More than Meets the Eye: Hidden Structures in the Proteome

Hal Wasserman[1] and Erica Ollmann Saphire[1,2]

[1]Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California 92037; email: erica@scripps.edu

[2]Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037

Annu. Rev. Virol. 2016. 3:373–86

First published online as a Review in Advance on July 22, 2016

The *Annual Review of Virology* is online at virology.annualreviews.org

This article's doi: 10.1146/annurev-virology-100114-054923

Keywords

transformer, metamorphic protein, morpheein, moonlighting protein, matrix, Ebola virus

Abstract

A central dogma of molecular biology is that the sequence of a protein dictates its particular fold and the fold dictates its function. Indeed, the sequence → structure → function hypothesis has been a guiding principle by which scientists approach molecular biology. Every student knows that the genome encodes information for the progression from primary sequence to secondary, tertiary, and ultimately quaternary structure. Yet with a growing number of proteins, a fifth level has been identified: rearrangement of existing structures into distinct forms. Recent observations indicate that replication of Ebola virus depends on this fifth level. We believe other viruses with compact genomes and rapid evolution under selective pressure will be a rich source of examples of polypeptides that rearrange to gain added functions. In this review, we describe mechanisms by which viral, prokaryotic, and eukaryotic polypeptides have adopted alternate structures to control or gain function.

INTRODUCTION

As molecular biologists, we have learned a central dogma: that a protein sequence defines its singular and particular structure, which in turn defines its function. When a structure has been solved and made available in the Protein Data Bank (PDB), we assume we may use this structure as a template to interpret the protein's biological function(s), and we base our subsequent experiments on the structure. Recent observations, however, have identified proteins that do not conform to that canon. Instead, they yield multiple, independent structures, each of which confers a unique biological function. For these proteins, the phenotype is directly connected to the fold, yet the existence of different folds was not easily predictable.

We may be likely to come across such proteins more frequently in viruses than in cellular organisms. Viruses often have error-prone replication. They have a more rapid life cycle than their cellular hosts. To survive, they must adapt to selective pressure from immune surveillance and potentially to different cellular environments. Many contain very few genes (arenaviruses have four, paramyxoviruses as few as six, filoviruses seven), yet their limited set of gene products must achieve more than four, six, or seven functions in the virus life cycle. These viruses' inherently low genomic complexity, coupled with selective pressure, may press evolution of multifunctionality from each of the few gene products they encode. Plasticity in the physical structures encoded by their genomes may be particularly advantageous: Single polypeptides that adopt multiple forms could allow more functions to be achieved from fewer genes. Indeed, in Ebola virus, a serendipitous discovery of the particularly dramatic conformational changes achieved by its matrix protein indicated to us that structural rearrangement to acquire additional functions could occur on a larger scale than generally assumed. Findings like these provide a compelling launchpad to consider whether such proteins exist in other viral genomes.

A FUNCTION MULTIPLIER, THE FILOVIRUS MATRIX PROTEIN VP40

The filoviruses are enveloped, pleomorphic viruses. One of their seven genes is *VP40*, which encodes the protein that assembles the matrix underneath the viral membrane and gives the virus its characteristic filamentous shape (1, 2). Expression of the VP40 protein alone is necessary and sufficient for assembly and budding of virus-like particles with the morphology of authentic filoviruses (3). But VP40 also plays a role in regulating viral transcription and replication, although the precise mechanism is still unclear (4). Marburg virus VP40 also suppresses host interferon signaling (5).

The first crystal structure of Ebola virus VP40, determined in the year 2000, described the fold of a VP40 monomer (6). An additional structure showed how VP40 could assemble into octameric rings, with RNA serendipitously found in electron density maps at the center of the ring (7, 8). The team of researchers that solved this structure showed that mutations that prevent ring assembly did not affect virus assembly. Hence, the roles of the ring and of the bound RNA were not yet clear, and the authors proposed that there could be an "additional function in the life cycle of the virus besides promoting virus assembly and budding off the plasma membrane" (7, p. 423). A second function for VP40 was not known for seven more years (4). In terms of virus assembly, the ring structure was the only model available for how copies of VP40 might oligomerize, and yet there was already evidence to the contrary (7).

