecular and cellular pathogenesis of autosomal recessive polycystic kidney disease. *Braz. J. Med. Biol. Res.* 39, 1537–1548. <http://dx.doi.org/10.1590/S0100-879X2006001200004>.
- Merloni, A., Dobrovolska, O., Zambelli, B., Agostini, F., Bazzani, M., Musiani, F., Ciurli, S., 2014. Molecular landscape of the interaction between the urease accessory proteins UreE and UreG. *Biochim. Biophys. Acta* 1844, 1662–1674. <http://dx.doi.org/10.1016/j.bbapap.2014.06.016>.
- Messina, J.P., Brady, O.J., Scott, T.W., Zou, C., Pigott, D.M., Duda, K.A., Bhatt, S., Katzelnick, L., Howes, R.E., Battle, K.E., Simmons, C.P., Hay, S.I., 2014. Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol.* 22, 138–146. <http://dx.doi.org/10.1016/j.tim.2013.12.011>.
- Meyer, M., Dimroth, P., Bott, M., 1997. In vitro binding of the response regulator CitB and of its carboxy-terminal domain to A + T-rich DNA target sequences in the control region of the divergent citC and citS operons of *Klebsiella pneumoniae*. *J. Mol. Biol.* 269, 719–731. <http://dx.doi.org/10.1006/jmbi.1997.1076>.
- Michaelsen-Preusse, K., Zessin, S., Grigoryan, G., Scharkowski, F., Feuge, J., Remus, A., Korte, M., 2016. Neuronal profilins in health and disease: relevance for spine plasticity and Fragile X syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3365–3370. <http://dx.doi.org/10.1073/pnas.1516697113>.
- Miertzschke, M., Koerner, C., Vetter, I.R., Keilberg, D., Hot, E., Leonardy, S., Søgaard-Andersen, L., Wittinghofer, A., 2011. Structural analysis of the Ras-like G protein MglA and its cognate GAP MglB and implications for bacterial polarity. *EMBO J.* 30, 4185–4197. <http://dx.doi.org/10.1038/emboj.2011.291>.
- Miller, M., 2009. The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. *Curr. Protein Pept. Sci.* 10, 244–269.
- Miller, M.J., Kangas, C.D., Macklin, W.B., 2009. Neuronal expression of the proteolipid

- protein gene in the medulla of the mouse. *J. Neurosci. Res.* 87, 2842–2853. <http://dx.doi.org/10.1002/jnr.22121>.
- Millhouse, S., Manley, J.L., 2005. The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. *Mol. Cell. Biol.* 25, 533–544. <http://dx.doi.org/10.1128/MCB.25.2.533-544.2005>.
- Mohan, R., Venugopal, S., 2012. Computational structural and functional analysis of hypothetical proteins of *Staphylococcus aureus*. *Bioinformatics* 8, 722–728. <http://dx.doi.org/10.6026/97320630008722>.
- Muszewska, A., Taylor, J.W., Szczesny, P., Grynberg, M., 2011. Independent subtilases expansions in fungi associated with animals. *Mol. Biol. Evol.* 28, 3395–3404. <http://dx.doi.org/10.1093/molbev/msr176>.
- Muto, S., Takada, T., Matsumoto, K., 2001. Biological activities of human mannose-binding lectin bound to two different ligand sugar structures, Lewis A and Lewis B antigens and high-mannose type oligosaccharides. *Biochim. Biophys. Acta* 1527, 39–46.
- Nagarathnam, B., Kalaimathy, S., Balakrishnan, V., Sowdhamini, R., 2012. Cross-genome clustering of human and *C. elegans* G-protein coupled receptors. *Evol. Bioinform.* 8, 229–259. Online. <http://dx.doi.org/10.4137/EBO.S9405>.
- Nolde, S.B., Vassilevski, A.A., Rogozhin, E.A., Barinov, N.A., Balashova, T.A., Samsonova, O.V., Baranov, Y.V., Feofanov, A.V., Egorov, T.A., Arseniev, A.S., Grishin, E.V., 2011. Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (*Echinochloa crus-galli*). *J. Biol. Chem.* 286, 25145–25153. <http://dx.doi.org/10.1074/jbc.M110.200378>.
- Oddone, A., Lorentzen, E., Basquin, J., Gasch, A., Rybin, V., Conti, E., Sattler, M., 2007. Structural and biochemical characterization of the yeast exosome component Rrp40. *EMBO Rep.* 8, 63–69. <http://dx.doi.org/10.1038/sj.embor.7400856>.
