Viral infection and human disease – insights from minimotifs

Krishna Kadaveru¹,¹Jay Vyas, Martin R. Schiller¹

¹University of Connecticut Health Center, Department of Molecular, Microbial, and Structural Biology, Biological Systems Modeling Group, 263 Farmington Ave., Farmington, CT, 06030-3305

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Viruses use short functional motifs to hijack host proteins
4. Motifs mutated in human disease
   4.1. Binding and trafficking motifs
      4.1.1. PDZ binding motifs
      4.1.2. 14-3-3 binding motifs
      4.1.3. SH2 binding motifs
      4.1.4. Ubiquitin ligase binding motifs
      4.1.5. Ankyrin G binding motif
      4.1.6. IQ calmodulin binding motif
      4.1.7. SH3 binding motif
      4.1.8. Actin binding motif
      4.1.9. Arf4 binding motif
      4.1.10. Nuclear hormone receptor binding motif
      4.1.11. alpha-actinin binding motif
      4.1.12. 5-phosphoribosyl-1-pyrophosphate binding motif
      4.1.13. Trafficking motifs
   4.2. Posttranslational modifications
      4.2.1. Phosphorylation sites
      4.2.2. Protease cleavage sites
      4.2.3. N-glycosylation sites
      4.2.4. Discussion of motifs mutated in disease
5. Motif mimetic drugs
   5.1. Receptor agonists
   5.2. Posttranslational modification inhibitors
      5.2.1. Kinase inhibitors
      5.2.2. Protease inhibitors
      5.2.3. Other posttranslational modifications
5.3. Inhibitors of protein-protein interactions in treating disease
6. Perspective on short functional motifs in human health
   6.1. Molecular vulnerability of the human body to short motifs
6.2. Insights into drugs design strategies
7. Acknowledgements
8. References

1. ABSTRACT

Short functional peptide motifs cooperate in many molecular functions including protein interactions, protein trafficking, and posttranslational modifications. Viruses exploit these motifs as a principal mechanism for hijacking cells and many motifs are necessary for the viral life-cycle. A virus can accommodate many short motifs in its small genome size providing a plethora of ways for the virus to acquire host molecular machinery. Host enzymes that act on motifs such as kinases, proteases, and lipidation enzymes, as well as protein interaction domains, are commonly mutated in human disease, suggesting that the short peptide motif targets of these enzymes may also be mutated in disease; however, this is not observed. How can we explain why viruses have evolved to be so dependent on motifs, yet these motifs, in general do not seem to be as necessary for human viability? We propose that short motifs are used at the system level. This system architecture allows viruses to exploit a motif, whereas the viability of the host is not affected by mutation of a single motif.

2. INTRODUCTION

Scientific inquiry over the past century has yielded a profound increase in our understanding of human disease along with advancement in its therapy. Diseases arise from inherited genetic or epigenetic disorders that are not embryonic lethal, or from environmental factors such as toxins and pathogens. Many genetic diseases are caused by mutations in coding regions of proteins imparting either a
loss of normal function or gain of an abnormal function. Since domains are the fundamental folding and functional units of proteins, it is logical to expect that disruption of domains could lead to disease. Many such mutations in domains are observed and can lead to aberrant protein folding, aberrant aggregation, or have direct consequences on enzymatic, binding, or other functions of proteins. These mutations represent a major mechanistic class of disease-causing mutations.

In addition to domains, proteins also contain short motifs (generally shorter than 10 amino acids) that are the targets of protein domains. These motifs enable critical cellular functions such as binding to other targets, protein trafficking, regulation, and posttranslational modifications. Since proteins use these motifs for a wide array of functions, one might expect that these short motifs, like their functional domain counterparts, would commonly be exploited by pathogens and mutated in disease. This review represents the first overview and consolidation of the role of short functional motifs involved in viral infection, human disease, and drug design.

There are 1000s of published short functional motifs that are most often defined by consensus sequence motifs or by position specific-scoring matrices. Throughout this manuscript, consensus motifs are represented using the single letter amino acid (IUPAC) code and regular expression motif syntax used by Prosite (a domain database): “x” indicates any amino acid in this position, residues in “[ ]” indicate allowable residues at a position, “<” or “>” indicate location on either the N- or C-terminus, respectively (1). For example, a LxCxE consensus motif binds retinoblastoma protein. Minimotif Miner, DilinMot, Scansite, SlimDisc, and Eukaryotic Linear Motif (ELM) are web-based tools for exploring the role of short motifs in protein and proteomes (2-6).

3. VIRUSES USE SHORT FUNCTIONAL MOTIFS TO HI-JACK HOST PROTEINS

Other than the few genome-encoded enzymes, viruses must depend on their host’s machinery for nearly all aspects of their life-cycle. Therefore it is not surprising that viruses use short functional motifs for binding to host proteins and recruiting host proteins to posttranslationally modify and traffic viral proteins to specific subcellular compartments. Although some reviews have summarized how a specific motif is used by viruses, the broad-based exploitation of short motifs by viruses has not been previously summarized (7-9).

Table 1 shows a collection of motifs that are commonly used by viruses to acquire host molecular machinery. In general, many types of viruses use these motifs at all stages of their life-cycles, and utilization of these motifs is not limited to specific species, with motifs used by many different species of plants and animals. Viruses exploit the ability of host enzymes to posttranslationally modify their proteins. In addition to the sources in Table 1, several previous reviews cover the myristoylation, prenylation, N-glycosylation, phosphorylation, and cleavage of viral proteins by host enzymes (9-20).