When we began work on VP40, we had expressed and purified it in order to use it to pull down a different protein of interest. We purified VP40 without a fusion partner, using size exclusion chromatography coupled to multiangle light scattering (SEC-MALS). Surprisingly, we observed that VP40 came off the column at ~70 kDa, not 33 kDa. We were curious about this unexpected

molecular weight and crystallized the protein. The fold, space group, and unit cell were identical to those solved in 2000 (6). However, considering that VP40 could be a dimer in solution, we looked into the crystal packing for potential dimer interfaces. We generated point mutations at each of them and found that an L117R point mutation at the N-to-N interface generated monomeric VP40 in solution. Therefore, the L117-involving N-to-N interface was probably the dimer interface in solution. More importantly, transfected VP40 bearing that same point mutation was unable to bud virus-like particles from cells, suggesting that the observed VP40 dimer is a building block of the viral matrix (9).

We crystallized VP40 in three distinct space groups. Notably, in each of these, the dimers assembled into long linear filaments by nearly identical C-to-C interfaces. Mutagenesis of this conserved C-to-C interface by which dimers build these conserved filaments indicated that that interface was also critical to matrix assembly.

It was tempting to predict that the linear filament observed in crystal packing built the filamentous viral matrix. Two pieces of biological information, however, suggested otherwise. A model like our initial one, in which the matrix is built by subsequent addition of dimers, should yield oligomers of two, four, six, eight, ten, etc. when VP40 interacts with lipid membranes. However, it had instead been reported that incubation of VP40 with liposomes induced assembly into hexamers, with oligomers observed in a size pattern of six, twelve, eighteen, etc. (10). Furthermore, in the filament we observed in the structures, the N- and C-terminal domains remained associated with each other. The second piece of evidence in the literature was that association of VP40 with membranes triggers a conformational change (11). Our initial filament model had no conformational change in VP40, suggesting that an additional, different form of VP40 built the viral matrix.

We wondered whether the side-by-side association reflected some initial state, upon which membrane interaction would drive a conformational change and assembly into hexameric building blocks. Indeed, we found that dextran sulfate, which outcompetes the phosphatidylserine ligand of VP40 in the cell membrane (10, 12), caused VP40 to conformationally rearrange and to reassemble into hexamers. In the hexamer, the central four molecules have undergone separation of their N- and C-terminal domains. The N-terminal domains interact by the N-to-N dimer interface and an additional N-to-N interface revealed upon rearrangement of the C-terminal domain. Hexamers assemble with each other to make zigzagging filaments by the same C-to-C interface previously observed. These hexamers exhibit dimensions similar to those observed for the matrix of actual virions.

It thus appeared that wild-type VP40 made dimer, octamer, and hexamer structures, with rearrangement of both tertiary and quaternary structure among them (**Figure 1**). In the course of this analysis, point mutations were discovered that would force or prevent single structures of VP40. Because assays for different functions of VP40 (virus assembly, transcriptional repression, etc.) had already been worked out by others (3, 13–15), the point mutations became tools by which functions could be assigned to structures: The dimer is critical for cellular trafficking and is a precursor for viral assembly. The rearranged hexamer is a building block of the viral matrix, and the octameric ring binds RNA and controls transcription. Studies from the Kawaoka lab identified a temporal and spatial location for each VP40 structure (16). Early in infection, perinuclear VP40 dominates; we believe this to be the octameric ring. Later in infection, cell-surface VP40 predominates; we believe this to be filamentous matrices forming from the hexameric building blocks. Control of the different structures and functions of VP40 is probably achieved by some yet-to-be-identified biological trigger. It is not yet known whether this trigger is a viral factor, a host factor, or the influence of increasing concentration of VP40 at different time points after infection.