- Okay, S., Ozcengiz, E., Ozcengiz, G., 2012. Immune responses against chimeric DNA and protein vaccines composed of plpEN-OmpH and PlpEC-OmpH from *Pasteurella multocida* A:3 in mice. *Acta Microbiol. Immunol. Hung.* 59, 485–498. <http://dx.doi.org/10.1556/AMicr.59.2012.4.6>.
- Oldenburg, J., Bevans, C.G., Müller, C.R., Watzka, M., 2006. Vitamin K epoxide reductase complex subunit 1 (VKORC1): the key protein of the vitamin K cycle. *Antioxid. Redox Signal.* 8, 347–353. <http://dx.doi.org/10.1089/ars.2006.8.347>.
- Onuchic, L.F., Furu, L., Nagasawa, Y., Hou, X., Eggermann, T., Ren, Z., Bergmann, C., Senderek, J., Esquivel, E., Zeltner, R., Rudnik-Schöneborn, S., Mrug, M., Sweeney, W., Avner, E.D., Zerres, K., Guay-Woodford, L.M., Somlo, S., Germino, G.G., 2002. PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel beta-helix 1 repeats. *Am. J. Hum. Genet.* 70, 1305–1317. <http://dx.doi.org/10.1086/340448>.
- Palfi, A., Millington-Ward, S., Chadderton, N., O'Reilly, M., Goldmann, T., Humphries, M.M., Li, T., Wolfrum, U., Humphries, P., Kenna, P.F., Farrar, G.J., 2010. Adeno-associated virus-mediated rhodopsin replacement provides therapeutic benefit in mice with a targeted disruption of the rhodopsin gene. *Hum. Gene Ther.* 21, 311–323. <http://dx.doi.org/10.1089/hum.2009.119>.
- Parekh, B.S., Maniatis, T., 1999. Virus infection leads to localized hyperacetylation of histones H3 and H4 at the IFN-beta promoter. *Mol. Cell* 3, 125–129.
- Park, Y.-H., Jeong, M.S., Jang, S.B., 2014. Death domain complex of the TNFR-1, TRADD, and RIP1 proteins for death-inducing signaling. *Biochem. Biophys. Res. Commun.* 443, 1155–1161. <http://dx.doi.org/10.1016/j.bbrc.2013.12.068>.
- Parsons, J.B., Broussard, T.C., Bose, J.L., Rosch, J.W., Jackson, P., Subramanian, C., Rock, C.O., 2014. Identification of a two-component fatty acid kinase responsible for host fatty acid incorporation by *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 10532–10537. <http://dx.doi.org/10.1073/pnas.1408797111>.
- Patel, M., Côté, J.-F., 2013. Ras GTPases' interaction with effector domains: breaking the families' barrier. *Commun. Integr. Biol.* 6, e24298. <http://dx.doi.org/10.4161/cib.24298>.
- Patel, S., 2017a. A critical review on serine protease: key immune manipulator and pathology mediator. *Allergol Immunopathol (Madr)*. <http://dx.doi.org/10.1016/j.aller.2016.10.011>.
- Patel, S., 2017b. Letter to the editor on “The necessity of overhaul in perception of microbiological culture methods.”. *Microb. Pathog.* <http://dx.doi.org/10.1016/j.micpath.2016.12.001>.
- Patel, S., 2016a. Letter to the Editor on “The urgency of annotating pathogen hypothetical genes.”. *Gene Reports*. <http://dx.doi.org/10.1016/j.genrep.2016.07.004>.
- Patel, S., 2016b. Drivers of bacterial genomes plasticity and roles they play in pathogen virulence, persistence and drug resistance. *Infect. Genet. Evol.* 45, 151–164. <http://dx.doi.org/10.1016/j.meegid.2016.08.030>.
- Patel, S., 2016c. In silico analysis of Hepatitis C virus (HCV) polyprotein domains and their comparison with other pathogens and allergens to gain insight on pathogenicity mechanisms. *Comput. Biol. Chem.* 65, 91–102. <http://dx.doi.org/10.1016/j.compbiolchem.2016.10.006>.
- Patel, S., Patel, S., 2016. Analysis of Ebola virus polymerase domains to find strain-specific differences and to gain insight on their pathogenicity. *VirusDisease* 27, 242–250. <http://dx.doi.org/10.1007/s13337-016-0334-8>.