Viruses also take advantage of a number of known short binding motifs for binding to host proteins. In addition to the sources presented in Table 1, previous reviews cover the viral use of motifs to bind host SH2 domains, PDZ domains, retinoblastoma protein (Rb), and Integrin receptors (21-27). Many viral proteins are known to have immunoreceptor tyrosine based motifs (ITAMs) that are usually found in transmembrane immune receptors.

Quite often viral consensus binding or posttranslational modification motifs are critical for the viral life-cycle and viral replication. A few of the many examples are provided. A PxQP motif in HIV Nef activates Src family kinases and is important for HIV replication (28, 29). A PPPY motif in Marbug Virus VP40 protein is important for its interaction with Tsg101 and viral budding (30). The N-myristoylation motif on the East African Cassava Mosaic Cameroon virus AC4 protein is responsible for plasma membrane targeting and pathogenicity (31). The carboxy-terminal PDZ binding motif of the E6 human papillomavirus protein is important for viral proliferation and maintenance of viral copy number (32).

There are many motifs in viral proteins that are important for the viral life-cycle but for which host targets are not yet known or perhaps mediate interaction between two viral proteins. For example, a YPPL motif in F13L vaccinia virus protein is required for efficient release of extracellular enveloped viruses (33). A KKR motif in cytoplasmic tail of Nef in Nipah Virus is important for its fusogenic activity (34). A PGQM motif in Gag is required for HIV infection (35, 36). A WxxF motif found in Vif mediates interaction with Vpr (37).

So why are short functional motifs so critical for viral function? One possibility is that viruses select motifs because motifs are often involved in the regulation of host molecules; the virus needs to deregulate pathways and then re-regulate these pathways to suit its needs. Therefore, motifs provide ready access to switching on and off specific cell processes, effectively allowing reprogramming of the cell. A single consensus motif may also provide the virus access to many factors that bind this motif. For example, a viral PxXP motif may target an array of different host factors containing SH3 domains. Short motifs may also provide a means by which the virus uses its small proteome to acquire many different host functions.

4. MOTIFS MUTATED IN HUMAN DISEASE

4.1. Binding and trafficking motifs

Since short peptide motifs are so important for the viral life-cycle and many short motifs are functional in mammalian proteins, it is a reasonable conjecture that mutation of motifs might represent a considerable fraction of disease-causing mutations in humans. However, it is not yet known the degree to which the short functional motifs are mutated in disease. We reviewed the literature and
identified 20 diseases that are associated, at least in part, with mutations that are in, or juxtaposed to short consensus motifs (Table 2).

### 4.1. PDZ binding motifs

PDZ domains bind to different classes of short consensus motifs at the C-termini of proteins (38). Ushers Syndrome and Hypercalciuria are known to be caused by mutations of PDZ binding motifs that bind to proteins with common features including cardiac abnormalities, Facial Dysmorphia and short stature. All residues in Raf-1 that are important for binding 14-3-3 (R²⁵⁶→STP²⁶¹) are mutated in Noonan Syndrome, and a S257L mutation has been observed in a patient with LEAOPARD Syndrome. Some of these mutations were shown to block binding to 14-3-3. Since 14-3-3 binding is a negative regulator of Raf-1 kinase activity, these mutations produce a gain of function in Raf-1 kinase. Ser259, a key residue in this motif that is also phosphorylated and mutated to Ala in ovarian cancer, but the effect of 14-3-3 binding to Raf-1 in ovarian cancer has not been explored (44).

### 4.1.3. SH2 binding motifs

SH2 domains bind phosphotyrosine residues. Mutation of an SH2 binding motif has not been shown to cause disease, however there are several reports that suggest SH2 motifs play a role in human disease. Diabetes mellitus can be caused by rare mutations in IRS-1 that lead to insulin resistance. Mutation of T³⁰⁶→RxxYxxM in IRS-1 alters a residue near the consensus YxxM motif for binding to the SH2 domain in the p85 subunit of phosphotyrosinesthol 3-kinase (PI3K) (45). This mutation effects association of IRS-1 with PI3K and insulin-induced PI3K activity, suggesting a role for a residue that is located four residues N-terminal to the critical tyrosine.

Likewise, a similarly important role for residues N-terminal to the tyrosine is suggested in Lupus, an autoimmune disease where the immune system attacks healthy cells (46). In systemic lupus erythematosus, two polymorphisms in Leukocyte Tyrosine Kinase (Ltk) are more prevalent in patients with disease than in control
### Table 2. Binding and trafficking motifs mutated in disease