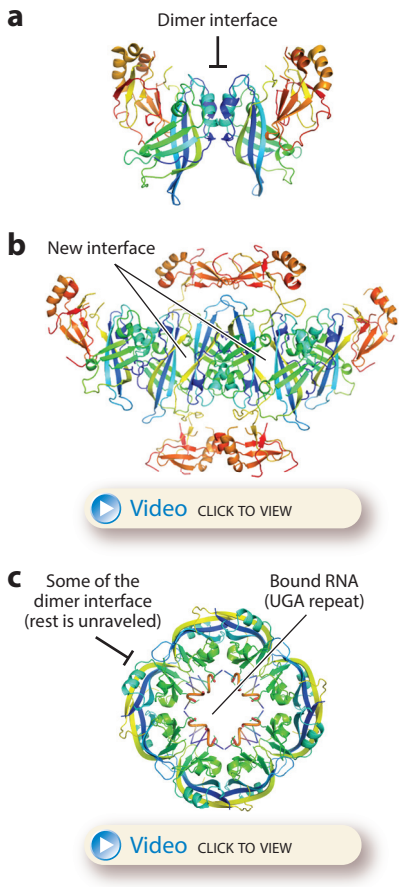

Figure 1

Three structures of VP40, colored blue to red from N to C terminus. (*a*) A butterfly-shaped dimer of VP40 is essential for trafficking (9). The fold of the protomer was determined in 2000 (6, 78). (*b*) A hexamer, formed by rearrangement of N-terminal (*blue and green*) and C-terminal (*orange*) domains and inversion of central subunits, polymerizes into filaments to build and bud virions (9). The central four C-terminal domains are disordered in the crystal structure. **Video 1** models assembly of the hexamer from the dimer. (*c*) An octameric ring binds RNA (7) and controls transcription (15). No other form of VP40 is known to bind RNA. Some of the residues that form the dimer interface in panel *a* are positioned on the outside of the ring; another 70 are disordered. The C-terminal domains are absent from this crystal structure. **Video 2** models assembly of the ring from the dimer. **Videos 1** and **2** reproduced from Reference 9 with permission from Cell Press.

Marburg virus VP40 similarly forms a dimer in solution and likely a filament as well (17), and earlier work demonstrated that it forms an oligomeric ring (8). Interestingly, Marburg virus VP40 has also been noted to exhibit an immunosuppressive function (18). This immunosuppressive function can be achieved by the N-terminal domain expressed alone (17). In the absence of the C-terminal domain, the N-terminal domain forms a ring. Therefore, the ring may be the Marburg virus VP40 structure that suppresses interferon signaling, though it is not yet clear whether the transcriptional control, immunosuppressive, and RNA-binding functions of the ring are related. Here some interesting evolutionary questions are posed: How did VP40 acquire the ability to

adopt a second (or third) structure, while maintaining fidelity and function of the first structure? Are other matrix proteins transformers as well? Where else in biology do such proteins exist, and how do they achieve their additional structural forms?

OTHER SERENDIPITOUS EXAMPLES

We describe here both early and more recent examples of proteins found to be multistructural. Thus far, nearly all of these examples have been found serendipitously. We still lack the tools to predict transformer behavior bioinformatically (as can be done for intrinsically disordered proteins). It is hoped, however, that further study of the examples that have fallen into our laps will help develop the necessary bioinformatics tools by which we could predict transformer behavior.

Concept Differentiation

The terms metamorphic protein (19), morpheein (20), moonlighting protein (21), and transformer have all been used to describe proteins that change structure or function. "Metamorphic protein" typically describes proteins that change structure. A change in structure often occurs in the course of carrying out a protein's single function (e.g., a serpin or a viral surface glycoprotein), but not necessarily (e.g., lymphotactin, as described below). "Morpheein" typically describes a protein that changes oligomeric state (particularly so that oligomeric assembly can turn a function on or off). "Moonlighting protein" describes a protein that changes function, often using an unchanging structure to achieve different functions by binding distinct ligands or occupying different organelles. The semantics are somewhat fluid, and all these terms and multiple reviews in which they have been used wonderfully describe the panoply of proteins on Earth. We favor "transformer" to describe the intersection of all these concepts: proteins such as VP40 that change conformation (and perhaps oligomeric state) in order to select among entirely different functions. "Transformer" also evokes the toys of the same name, which change from a robot to a vehicle to achieve different functionalities.

Intrinsically disordered proteins change both structure and function. These, however, are different in that their adaptability arises from inherent disorder, which can be predicted by bioinformatics analysis of sequence.

Proteins that are not intrinsically disordered but nonetheless change structure in order to change their function may achieve this switch by presence or absence of an additional peptide, by changing the fold of the polypeptide, or by differential oligomerization.