- Pavicic-Kaltenbrunner, V., Mishima, M., Glotzer, M., 2007. Cooperative assembly of CYK-4/MgcRacGAP and ZEN-4/MKLP1 to form the centralspindlin complex. *Mol. Biol. Cell* 18, 4992–5003. <http://dx.doi.org/10.1091/mbc.E07-05-0468>.
- Pearson, W.R., 2013. An introduction to sequence similarity (“homology”) searching. *Curr. Protoc. Bioinformatics* Chapter 3, Unit3.1. <http://dx.doi.org/10.1002/0471250953.bi0301s42>.
- Ponting, C.P., 2002. Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* 30, 3643–3652.
- Ponting, C.P., Schultz, J., Milpetz, F., Bork, P., 1999. SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res.* 27, 229–232. <http://dx.doi.org/10.1093/nar/27.1.229>.
- Poritsanos, N.J., Mizuno, T.M., Lautatzis, M.-E., Vrontakis, M., 2009. Chronic increase of circulating galanin levels induces obesity and marked alterations in lipid metabolism similar to metabolic syndrome. *Int. J. Obes.* 33, 1381–1389. <http://dx.doi.org/10.1038/ijo.2009.187>.
- Prigozhina, N.L., Walker, R.A., Oakley, C.E., Oakley, B.R., 2001. Gamma-tubulin and the C-terminal motor domain kinesin-like protein, KLP4, function in the establishment of spindle bipolarity in *Aspergillus nidulans*. *Mol. Biol. Cell* 12, 3161–3174.
- Puliyappadamba, V.T., Wu, W., Bevis, D., Zhang, L., Polin, L., Kilkuskie, R., Finley, R.L., Larsen, S.D., Levi, E., Miller, F.R., Wali, A., Rishi, A.K., 2011. Antagonists of anaphase-promoting complex (APC)-2-cell cycle and apoptosis regulatory protein (CARP)-1 interaction are novel regulators of cell growth and apoptosis. *J. Biol. Chem.* 286, 38000–38017. <http://dx.doi.org/10.1074/jbc.M111.222398>.
- Raasi, S., Orlov, L., Fleming, K.G., Pickart, C.M., 2004. Binding of polyubiquitin chains to ubiquitin-associated (UBA) domains of HHR23A. *J. Mol. Biol.* 341, 1367–1379. <http://dx.doi.org/10.1016/j.jmb.2004.06.057>.
- Redeker, V., Melki, R., Promé, D., Le Caer, J.P., Rossier, J., 1992. Structure of tubulin C-terminal domain obtained by subtilisin treatment. The major alpha and beta tubulin isotypes from pig brain are glutamylated. *FEBS Lett.* 313, 185–192.
- Reinés, A., Bernier, L.-P., McAdam, R., Belkaid, W., Shan, W., Koch, A.W., Séguéla, P., Colman, D.R., Dhaunchak, A.S., 2012. N-cadherin prodomain processing regulates synaptogenesis. *J. Neurosci.* 32, 6323–6334. <http://dx.doi.org/10.1523/JNEUROSCI.0916-12.2012>.
- Rigden, D.J., Jedrzejewski, M.J., Galperin, M.Y., 2003. An extracellular calcium-binding domain in bacteria with a distant relationship to EF-hands. *FEMS Microbiol. Lett.* 221, 103–110. [http://dx.doi.org/10.1016/S0378-1097\(03\)00160-5](http://dx.doi.org/10.1016/S0378-1097(03)00160-5).
- Riquelme, M.A., Kar, R., Gu, S., Jiang, J.X., 2013. Antibodies targeting extracellular domain of connexins for studies of hemichannels. *Neuropharmacology* 75, 525–532. <http://dx.doi.org/10.1016/j.neuropharm.2013.02.021>.
- Robinson, D.A., Monk, A.B., Cooper, J.E., Feil, E.J., Enright, M.C., 2005. Evolutionary genetics of the accessory gene regulator (agr) locus in *Staphylococcus aureus*. *J. Bacteriol.* 187, 8312–8321. <http://dx.doi.org/10.1128/JB.187.24.8312-8321.2005>.
- Sanchez, R., Zhou, M.-M., 2009. The role of human bromodomains in chromatin biology and gene transcription. *Curr. Opin. Drug Discov. Devel.* 12, 659–665.