<table>
<thead>
<tr>
<th>Disease</th>
<th>Protein</th>
<th>Motif Mutation</th>
<th>Motif function</th>
<th>Loss or gain of function</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usher Syndrome</td>
<td>SANS</td>
<td>DTxL&gt;→VTxL&gt;</td>
<td>PDZ binding motif</td>
<td>LOF</td>
<td>(41)</td>
</tr>
<tr>
<td>Idiopathic Hypercalciuria</td>
<td>Claudin 15</td>
<td>TRxV&gt;→RRxV&gt;</td>
<td>PDZ binding motif for ZO-1</td>
<td>LOF</td>
<td>(39)</td>
</tr>
<tr>
<td>Congenital Adrenal Hypoplasia</td>
<td>DAX-1</td>
<td>LLAxL→LLxPxl</td>
<td>Nuclear Receptor Binding motif</td>
<td>LOF</td>
<td>(130, 131)</td>
</tr>
<tr>
<td>Autoimmune Polyendocrinopathy-</td>
<td>autoimmune regulator</td>
<td>LxxLL→ deletion</td>
<td>Nuclear Receptor Binding motif</td>
<td>LOF</td>
<td>(63)</td>
</tr>
<tr>
<td>Systemic Lupus Erythematosus</td>
<td>Lk</td>
<td>GxxYxxM→ExxYxxM</td>
<td>PI3-kinase SH2 binding motif</td>
<td>GOF</td>
<td>(46)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>Ire-1</td>
<td>TxxxYxxM→RxxYxxM</td>
<td>PI3-kinase SH2 binding motif</td>
<td>GOF</td>
<td>(45)</td>
</tr>
<tr>
<td>Pleiotropic Malformation Syndrome</td>
<td>STRA6</td>
<td>VxxL→VxxL</td>
<td>Stat5 SH2 binding motif</td>
<td>LOF</td>
<td>(68)</td>
</tr>
<tr>
<td>Lesch-Nyhan Syndrome</td>
<td>Hypoxanthine</td>
<td>VLIxEIDTDG → VLIxEIDTDG</td>
<td>phosphoribosylpyrophosphate binding motif</td>
<td>LOF</td>
<td>(65, 66)</td>
</tr>
<tr>
<td>Idiopathic Dilated Cardiomyopathy*</td>
<td>Nebulette</td>
<td>SxKxxYK or SxKxxYK</td>
<td>Actin binding motif</td>
<td>LOF</td>
<td>(59)</td>
</tr>
<tr>
<td>Retinitis Pigmentosa</td>
<td>Rhodopsin</td>
<td>sxxYK→ sxxYK or sxxYK</td>
<td>Arf6 binding motif</td>
<td>LOF</td>
<td>(60, 61)</td>
</tr>
<tr>
<td>Noonan Syndrome</td>
<td>Raf-1</td>
<td>RxxSTP→RxxSTP</td>
<td>14-3-3 binding motif</td>
<td>GOF</td>
<td>(43)</td>
</tr>
<tr>
<td>LEOPARD Syndrome</td>
<td>Raf-1</td>
<td>RxxSTP→RxxSTP</td>
<td>14-3-3 binding motif</td>
<td>GOF</td>
<td>(43)</td>
</tr>
<tr>
<td>Ovarian Cancer</td>
<td>RAF-2</td>
<td>RxxSTP→RxxSTP</td>
<td>14-3-3 binding motif</td>
<td>GOF</td>
<td>(44)</td>
</tr>
<tr>
<td>Myobrmial Myopathy</td>
<td>ZASP</td>
<td>26 residue motif</td>
<td>α-actinin binding motif</td>
<td>LOF</td>
<td>(42)</td>
</tr>
<tr>
<td>Colorectal carcinogenesis, Melanomas</td>
<td>β-catenin</td>
<td>DSxGHDs→DSxGHDs</td>
<td>β-TrCP ubiquitin ligase adaptor recognition domain</td>
<td>LOF</td>
<td>(53, 54)</td>
</tr>
<tr>
<td>Liddle’s Syndrome</td>
<td>ENaC β subunit</td>
<td>PPSxY→PSxY, PPSxY, PPSxY, PPSxY, PPSxY</td>
<td>Nedd4 WW domain binding motif</td>
<td>LOF</td>
<td>(49, 50)</td>
</tr>
<tr>
<td>Brugada Syndrome</td>
<td>Nav 1.5 Channel</td>
<td>VPxAxESD→VPIAxKSD</td>
<td>Ankyrin G binding motif</td>
<td>LOF</td>
<td>(55)</td>
</tr>
<tr>
<td>Rheumatoid Arthritis</td>
<td>PTP-N22</td>
<td>PPLPxxTRxxxIV → PPLPxxTRxxxIV</td>
<td>Csk SH3 domain binding motif</td>
<td>LOF</td>
<td>(57)</td>
</tr>
<tr>
<td>Tricho-rhino-phalangeal Syndromes</td>
<td>TRPS1 transcription factor</td>
<td>RRRTRKR→RRRTRKR</td>
<td>Nuclear localization motif</td>
<td>LOF</td>
<td>(67)</td>
</tr>
</tbody>
</table>

* Polymorphisms are risk factors for this motif.

Subjects (46). These mutations surround a Y753xxM motif for binding the SH2 domain of the p85 subunit of PI3K. One polymorphism results in a gain of function mutation of G750xxYxxM→E750xxYxxM that enhances binding to PI3K, suggesting that residue 750 is important for SH2 domain binding. In support of this observation, examination of the structure of the complex between PI3 kinase SH2 domain and a cKit peptide that contains the sequence TxxYxxM shows contacts between the Thr residue of cKit and Asn56 of the SH2 domain (47).

Pleiotropic Malformation Syndrome results from mutations in STRA6 a protein identified based on its...
Short functional motifs in human pathogenesis

retinoic acid-induced expression (48). A Y^545LLL^646 →Y^545MLL mutation in STRA6 has been identified in a consensus binding motif for the Stat5 SH2 domain, however the effect of this mutation of Stat5 and Stat5 signaling needs to be validated.