Presence or Absence of a Propeptide

For some proteins, an additional propeptide sequence causes a different tertiary or quaternary assembly when it is present or absent by allowing or preventing certain interactions. Two examples of proteins that exhibit this behavior are the hepatitis B capsid protein and the bacterial matrix metallopeptidase (MMP) protein karilysin. One form of the hepatitis B capsid protein, termed HBcAg, assembles the viral capsid after a proposed allosteric structural change (22–25) (**Figure 2*a***). The alternate form of the protein, termed HBeAg, is prevented from assembling like HBcAg by an additional 10-residue propeptide (26, 27) (**Figure 2*b***). HBeAg plays no role in capsid assembly but instead attenuates the immune response (28). Karilysin is a bacterial MMP. This enzyme includes a 14-residue sequence that binds across the active site, thereby keeping the enzyme inactive; this sequence also assists in protein folding and stabilization. Later, a maturation process, which includes cleaving away the 14-residue sequence followed by rearrangement of the residues surrounding the active site, accomplishes enzyme activation (29).

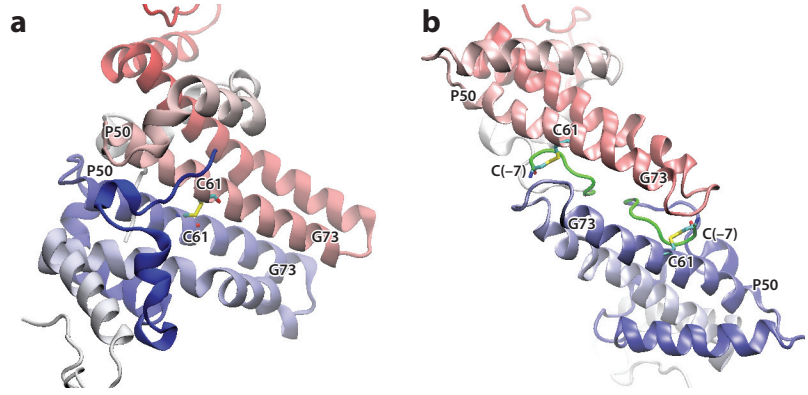

Figure 2

Different dimers formed by the hepatitis B virus capsid protein. (*a*) The HBcAg form of the protein forms a dimer critical for capsid assembly (PDB: 3KXS) (25). One monomer is colored blue to white from N to C terminus. The other monomer is colored white to red from N to C terminus. (*b*) The alternate HBeAg antigen contains an additional N-terminal propeptide (*green*) that prevents assembly by the interface of HBcAg, causing assembly via an alternate dimer (PDB: 3V6Z) (27). Note that intramolecular C(−7)-C61 disulfide bonds form, preventing formation of the intermolecular C61-C61 disulfide bond found in panel *a*.

Fold Switch

Other proteins are known to switch the fold of individual monomers. KaiB, a cyanobacterial clock protein, was recently found to switch between two folds in its function as an oscillator (30), rethreading the C-terminal half of its polypeptide chain to change the fold from a βαββααβ pattern to a βαβαββα pattern. This fold switch allows KaiB to switch from inactive to active states on a timescale that matches that of Earth's day/night cycle. Although this fold switch is now known to be essential to regulation of circadian rhythm, the existence of the fold switch and of a second fold of the KaiB polypeptide was not predictable from the sequence, the first structure, or any existing method. As for VP40, the second fold was found only via chance observations. However, this fold switch appears to be an on/off of a single function rather than a switch between function A and function B.

Fold Switch and Oligomerization

Prions and amyloid are well-known molecules for which a change in fold and propagated oligomerization alter polypeptide function (31–34). One difference is that prion changes appear irreversible, whereas those of transformers and metamorphic proteins could be reversible and in equilibrium. An additional difference is that the malignant PrP^{Sc} structure does not benefit the organism that encodes the PrP protein. Instead, the PrP^{Sc} structure reflects a disease state, and mutations that block that conversion are beneficial. In contrast, Ebola virus requires all the different forms of VP40 in order to complete its life cycle. The yeast Mod5 protein, however, does confer benefit in its amyloid structure, namely adaption to environmental stress (35).