- Scharfenberg, F., Serek-Heuberger, J., Coles, M., Hartmann, M.D., Habeck, M., Martin, J., Lupas, A.N., Alva, V., 2015. Structure and evolution of N-domains in AAA metalloproteases. *J. Mol. Biol.* 427, 910–923. <http://dx.doi.org/10.1016/j.jmb.2014.12.024>.
- Schrieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., Passmore, L.A., 2014. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit G1c7. *Nat. Struct. Mol. Biol.* 21, 175–179. <http://dx.doi.org/10.1038/nsmb.2753>.
- Schyth, B.D., Bela-Ong, D.B., Jalali, S.A.H., Kristensen, L.B.J., Einer-Jensen, K., Pedersen, F.S., Lorenzen, N., 2015. Two virus-induced MicroRNAs known only from teleost fishes are orthologues of MicroRNAs involved in cell cycle control in humans. *PLoS One* 10, e0132434. <http://dx.doi.org/10.1371/journal.pone.0132434>.
- Searle, M.S., Garner, T.P., Strachan, J., Long, J., Adlington, J., Cavey, J.R., Shaw, B., Layfield, R., 2012. Structural insights into specificity and diversity in mechanisms of ubiquitin recognition by ubiquitin-binding domains. *Biochem. Soc. Trans.* 40, 404–408. <http://dx.doi.org/10.1042/BST20110729>.
- Shashidhar, P.K., Shashikala, G.V., 2012. Low dose adrenocorticotrophic hormone test and adrenal insufficiency in critically ill acquired immunodeficiency syndrome patients. *Indian J. Endocrinol. Metab.* 16, 389–394. <http://dx.doi.org/10.4103/2230-8210.95680>.
- Shibahara, K., Verreault, A., Stillman, B., 2000. The N-terminal domains of histones H3 and H4 are not necessary for chromatin assembly factor-1-mediated nucleosome assembly onto replicated DNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7766–7771.
- Shin, J.L., Solan, J.L., Lampe, P.D., 2001. The regulatory role of the C-terminal domain of connexin43. *Cell Commun. Adhes.* 8, 271–275.
- Shoseyov, O., Shani, Z., Levy, I., 2006. Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* 70, 283–295. <http://dx.doi.org/10.1128/MMBR.00028-05>.
- Shvetsov, A., Berkane, E., Chereau, D., Dominguez, R., Reisler, E., 2009. The actin-binding domain of cofilin is dynamic and unstructured and affects lateral and longitudinal contacts in F-actin. *Cell Motil. Cytoskeleton* 66, 90–98. <http://dx.doi.org/10.1002/cm.20328>.
- Siegfried, J.M., Kasprzyk, P.G., Treston, A.M., Mulshine, J.L., Quinn, K.A., Cuttitta, F., 1992. A mitogenic peptide amide encoded within the E peptide domain of the insulin-like growth factor IB prohormone. *Proc. Natl. Acad. Sci. U. S. A.* 89, 8107–8111.
- Singh, A., Karrar, S., 2014. The role of intracellular organisms in the pathogenesis of inflammatory arthritis. *Int. J. Inflamm.* 2014, 158793. <http://dx.doi.org/10.1155/2014/158793>.
- Smith, A.M., Hansey, C.N., Kaeppler, S.M., 2012. TCUP: a novel hAT transposon active in maize tissue culture. *Front. Plant Sci.* 3, 6. <http://dx.doi.org/10.3389/fpls.2012.00006>.
- Song, H.K., Mulrooney, S.B., Huber, R., Hausinger, R.P., 2001. Crystal structure of *Klebsiella aerogenes* UreE, a nickel-binding metallochaperone for urease activation. *J. Biol. Chem.* 276, 49359–49364. <http://dx.doi.org/10.1074/jbc.M108619200>.
- Stenmark, H., 2009. Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Biol.* 10, 513–525. <http://dx.doi.org/10.1038/nrm2728>.
- Stevenson, L., Allen, W.L., Turkington, R., Jithesh, P.V., Proutski, I., Stewart, G., Lenz, H.-J., Van Schaeybroeck, S., Longley, D.B., Johnston, P.G., 2012. Identification of galanin and its receptor GalR1 as novel determinants of resistance to chemotherapy and potential biomarkers in colorectal cancer. *Clin. Cancer Res.* 18, 5412–5426. <http://dx.doi.org/10.1158/1078-0432.CCR-12-1780>.