4.1.4. Ubiquitin ligase binding motifs

Liddle’s Syndrome presents with salt-sensitive hypertension and is associated with mutations in the β and γ subunits of the epithelial sodium channel (ENaC). The majority of missense mutations are concentrated in a cytoplasmic P^516LPPxX^518 motif near the C-terminus of the β subunit (49, 50). Mutation of this motif increases the channel activity several fold and leads to retention of the activated channel at the cell surface. This short motif binds to the ubiquitin ligase Ned44 through its WW domain and plays a role in stability of the channel (51, 52). Ubiquitin ligation is also effected in melanoma cell lines, where a D^275SGHDS^277 →DSGHDF mutation has been observed in β-catenin. The mutated SxxxxS consensus motif in this sequence is normally phosphorylated and targeted for degradation by β-TrCP, an adaptor for a ubiquitin ligase (53, 54).

4.1.5. Ankyrin G binding motif

Brugada Syndrome is a fatal cardiac arrhythmia where an E1053K mutation has been identified in the Nav 1.5 sodium channel (55). This mutation alters a V^1044PIQxxxESD^1055 motif in the cytosolic loop that binds to an unknown target on the surface of cardiomyocytes.

4.1.6. IQ Calmodulin binding motif

Mutations in Myosin 7A have been identified in patients with Nonsyndromic Hearing Loss (56). A R853C mutation disrupts an IQ motif (V^843QxxxRGxxxR^853) known for binding to calmodulin. Experiments examining Calmodulin dependent vasoconstriction support a functional role for this motif in Calmodulin binding and a loss of function in hearing loss.

4.1.7. SH3 binding motif

Mutations in Protein Tyrosine Phosphatase (PTPN22) have been identified in patients with Rheumatoid Arthritis (57). The R620W mutation alters the P^615PLPxRTxxxxIV^627 consensus motif for binding the Csk SH3 domain and reduces its affinity for Csk (58).

4.1.8. Actin binding motif

Polymorphisms in the Nebulette protein are associated with an increased risk for Idiopathic Dilated Cardiomyopathy (59). Two of these polymorphisms (Asn654 or Lys654 are located in a actin binding motif S^653[DEQNS]xxYK^658) and may affect binding to actin (59).

4.1.9. Arf4 binding motif

Mutations in the C-terminus of Rhodopsin cause Retinitis Pigmentosa (60, 61). Missense mutations (V345M and P347S) within, and a truncation mutation that eliminates a five residue xVxPx motif from the C-terminus, blocks binding of Rhodopsin to the Arf4 GTPase, an interaction important for trafficking of Rhodopsin to the rod outer segment.

4.1.10. Nuclear hormone receptor binding motif

Mutations in two similar hormone receptor binding motifs have been identified. L295P and L297P mutations in the L^284LxLxL^299 motif in DAX-1 were identified in Congenital Adrenal Hypoplasia. This motif is thought to be involved in recruitment of transcription factor co-regulators. The more conventional motif for nuclear hormone receptor binding is a LxxLL (62). While not conclusive, a frameshift mutation at position Ala^505 near the C-terminus of autoimmune regulator protein deletes a L^516xxLL^520 motif which may be involved in Autoimmune Polyendocrinopathy-candidiasis-ectodermal Dysplasia (63).

4.1.11. α-actinin binding motif

A 26 amino acid region of ZASP contains a ZM motif with the consensus definition of Q[YF]NxPx[ML]YSxxx[IL] (64). Mutations A147T and A165V surrounding this motif are mutated in ZASP patients with Myofibrillar Myopathy (42). While the ZM motif binds to α-actinin, it is not known whether these mutations impair α-actinin binding.

4.1.12. 5-phosphoribosyl-1-pyrophosphate binding motif

Hypoxanthine-guanine Phosphoribosyltransferase is an enzyme that binds 5-phosphoribosyl-1-pyrophosphate through a V^1341LIVEDHIDTG^1350 motif in its active site (65, 66). In Lesch-Nyhan Syndrome, 11 different mutations spanning the majority of these residues abrogate or impair the enzymatic activity.

4.1.13. Trafficking motifs

Several protein trafficking motifs are known to be mutated in disease. TRPS1 is a transcription factor that is deleted or mutated in Tricho-rhino-phantalangeal Syndromes (TRPS) (67). The last Arg of known RRRTRKR nuclear localization motif is mutated. R952H or R952C mutations in these patients blocks nuclear translocation of TRPS1 in a transfected cell line. In HCN2 and Ether-a-go-go Related Gene (HERG), a C-terminal RxR motif is involved in endoplasmic reticulum retention (68, 69). In several patients with long QT syndrome the C-terminus is truncated, and mutation of these residues reduces channel conductance in transfected cells, thus the effects of the deletions are thought to be due to defective trafficking of the mutant protein through this ER retention motif. Although not directly mutated, a C-terminal dileucine motif involved in trafficking is lost in C-terminal deletions of ATP7A in Menkes Disease and a C-terminal peroxisomal SKL> targeting motif is lost in Maclonal-CoA Decarboxylase, the gene mutated in Malonyl-CoA Decarboxylase Deficiency (70, 71).