Another example of a fold switch with oligomerization is the human chemokine lymphotactin. Lymphotactin exists in an equilibrium between two conformations, termed Ltn10 and Ltn40, which under physiological conditions (37°C, 150 mM NaCl) are equally populated and readily

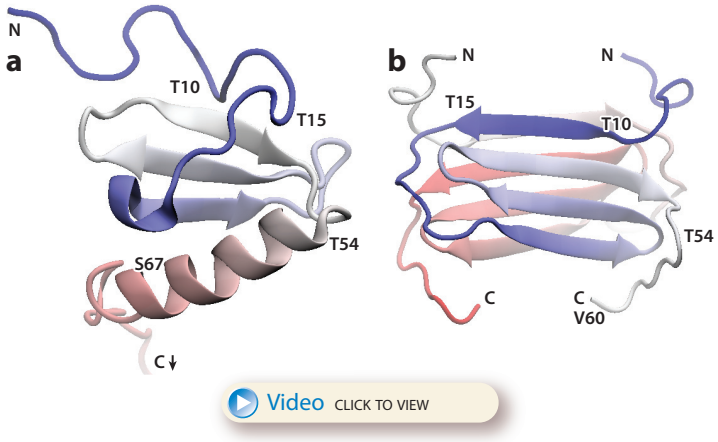

Figure 3

Alternative structures for lymphotactin. (*a*) The originally discovered, monomeric form (PDB: 1J8I) (37). Residues 1–93 are included in the structure, though residues 80–93 (part of the long C-terminal tail) are not shown here. (*b*) The more recently discovered, dimeric form (PDB: 2JP1) (40); the second chain of the dimer is seen in the background. Residues 1–60 are included in the structure. In panel *a*, there is an α-helix at residues 54–67, which is not present in panel *b*. Moreover, in panel *a*, there is a three-strand β-sheet; in panel *b*, there is a four-strand β-sheet, including a new first strand at residues 10–15. **Video 3** depicts a morph between the two folds.

interconvert. At 10°C, 200 mM NaCl, Ltn10 dominates; at 40°C, 0 M NaCl, Ltn40 dominates (36). Point mutations have also been employed to favor one configuration over the other, and hence to determine the functions of each configuration.

The Ltn10 form of lymphotactin, solved in 2001, is a monomeric canonical chemokine fold (37) (**Figure 3*a***). In this fold, lymphotactin is identified as the prototypical and indeed the only member of the C family of chemokines (38) and is responsible for receptor activation (39). This fold has strong structural similarities to those of other known chemokines, except that it lacks one of two disulfide bonds that are characteristic of the other chemokine families. Notable secondary structure features are an N-terminal tail followed by a three-strand antiparallel β-sheet, a four-turn α-helix, and a C-terminal tail. Ltn10 forms a monomer and serves as an agonist to the chemokine receptor XCR1.

The alternative fold of the same sequence, Ltn40, was solved in 2007. In contrast to the Ltn10 form, the Ltn40 fold of the same sequence is not closely related to that of any known chemokine (40) (**Figure 3*b***). The notable secondary structure feature is a four-strand antiparallel β-sheet. Part of the N-terminal tail coil in Ltn10 has been replaced by a new initial β-strand (residues 10–15). Moreover, the α-helix of Ltn10 (residues 54–67) has been replaced by C-terminal tail coil in Ltn40 (and residues 61–67 are also disordered). Overall structural changes are substantial: "[V]irtually all Ltn10 tertiary interactions are replaced with different tertiary and quaternary contacts in Ltn40. Except for a few residues in β1 and β3, interconversion between the two structures repositions every side chain relative to the hydrophobic core, as if the protein was being turned inside out" (40, p. 5058; 41). Further, unlike Ltn10, Ltn40 forms a dimer. Ltn40 is also distinct in function from Ltn10. Ltn10 binds XCR1 and thereby induces chemotaxis, whereas Ltn40 does not. Ltn40, however, more strongly binds glycosaminoglycans, which are common binding partners for chemokines on the cell surface; this binding may help maintain a chemotactic

concentration gradient and may also serve as a lock to prevent accidental conversion of Ltn40 into Ltn10.