- Strich, R., Cooper, K.F., 2014. The dual role of cyclin C connects stress regulated gene expression to mitochondrial dynamics. *Microb. Cell* 1, 318–324.

- Su, V., Lau, A.F., 2009. Ubiquitin-like and ubiquitin-associated domain proteins: significance in proteasomal degradation. *Cell. Mol. Life Sci.* 66, 2819–2833. <http://dx.doi.org/10.1007/s00018-009-0048-9>.
- Suzuki, M., Danilchanka, O., Mekalanos, J.J., 2014. *Vibrio cholerae* T3SS effector VopE modulates mitochondrial dynamics and innate immune signaling by targeting Miro GTPases. *Cell Host Microbe* 16, 581–591. <http://dx.doi.org/10.1016/j.chom.2014.09.015>.
- Szymczyzna, B.R., Bowman, J., McCracken, S., Pineda-Lucena, A., Lu, Y., Cox, B., Lambermon, M., Graveley, B.R., Arrowsmith, C.H., Blencowe, B.J., 2003. Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. *Genes Dev.* 17, 461–475. <http://dx.doi.org/10.1101/gad.1060403>.
- Tambursten, M.V., Reppe, S., Spahr, A., Sabetrasekh, R., Kvalheim, G., Slaby, I., Syversen, U., Lyngstadaas, S.P., Reseland, J.E., 2010. Ameloblastin promotes bone growth by enhancing proliferation of progenitor cells and by stimulating immunoregulators. *Eur. J. Oral Sci.* 118, 451–459. <http://dx.doi.org/10.1111/j.1600-0722.2010.00760.x>.
- Tambursten, M.V., Reseland, J.E., Spahr, A., Brookes, S.J., Kvalheim, G., Slaby, I., Snead, M.L., Lyngstadaas, S.P., 2011. Ameloblastin expression and putative autoregulation in mesenchymal cells suggest a role in early bone formation and repair. *Bone* 48, 406–413. <http://dx.doi.org/10.1016/j.bone.2010.09.007>.
- Terawaki, S., Camosseto, V., Prete, F., Wenger, T., Papadopoulos, A., Rondeau, C., Combes, A., Rodriguez Rodrigues, C., Vu Manh, T.-P., Fallet, M., English, L., Santamaria, R., Soares, A.R., Weil, T., Hammad, H., Desjardins, M., Gorvel, J.-P., Santos, M.A.S., Gatti, E., Pierre, P., 2015. RUN and FYVE domain-containing protein 4 enhances autophagy and lysosome tethering in response to Interleukin-4. *J. Cell Biol.* 210, 1133–1152. <http://dx.doi.org/10.1083/jcb.201501059>.
- Thauerer, B., Zur Nedden, S., Baier-Bitterlich, G., 2014. Protein kinase C-related kinase (PKN/PRK). Potential key-role for PKN1 in protection of hypoxic neurons. *Curr. Neuropharmacol.* 12, 213–218. <http://dx.doi.org/10.2174/1570159X11666131225000518>.
- Thiel, G., Al Sarraj, J., Vinson, C., Stefano, L., Bach, K., 2005. Role of basic region leucine zipper transcription factors cyclic AMP response element binding protein (CREB), CREB2, activating transcription factor 2 and CAAT/enhancer binding protein alpha in cyclic AMP response element-mediated transcription. *J. Neurochem.* 92, 321–336. <http://dx.doi.org/10.1111/j.1471-4159.2004.02882.x>.
- Töröcsik, B., Angelastro, J.M., Greene, L.A., 2002. The basic region and leucine zipper transcription factor Mafk is a new nerve growth factor-responsive immediate early gene that regulates neurite outgrowth. *J. Neurosci.* 22, 8971–8980.
- Tran, H.T., Barnich, N., Mizoguchi, E., 2011. Potential role of chitinases and chitin-binding proteins in host-microbial interactions during the development of intestinal inflammation. *Histol. Histopathol.* 26, 1453–1464.
- Tripathi, L.P., Sowdhamini, R., 2008. Genome-wide survey of prokaryotic serine proteases: analysis of distribution and domain architectures of five serine protease families in prokaryotes. *BMC Genomics* 9, 549. <http://dx.doi.org/10.1186/1471-2164-9-549>.