4.2. Posttranslational modification motifs

Several mutations in phosphorylation sites, protease cleavage sites, and N-glycosylation sites are known to play a role in disease.
4.2.1. Phosphorylation sites

Kinases are one of the largest gene families in mammalian proteomes, and it is not surprising that deregulation of kinase activity is a common theme in disease. One treatment strategy has been to develop inhibitors of kinases. This has been particularly effective in cancer where drugs inhibiting a number of different kinases are effective and in current practice. Most well-studied kinases are known to phosphorylate multiple substrates, thus it is not known whether the efficacy of these drugs is due to broad inhibition of phosphorylation of many substrates or through selective inhibition of one or several key regulatory phosphorylation events.

Given that many phosphorylation consensus sequences can be abolished or created by introduction of a missense mutation and the fact that kinase inhibitors are effective for diseases, we expected that there would be many examples of disease causing mutations in consensus sequences of kinase substrates that affect substrate phosphorylation. While this is a very broad subject and it is difficult to be comprehensive, a thorough literature search revealed only seven known diseases in which mutation of phosphorylation sites were known to be mutated in disease (Table 3).

Patients with familial advanced sleep phase syndrome contain a S662G mutation in the Per2 gene which is within a domain that binds to Casein Kinase Iε (72). This mutation blocks CK1ε mediated phosphorylation of Per2. Maturity-onset Diabetes of the Young can be caused by a gain of function G115S mutation in the hepatocyte nuclear factor 4α transcription factor. Mutation of T58 in cMyc is a mutational hotspot in many lymphomas. A S470N polymorphism in Synapsin III is observed (48). This mutation causes a R655C mutation or an insertion after R1111; sequence with a T→I mutation or an insertion after R1111, sequence for Calmodulin Kinase II phosphorylation (77). Several of these mutations cluster around a R1111EMT1114 sequence with a T→I mutation or an insertion after R1111, sequence for Calmodulin Kinase II phosphorylation (77). This mutation leads to reduced phosphorylation of IKAP in biosynthetic labeling experiments. In Plieotrophic Malformation Syndrome, a homoyzgyous mutation in STRA6 is observed (48). This mutation causes a R655C residue change that eliminates a consensus phosphorylation site (RxT) for Protein Kinase A.

We were surprised at how few phosphorylation sites were known to be mutated in disease. This result supports the hypothesis that deregulation of a kinase results in phosphorylation of multiple downstream targets and that more than one target is necessary for causing disease. We can therefore generalize that activating a kinase likely causes phosphorylation of multiple downstream targets and can cause disease, but phosphomimetic mutations of a
single phosphorylation site that is a target of one of these kinases does not lead to disease with the exception of the gain of function site in Hepatocyte nuclear factor 4α transcription factor as noted above. This suggests that drugs that target kinases act at a systemic level by blocking multiple important phosphorylation events.

4.2.2. Protease cleavage sites

As discussed above for kinases, another general class of therapeutic compounds that block enzyme activity is inhibitors of proteases. For example, inhibitors targeting the HIV protease are one of the currently used strategies to control HIV infection (78). Most proteases cleave many substrates. For example, all prohormones are processed at dibasic cleavage sites by a small number of Prohormone Convertases.

We wanted to know if there were any diseases that resulted from incorrect protease processing of a specific site. Five examples were identified and are shown in Table 3. FGF23 is processed from a prohormone at a R216L sequence to activate the zymogen. A R216L mutant of APP has a KM → NL mutation in the β-secretase cleavage site and are involved in Alzheimer’s disease (87). Thus, altered processing of APP identifies another motif targeted in disease. While the Swedish mutant of APP has a KM→NL mutation in the β-secretase cleavage site, these changes do not alter the proteolytic activity toward this site (88).

4.2.3. N-glycosylation sites

Aberrant N-glycosylation seems to be common in disease. There are 32 known gain of function N-glycosylation mutations that span a wide array of disease and have previously been reviewed (89). There are a few loss of function glycosylation mutations. In Retinitis Pigmentosa a Ser/Asn substitution in the NxS N-glycosylation consensus sequence eliminates an N-glycosylation site. Thus, altered processing of APP identifies another motif targeted in disease. While the Swedish mutant of APP has a KM→NL mutation in the β-secretase cleavage site, these changes do not alter the proteolytic activity toward this site (88).

4.2.4. Discussion of motifs mutated in disease

Our review of the literature identified 20 binding motifs and 15 motifs for posttranslational modifications that are mutated in human disease. There were no unifying themes regarding type of motif, disease, organ, etc. Since many motifs are known to cause disease, the 35 binding and posttranslational modification motifs summarized in this review suggest that a disproportionately small number of short functional motifs are mutated in disease. However, more than 32 N-glycosylation motifs are associated with disease (89). Since N-glycosylation motifs play an important role in folding of proteins that transit through the secretory pathway, this motif is unique among short consensus motifs because it plays a role in protein folding and misfolding is a major disease mechanism. There are a number of possible explanations as to why so few motifs are mutated in disease which are further discussed in section 6.

5. MOTIF MIMETIC DRUGS

5.1. Receptor agonists

The majority of drugs approved by the Food and Drug Administration (FDA) target receptors and enzymes. A few peptide hormone receptors bind to short peptide hormones that can be considered motifs. Examples include...
Short functional motifs in human pathogenesis

Figure 1. Chemical structures of FDA-approved motif mimetic drugs. Comparisons of the chemical structures of small compounds (Morphine and Aggrastat® (Tirofiban)) to the peptide motifs they mimic (Enkephalins and RGD Integrin ligands), respectively. The R group in Enkephalins can be Met for Met-enkaphalin, Leu for Leu-enkephalin, or other peptide for longer Enkephalin peptides such as β-endorphin.