As is found for many known transformer proteins, the two conformations lead to different oligomerization states (in this case, monomer versus dimer). Lymphotactin also appears to be a true transformer protein in that associated with its two conformations are two distinct functions (XCR1 agonism and binding of glycosaminoglycans). We speculate that factors present in the cell deliberately select a preponderance of Ltn10, a preponderance of Ltn40, or a more balanced equilibrium, depending on which function is needed at a given time, in order to regulate lymphotactin's T cell recruitment. However, as is often true in our current understanding of multistructural or transformer proteins, we do not know what these factors are or how exactly they trigger selective formation of Ltn10 or Ltn40. The above rules involving temperature, salt, and point mutations, though useful in laboratory work, are evidently not the true selective factors found under physiological conditions. However, studying how these laboratory factors act on lymphotactin may provide significant leads.

A Potential Transformer

NS1 is a nonstructural protein of the influenza A virus. It has multiple immunosuppressive functions, some achieved by binding RNA. Each NS1 protomer consists of an RNA-binding domain and an effector domain, joined by a linker. All known structures of NS1 dimerize, but configurations of the dimer are distinct. The first structure was solved in 2008 (PDB: 3F5T) from an H5N1 strain (42) that encoded a shorter, partially disordered interdomain linker. The resulting NS1 dimer is linear, with the effector domains flanking the joined RNA-binding domains (**Figure 4*a***).

An alternative structure was solved in 2014 (PDB: 4OPH) from an H6N6 strain (43) (**Figure 4*b***). The linkers in this strain are full length (including a TIASV sequence at residues 80–84) and are completely ordered. The region encoding residues 74–77 (disordered in the H5N1 structure) are visible and form a β-turn, which seemingly helps to reposition the effector domains. The resulting configuration is more compact than the H5N1 structure, with the two effector domains placed around the joined RNA-binding domains roughly 120° apart, rather than 180°. Moreover, a surface loop in each effector domain interfaces with the RNA-binding domain of the same protomer, further stabilizing the distinctive positioning of the effector domains.

PDB 4OPA is an experimental variant of 4OPH, in which the genetic sequence of 4OPH has been modified to shorten the linkers by deleting the TIASV sequence at residues 80–84. This genetic sequence is thus intermediate between those of 4OPH and 3F5T, and one might expect the resulting structure to be similarly intermediate between the 4OPH and 3F5T structures. Surprisingly, that is not true. As in 4OPH, the linkers are completely ordered, with a distinctive β-turn at residues 74–77 (43) (**Figure 4*c***). But the resulting dimer configuration is even more compact than 4OPH, with the effector domains so close together that they interface with each other, augmenting the dimer interface.

A wide range of effector domain–effector domain interactions have been observed (42–48) and analyzed (43, 45, 46, 49–51), with higher-order structures potentiating dsRNA binding and virulence (52). As for VP40, different conformations and assemblies of NS1 predominate at different times and in different subcellular environments (46, 49). Further, it is thought that the interface buried by one type of NS1 dimer must instead be exposed in order for the protein to perform other functions, namely interaction with cleavage and polyadenylation specificity factor 30 (CPSF30) in suppression of host antiviral gene expression (53, 54) and binding of the inter-SH2 domain of the