- Tveit, A.T., Urlich, T., Svenning, M.M., 2014. Metatranscriptomic analysis of arctic peat soil microbiota. *Appl. Environ. Microbiol.* 80, 5761–5772. <http://dx.doi.org/10.1128/AEM.01030-14>.
- Uhlíková, H., Obořil, M., Klempová, J., Šedo, O., Zdráhal, Z., Kašparovský, T., Skládal, P., Lochman, J., 2016. Elicitin-induced distal systemic resistance in plants is mediated through the protein-protein interactions influenced by selected lysine residues. *Front. Plant Sci.* 7, 59. <http://dx.doi.org/10.3389/fpls.2016.00059>.
- Upadhyay, A.A., Fleetwood, A.D., Adebali, O., Finn, R.D., Zhulin, I.B., 2016. Cache domains that are homologous to, but different from PAS domains comprise the largest superfamily of extracellular sensors in prokaryotes. *PLoS Comput. Biol.* 12, e1004862. <http://dx.doi.org/10.1371/journal.pcbi.1004862>.
- van Dijk, T.B., Gillemans, N., Stein, C., Fanis, P., Demmers, J., van de Corput, M., Essers, J., Grosveld, F., Bauer, U.-M., Philipsen, S., 2010. Friend of Prmt1, a novel chromatin target of protein arginine methyltransferases. *Mol. Cell Biol.* 30, 260–272. <http://dx.doi.org/10.1128/MCB.00645-09>.
- Van Kesteren, R.E., Smit, A.B., De Lange, R.P., Kits, K.S., Van Golen, F.A., Van Der Schors, R.C., De With, N.D., Burke, J.F., Geraerts, W.P., 1995. Structural and functional evolution of the vasopressin/oxytocin superfamily: vasopressin-related conopressin is the only member present in *Lymnaea*, and is involved in the control of sexual behavior. *J. Neurosci.* 15, 5989–5998.
- van Tonder, A.J., Mistry, S., Bray, J.E., Hill, D.M.C., Cody, A.J., Farmer, C.L., Klugman, K.P., von Gottberg, A., Bentley, S.D., Parkhill, J., Jolley, K.A., Maiden, M.C.J., Brueggemann, A.B., 2014. Defining the estimated core genome of bacterial populations using a Bayesian decision model. *PLoS Comput. Biol.* 10, e1003788. <http://dx.doi.org/10.1371/journal.pcbi.1003788>.
- Veluchamy, A., Mary, S., Acharya, V., Mehta, P., Deva, T., Krishnaswamy, S., 2009. HNHdb: a database on pattern based classification of HNH domains reveals functional relevance of sequence patterns and domain associations. *Bioinformatics* 4, 80–83.
- Vyas, J.M., Van der Veen, A.G., Ploegh, H.L., 2008. The known unknowns of antigen processing and presentation. *Nat. Rev. Immunol.* 8, 607–618. <http://dx.doi.org/10.1038/nri2368>.
- Wang, G., Wang, Y., Shen, C., Huang, Y., Huang, K., Huang, T.H.M., Nephew, K.P., Li, L., Liu, Y., 2010. RNA polymerase II binding patterns reveal genomic regions involved in microRNA gene regulation. *PLoS One* 5, e13798. <http://dx.doi.org/10.1371/journal.pone.0013798>.
- Wang, Q., Burles, K., Couturier, B., Randall, C.M.H., Shisler, J., Barry, M., 2014. Ectromelia virus encodes a BTB/kelch protein, EVM150, that inhibits NF-κB signaling. *J. Virol.* 88, 4853–4865. <http://dx.doi.org/10.1128/JVI.02923-13>.
- Watson, J.R., Fox, H.M., Nietlisbach, D., Gallop, J.L., Owen, D., Mott, H.R., 2016. Investigation of the interaction between Cdc42 and its effector TOCA1: handover of Cdc42 to the actin regulator N-WASP is facilitated by differential binding affinities. *J. Biol. Chem.* <http://dx.doi.org/10.1074/jbc.M116.724294>.
- Wendel, A.A., Lewin, T.M., Coleman, R.A., 2009. Glycerol-3-phosphate acyltransferases: rate limiting enzymes of triacylglycerol biosynthesis. *Biochim. Biophys. Acta* 1791, 501–506. <http://dx.doi.org/10.1016/j.bbalip.2008.10.010>.