Enkephalins, Vasopressin, Oxytocin, Thyroid Releasing Hormone, Cholecystokinin, and Somatostatin-14 peptides. In addition, evolutionary conservation of sequences for peptide hormones often shows a conserved cluster of ~5 residues, suggesting that a short motif core of peptide hormone residues may be responsible for interaction with their receptors. Consistent with this hypothesis mapping of important contact residues for growth hormone binding to the Growth Hormone Receptor shows a similar size core binding motif (93, 94). In several cases, peptide hormones are therapeutically used: Somatuline Depot® (Histrelin) is a Gonadotropin Releasing Hormone (GnRH) agonist used for treating the early onset of puberty. Demerol® (Meperidine) is an Opiate agonist for the κ-opioid receptor; Morphine, Fentanyl, Methadone, Oxycodone, and Codeine are agonists for the μ-opioid receptor. A comparison of Morphine as a mimic of Enkaphalin peptides is shown in Figure 1.

5.2. POSTTRANSLATIONAL MODIFICATION INHIBITORS

5.2.1. Kinase inhibitors

Despite the many important functions of protein phosphorylation in cells, there are no known drugs that directly target protein phosphorylation sites in proteins. However, over the past decade a number of kinase inhibitors have emerged for treating various cancers and immunosupression. Stutnet® (Sunitinib) is a receptor tyrosine kinase (RTK) inhibitor for treating gastrointestinal tumors. Similarly, Sprycel® (Dasatinib) is a broad-based tyrosine kinase inhibitor that inhibits Bcr-Abl, Src, and other protein tyrosine kinases for treating chronic myeloid leukemia (CML) and acute lymphoblastic leukemia (ALL). Gleevec® (Imatinib Mesylate) is a general protein tyrosine kinase inhibitor approved for use in CML and Iressa® (Gefitinib) is a tyrosine kinase inhibitor for lung cancer (95). Another broad-based Ser/Thr and Tyr kinase inhibitor is Nexavar® (Sorafenib) which is used for advanced Renal Cell Carcinoma. Tarceva® (Erlotinib) is an Epidermal Growth Factor Receptor inhibitor used for treating Non-small Cell Lung Cancers (96). The majority of these small compound inhibitors of kinase activity are relatively non-specific, yet therapeutically effective. A number of neutralizing antibodies to specific RTKs are also in clinical use for various types of cancer. These have advantages in specificity, but therapeutic antibodies are limited by delivery when compared to small molecule drugs. Beyond cancer, Siromilus® (Rapamycin) is used to block organ rejection by inhibiting mTOR, a kinase involved in cytokine-stimulated T-cell proliferation (97).

Protein kinases, including the targets of most of the aforementioned inhibitors, generally phosphorylate many cellular proteins. Since many of these inhibitors have specificity limited to a precise subset of kinases in the kinome, it is likely that these drugs inhibit multiple phosphorylation events in cells. It is not yet clear whether the benefit of these drugs is from a single phosphorylation site or inhibition of multiple sites. Since very few known phosphorylation site mutations lead to disease (previous section) and no known drugs target phosphorylation sites, the actions of these drugs are likely mediated through perturbation of a number of different phosphorylation sites and act at a systemic level.

Non-peptide agonists have been identified that bind to the Insulin, Thrombopoietin, Wassopressin, Somatostatin, Bradykinin, Cholecystokinin, Angiotensin, Melatonin, Growth Hormone, Opioid, and Glucagon-like Peptide 1 receptors. Thus, in the case of receptors small molecule agonists that mimic core regions of peptide hormones seems to be a viable strategy for drug design. Several examples are provided. Supprelin LA® (Histrelin) is a Gonadotropin Releasing Hormone (GnRH) agonist used for treating the early onset of puberty. Demerol® (Meperidine) is an Opiate agonist for the κ-opioid receptor; Morphine, Fentanyl, Methadone, Oxycodone, and Codeine are agonists for the μ-opioid receptor. A comparison of Morphine as a mimic of Enkaphalin peptides is shown in Figure 1.
Short functional motifs in human pathogenesis

5.2.2. Protease inhibitors

Proteases cleave recognition motifs in protein substrates. Several mutations in protease motifs lead to disease (previous section), suggesting that at least some proteases are specific and might make good targets for drugs. There are several FDA approved protease inhibitors. Tritace® (Ramipril), Vasotec® (Enalapril), Accupril® (Quinapril), Monopril®, Monopril® (Losartan), and Lotensin® (Benzapril) are a substrate mimic Angiotensin-converting Enzyme inhibitors (98). Novastan® (Argatroban) is a Thrombin inhibitor. Januvia® (Sitagliptin) is a DPP-4 dipeptidyl peptidase enzyme inhibitor for treating Type II Diabetes. There are numerous inhibitors of the HIV proteases and this approach has been one of the principal means for treating Autoimmune Immunodeficiency Syndrome (AIDS) (78).