- Wheatland, R., 2004. Molecular mimicry of ACTH in SARS — implications for corticosteroid treatment and prophylaxis. *Med. Hypotheses* 63, 855–862. <http://dx.doi.org/10.1016/j.mehy.2004.04.009>.
- Willett, J.W., Kirby, J.R., 2012. Genetic and biochemical dissection of a HisKA domain identifies residues required exclusively for kinase and phosphatase activities. *PLoS Genet.* 8, e1003084. <http://dx.doi.org/10.1371/journal.pgen.1003084>.
- Wilton, B.A., Campbell, S., Van Buuren, N., Garneau, R., Furukawa, M., Xiong, Y., Barry, M., 2008. Ectromelia virus BTB/kelch proteins, EVM150 and EVM167, interact with cullin-3-based ubiquitin ligases. *Virology* 374, 82–99. <http://dx.doi.org/10.1016/j.virol.2007.11.036>.
- Winter, E., Ponting, C.P., 2002. TRAM, LAG1 and CLN8: members of a novel family of lipid-sensing domains? *Trends Biochem. Sci.* 27, 381–383.
- Wong, H.C., Mao, J., Nguyen, J.T., Srinivas, S., Zhang, W., Liu, B., Li, L., Wu, D., Zheng, J., 2000. Structural basis of the recognition of the dishevelled DEP domain in the Wnt signaling pathway. *Nat. Struct. Biol.* 7, 1178–1184. <http://dx.doi.org/10.1038/82047>.
- Wong, K.R., Hughes, C., Koronakis, V., 1998. A gene, *yaeQ*, that suppresses reduced operon expression caused by mutations in the transcription elongation gene *rfaH* in *Escherichia coli* and *Salmonella typhimurium*. *Mol. Genet. Evol.* 257, 693–696.
- Wu, C., Hou, N., Piao, X., Liu, S., Cai, P., Xiao, Y., Chen, Q., 2015. Non-immune immunoglobulins shield *Schistosoma japonicum* from host immunorecognition. *Sci. Rep.* 5, 13434. <http://dx.doi.org/10.1038/srep13434>.
- Wu, J., Lei, G., Mei, M., Tang, Y., Li, H., 2010. A novel C53/LZAP-interacting protein regulates stability of C53/LZAP and DDRGK domain-containing Protein 1 (DDRK1) and modulates NF-κB signaling. *J. Biol. Chem.* 285, 15126–15136. <http://dx.doi.org/10.1074/jbc.M110.110619>.
- Wuichet, K., Søgaard-Andersen, L., 2015. Evolution and diversity of the Ras superfamily of small GTPases in prokaryotes. *Genome Biol. Evol.* 7, 57–70. <http://dx.doi.org/10.1093/gbe/evu264>.
- Xiao, J., Jin, S.-Y., Xue, J., Sorvillo, N., Voorberg, J., Zheng, X.L., 2011. Essential domains of a disintegrin and metalloprotease with thrombospondin type 1 repeats-13 metalloprotease required for modulation of arterial thrombosis. *Arterioscler. Thromb. Vasc. Biol.* 31, 2261–2269. <http://dx.doi.org/10.1161/ATVBAHA.111.229609>.
- Xiao, X., Liu, Y., Zhang, X., Wang, J., Li, Z., Pang, X., Wang, P., Cheng, G., 2014. Complement-related proteins control the flavivirus infection of *Aedes aegypti* by inducing antimicrobial peptides. *PLoS Pathog.* 10, e1004027. <http://dx.doi.org/10.1371/journal.ppat.1004027>.
- Xie, X., Garg, A., Keatinge-Clay, A.T., Khosla, C., Cane, D.E., 2016. Epimerase and reductase activities of polyketide synthase ketoreductase domains utilize the same conserved tyrosine and serine residues. *Biochemistry* 55, 1179–1186. <http://dx.doi.org/10.1021/acs.biochem.6b00024>.
- Yamashita, A., Kumazawa, T., Koga, H., Suzuki, N., Oka, S., Sugiura, T., 2010. Generation of lysophosphatidylinositol by DDHD domain containing 1 (DDHD1): possible involvement of phospholipase D/phosphatidic acid in the activation of DDHD1. *Biochim. Biophys. Acta* 1801, 711–720. <http://dx.doi.org/10.1016/j.bbalip.2010.03.012>.
- Yao, C.-K., Lin, Y.Q., Ly, C.V., Ohshima, T., Haueter, C.M., Moiseenkova-Bell, V.Y., Wensel, T.G., Bellen, H.J., 2009. A synaptic vesicle-associated Ca²⁺ channel promotes endocytosis and couples exocytosis to endocytosis. *Cell* 138, 947–960. <http://dx.doi.org/10.1016/j.cell.2009.06.033>.
- Yao, X.-J., Kobinger, G., Dandache, S., Rougeau, N., Cohen, E., 1999. HIV-1 Vpr-chloramphenicol acetyltransferase fusion proteins: sequence requirement for virion incorporation and analysis of antiviral effect. *Gene Ther.* 6, 1590–1599. <http://dx.doi.org/10.1038/sj.gt.3300988>.
- Ying, M., Chen, D., 2012. Tudor domain-containing proteins of *Drosophila melanogaster*. *Develop. Growth Differ.* 54, 32–43.
- Yoon, H., Lee, H., Kim, H.J., You, K.T., Park, Y.N., Kim, H., Kim, H., 2011. Tudor domain-containing protein 4 as a potential cancer/testis antigen in liver cancer. *Tohoku J. Exp. Med.* 224, 41–46.
- Yumimoto, K., Muneoka, T., Tsuboi, T., Nakayama, K.I., 2013. Substrate binding promotes formation of the Skp1-Cul1-Fbx13 (SCF(Fbx13)) protein complex. *J. Biol. Chem.* 288, 32766–32776. <http://dx.doi.org/10.1074/jbc.M113.511303>.
- Zaidi, N., Swinnen, J.V., Smans, K., 2012. ATP-citrate lyase: a key player in cancer metabolism. *Cancer Res.* 72, 3709–3714. <http://dx.doi.org/10.1158/0008-5472.CAN-11-4112>.
- Zhang, J.-L., Huang, Y., Qiu, L.-Y., Nickel, J., Sebald, W., 2007. von Willebrand factor type C domain-containing proteins regulate bone morphogenetic protein signaling through different recognition mechanisms. *J. Biol. Chem.* 282, 20002–20014. <http://dx.doi.org/10.1074/jbc.M700456200>.
- Zhang, T., Dayanandan, B., Rouiller, I., Lawrence, E.J., Mandato, C.A., 2011. Growth-arrest-specific protein 2 inhibits cell division in *Xenopus* embryos. *PLoS One* 6, e24698. <http://dx.doi.org/10.1371/journal.pone.0024698>.
- Zhang, Z., Roe, S.M., Diogon, M., Kong, E., El Alaoui, H., Barford, D., 2010. Molecular structure of the N-terminal domain of the APC/C subunit Cdc27 reveals a homodimeric tetratricopeptide repeat architecture. *J. Mol. Biol.* 397, 1316–1328. <http://dx.doi.org/10.1016/j.jmb.2010.02.045>.
- Zhou, M.-N., Kunttas-Tatli, E., Zimmerman, S., Zhouzheng, F., McCartney, B.M., 2011. Cortical localization of APC2 plays a role in actin organization but not in Wnt signaling in *Drosophila*. *J. Cell Sci.* 124, 1589–1600. <http://dx.doi.org/10.1242/jcs>.

- 073916.
- Zhou, Y.-F., Springer, T.A., 2014. Highly reinforced structure of a C-terminal dimerization domain in von Willebrand factor. *Blood* 123, 1785–1793. <http://dx.doi.org/10.1182/blood-2013-11-523639>.
- Zhulin, I.B., Nikolskaya, A.N., Galperin, M.Y., 2003. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. *J. Bacteriol.* 185, 285–294.
- Zottich, U., Da Cunha, M., Carvalho, A.O., Dias, G.B., Silva, N.C.M., Santos, I.S., do Nascimento, V.V., Miguel, E.C., Machado, O.L.T., Gomes, V.M., 2011. Purification, biochemical characterization and antifungal activity of a new lipid transfer protein (LTP) from *Coffea canephora* seeds with α -amylase inhibitor properties. *Biochim. Biophys. Acta* 1810, 375–383. <http://dx.doi.org/10.1016/j.bbagen.2010.12.002>.
- Zou, J., Chang, M., Nie, P., Secombes, C.J., 2009. Origin and evolution of the RIG-I like RNA helicase gene family. *BMC Evol. Biol.* 9, 85. <http://dx.doi.org/10.1186/1471-2148-9-85>.