5.2.3. Other posttranslational modifications

Many proteins are modified by lipidation of specific motifs in the amino and carboxy-termini of proteins. Two drugs in this class are approved by the FDA. Zovirax® (Acyclovir) is an antiviral myristoylation inhibitor that blocks N-terminal myristoylation of viral proteins, again highlighting the critical importance of motifs in viral infection. Valtrex® (Valacyclovir) is rapidly converted to Acyclovir and is for treating herpes simplex virus. Another lipidation motif is a CAAX> box (A = aliphatic residue) which is on the C-terminus of proteins that are modified by covalent attachment of farnsyl or geranylgeranyl group on the Cys residue. This modification is catalyzed by Farnesyl Transferase I and Geranylgeranyl Transferases I and II, respectively. Inhibitors for these enzymes are being explored for treatment of cancer (99, 100). Farnsyl Transferase inhibitors reached Phase III trials, but, by themselves produced modest effects and are now being explored for treatment in combination with other drugs and are currently in clinical trials. Inhibitors of the Geranylgeranyl Transferases are also being explored (12, 100).

5.3. Inhibitors of protein-protein interactions in treating disease

Despite an enormous research effort toward understanding and inhibiting protein-protein interactions, no drugs that disrupt intracellular protein-protein interactions are currently approved by the FDA for treatment of disease. This is likely due, in part, to the focus of big pharmaceutical companies on receptors, ion channels, and enzymes. Notwithstanding, this is an active area of research.

Potential extracellular targets are the Integrin Receptors. Proteins with an RGD motif bind to approximately ½ of the 20 known Integrin Receptors; these receptors mediate focal adhesion attachment to the extracellular matrix. Investigators have attempted to exploit this motif in various ways, most of which have been extensively reviewed elsewhere, but are summarized below.

Two RGD mimetic drugs are antagonists of the Integrin αIIbβ3 (Glycoprotein IIb/IIIa Receptor), which is involved in platelet thrombosis. Eptifibatide® (Integrilin) is a cyclic hexapeptide containing the RGD sequence that binds to the Glycoprotein IIb/IIIa Receptor (101-105). Co-administered with aspirin or Clopigoidrel (low molecular weight heparin), this cocktail is used to treat Unstable Angina or a Non-ST-segment-elevation Myocardial Infarction. Similarly, Aggrastat® (Tirofiban) is a low molecular weight synthetic RGD mimetic (Figure 1) used to treat Unstable Angina and Non-Q-wave Myocardial Infarction (101).

In addition to its approved medical uses, the RGD motif is also being explored in several ways for treatment of cancer. RGD peptides can also be used to treat tumors by targeting toxins, radioactive chemicals, antimitic drugs, and DNA replication inhibitors to tumors (106-113). Similarly, tumors and tissues can be visualized using RGD peptide conjugates (109, 108, 114). Cilengitide is a RGD cyclic pentapeptide in Phase II trial that inhibits angiogenesis (115-117). Other uses of RGD peptides are in increasing adhesion to solid supports for increasing adhesion of medical implants and grafts, and in treatment of Acute Renal Injury, and Osteoporosis (118-122).

Several other motifs are being explored as drug targets. Aminopeptidase N is a protease expressed on the cells surface of tumor blood vessels. A hexapeptide/Tumor Necrosis Factor conjugate containing a NGR motif acts as a tumor-homing conjugate (123). Some motifs that bind to intracellular domains are being explored as drug targets. PDZ motifs have well defined binding clefts and proteins containing PDZ domains are involved in pain perception, Epilepsy, Cancer, Schizophrenia, and Parkinson’s disease (124). Drugs that target PDZ-mediated interactions are under investigation. High affinity peptides that bind the SH3 domain of the Crk adaptor are under investigation as anticancer drugs (125). Peptides targeting the Grb2 and Crk SH2/SH3 adaptors are under investigation for treating several diseases (126). NK109 is an antitumor drug that binds to a PNxxxxP motif present in the C2 domain of Protein Kinase Ca (127). Nutlin-3 is a small molecule inhibitor that blocks the motif-driven interaction of MDM2 and MDMX with p53 (128). Treatment of mouse models of Retinoblastoma with Nutlin-3 and a Topoisomerase inhibitor Hycamatin® (topotecan hydrochloride) drastically reduced tumor burden.

6. PERSPECTIVE ON SHORT FUNCTIONAL MOTIFS IN HUMAN HEALTH

6.1. Molecular vulnerability of the human body to short motifs

Investigation of disease causing mutations, drugs, pathogens, and disease models teach us about the vulnerability of the human body. Disease causing mutations and FDA approved drugs suggest that enzymes, hormones, and receptors are particularly vulnerable sites. We questioned whether short functional motifs are also a point of human vulnerability. With the exception of N-glycosylation motifs which effects protein folding in the endoplasmic reticulum, our literature analysis revealed that few short functional motifs are known to be mutated in
Short functional motifs in human pathogenesis

human disease. We must consider several possible explanations: 1. mutation of many short motifs are embryonic lethal and most individuals harboring such mutations die 2. short motifs have unimportant functions 3. short motifs are important at the systemic level, but any single motif rarely effects viability 4. the majority of functional motifs have not been identified so that we are unaware of their role in disease, or 5. the majority of diseases have not been identified so we are not aware of the role of motifs in the yet-to-be-discovered disease.

To gain a brute-force understanding at the genome level and address these possible explanations, we examined some statistics of mouse models using the Mouse Genome Database (129). Of the 4,289 genes thus far deleted in targeted knock-out mice, 1,897 are lethal, suggesting that ~44% of mouse genes are necessary for survival. Of the 1,897 genes required for mouse viability, 1,742 were lethal at embryonic, perinatal, or postnatal stage and another 155 were lethal at a later stage in life. These 155 genes produce viable organisms with defects that cause death later in life and can be considered an equivalent of disease in humans. By extrapolation, the disease causing genes represent only a small percentage of mouse genes.

While we cannot rule out that a large proportion of human motif mutations result in embryonic lethality, this seems unlikely in consideration of the small proportion of genes that cause disease in mice. Although mutation of ~44 % of mice genes results in lethality, motifs often only diminish function of a protein and we did not identify any cases where mutation of a motif is similar to a null phenotype for the gene. We are in the early stages of short motif discovery and it is possible that many more mutated motifs will be identified as more motifs and more specific motif definitions are accumulated. In our literature analysis, a common theme was an observation that residues juxtaposed to a consensus sequence, or within a portion of a motif not thought to impart specificity, were mutated. This suggests that current motif definitions may not be exact and need improvement. However, to this point it seems like mutation of motifs that result in disease is a rare occurrence.

We favor the hypothesis that most motifs are important at the systemic level, but mutation of a single motif does not perturb the system enough to affect viability. Since few intracellular motifs are mutated in disease (sections 3 and 4), it appears that human cellular physiology is not critically dependent on any one motif for function. For example, if a kinase phosphorylates 50 targets and five are critical for viability then knocking out the kinase will cause death, but knocking out any one phosphorylation site will not. In this scenario, all five critical phosphorylation sites would need to be mutated to cause death. Effectively, this system architecture reduces vulnerability of the organism. However, opportunistic pathogens such as viruses could readily exploit cellular systems that are driven through short motif recognition and this seems to be the case (see section 2). There are a couple of specific cases, where this does not hold true. Asparagine-linked glycosylation motifs are involved in the protein folding quality control pathway of the endoplasmic reticulum and both gain and loss of function mutations for this motif are found in disease (89). We think this is a point of vulnerability because of its role in effecting folding, as misfolding is a primary disease mechanism.

The other case is that for extracellular peptide hormones that have core motifs that bind to receptors. Many of the cell-to-cell signaling systems are not redundant, vulnerable to mutation in disease, and may be one reason why such a larger proportion of drugs target receptors and peptide hormones.

One consideration regarding our current knowledge of motifs is that motifs may need to be much more specific than currently motif definitions describe. Many motifs have highly redundant motif definitions and are found at greater than 100,000 copies in the human proteome; clearly not all of these putative sites are used. This over-prediction of short motifs may be due to the fact that motifs must work within a specific context such as being on the protein surface, be localized to a specific compartment, or being expressed with its target in the same cell or during a specific point of development. This may hinder our ability to recognize motifs that are mutated in disease.

6.2. Insights into drugs design strategies.

The majority of FDA-approved drugs target enzymes, receptors, and channels. Many of these drugs are screened based on a specific target. Most of these targets are likely to have non-redundant functions in cells. Only a few drugs that mimic short motifs exist (Figure 1) and in general, drugs that block protein-protein interactions are not routinely used in treatment of disease. This could be because pharmaceutical companies do not explore protein-protein interactions as potential drug targets. However, as covered in Section 5, there has been reasonable interest in exploring a number of motifs as druggable targets. Since viruses have evolved to use motifs for essential functions by hijacking host proteins, one potential area where motif mimetic drugs may be useful is for treating viral infection and infection by other pathogens. Another way that motif mimetic drugs may be useful is as a cocktail targeting multiple targets. For example, an inhibitor of a tyrosine kinase can disrupt phosphorylation of many proteins and processes. If the few critical phosphorylation sites were known, a cocktail of motif mimetic drugs that that block phosphorylation of the key sites might be just as effective as the tyrosine kinase inhibitor, but present less problems with toxicity through non-desirable effects of the drug on other phosphorylation sites and cell processes. Such drug cocktails may be more effective by perturbing the systems in different ways. For example, in controlling the spread of HIV, a protease inhibitor cocktail that targets the HIV protease and reverse transcriptase is more effective than a single protease inhibitor. Much remains to be learned about motifs, but given their widespread role in cell function and physiology, they are like to be part of the future of drug design.
7. ACKNOWLEDGEMENTS

We would like to thank Dr. Michael Gryk for critically reading this manuscript and want to acknowledge members of the Minimotif Miner team for intellectual discussions related to topics of this review. We thank the NIH (GM079689 to M.R.S.) and the University of Connecticut Partnership for Excellence in Structural Biology (K.K.) for funding. Krishna Kadaveru and Jay Vyas contributed equally to this work.

8. REFERENCES


Short functional motifs in human pathogenesis


Short functional motifs in human pathogenesis


Short functional motifs in human pathogenesis


Short functional motifs in human pathogenesis


Key Words: Motif, Disease, Virus, Drugs, Posttranslational modification, Review

Send correspondence to: Martin R. Schiller, University of Connecticut Health Center, Department of Molecular, Microbial, and Structural Biology, Biological Systems Modeling Group, Farmington, CT 06030-3305, Tel: 860-679-4610, Fax: 860-679-1726, E-mail: Schiller@nso.uchc.edu

http://www.bioscience.org/current/vol13.htm