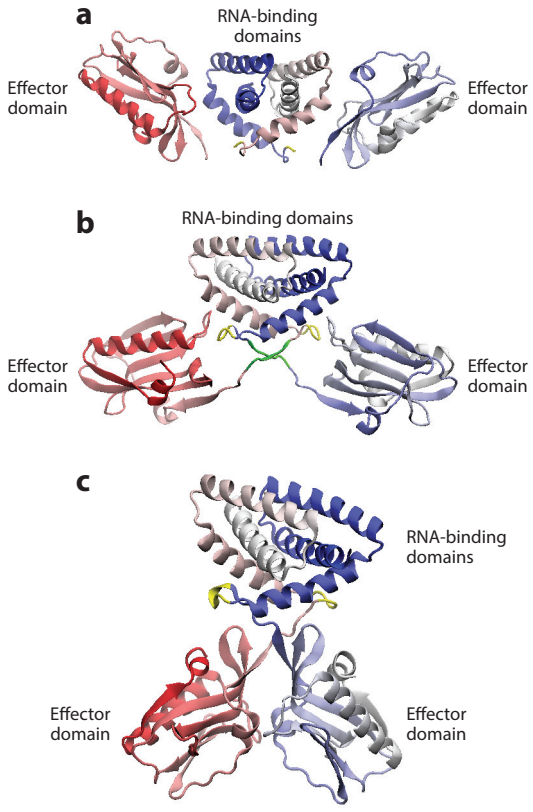

Figure 4

Three structures of influenza virus NS1. (*a*) NS1 from H5N1 (A/Vietnam/1203/2004), which contains a natural deletion in residues 80–84 of the linker (PDB: 3F5T) (42). Residues 75–79, also of the linker, are disordered. This dimer is linear, with the two effector domains flanking the central paired RNA-binding domains. (*b*) NS1 from H6N6 (A/blue-winged teal/MN/993/1980) (PDB: 4OPH) (43). As is typical for the H6N6 strain, residues 80–84 are present (*green*). Also, the entire linker is ordered, and there is a β-turn at residues 74–77 (*yellow*). This dimer is more compact, with the two effector domains closer to each other. The effector domains are also closer to the paired RNA-binding domains; indeed, a surface loop in each effector domain makes contact with the same protomer's RNA-binding domain. (*c*) NS1 from H6N6 (A/blue-winged teal/MN/993/1980) (PDB: 4OPA) (43) that contains its natural residues 74–77 (*yellow*), but from which residues 80–84 of the linker have been artificially deleted, in mimicry of the natural deletion of the sequence depicted in panel *a*. One might expect that this NS1 would be arranged like the similarly deleted 3F5T in panel *a* or the similarly sequenced 4OPH in panel *b*, or perhaps intermediate between the two. Unexpectedly, this dimer is even more compact than that of 4OPH. The two effector domains are now in direct contact, strengthening the dimer interface by connecting with each other.

p85b isoform of phosphoinositide 3-kinase (PI3K) in stimulation of lipid kinase activity in support of virus replication (55).

Given that the NS1 linker seems a natural candidate for multiform flexibility, we speculate that multiple NS1 configurations exist in equilibrium and may be dynamically triggered by factors that are currently unknown. Genetic differences among strains may alter the statistical preferences among preexisting possibilities, and so may have given rise to experimental visualization of one form over another.

TRANSFORMERS ELSEWHERE IN VIROLOGY?

Multifunctional proteins abound throughout virology. We wonder how many other multifunctional proteins are similarly multistructural but have yet to be identified as such. In some cases, perhaps it simply is not known that there is a distinct structure remaining to be solved, so no energy has been invested in solving it. Perhaps only one form is easily amenable to structural analysis. Perhaps formation of crystal contacts drove the equilibrium toward just one of the structures. Perhaps nuclear magnetic resonance was not suitable because of convoluted spectra or the size of the molecule. Perhaps a biological cofactor is required for transformation of the protein into its other structure, and that cofactor was not supplied in protein expression or purification.

Human immunodeficiency virus type 1 (HIV-1) integrase is another potential transformer but has thus far eluded structural analysis in its complete state, possibly because of conformational or oligomeric heterogeneity. Paramyxovirus matrix proteins, which are known to be multifunctional, and some of which have the same domain organization as VP40, could be transformers. Protein structures of some of these are available: Borna disease virus as a monomer (56), Newcastle disease virus as a dimer and in a pseudo-tetrameric array that may be involved in matrix assembly (57), human metapneumovirus as a dimer and helical filament (58), vesicular stomatitis virus (59, 60) and Lagos bat virus (60) as monomers, and respiratory syncytial virus as a monomer (61) and as the dimeric building block of the viral matrix (62). These structures provide a roadmap to determine, for proteins that are known to be multifunctional, whether there could be alternate structures associated with the different functions.

FOR STRUCTURAL BIOLOGY IN GENERAL

We believe transformer proteins will be more often found in viruses, for which error-prone replication, a rapid life cycle, and constant selective pressure may have forced adaptability in a compact genome. However, the behavior of these proteins depends on thermodynamics of polypeptide folding and assembly that are common to both viruses and the cells they infect. This one gene–multiple structure phenomenon is another, as-yet-unexplored strategy by which additional functional complexity is encoded in the genome. Armed with the insight that additional hidden structures may exist, and the necessary strategies to pursue them, we may discover a variety of nonviral transformer proteins. By finding such proteins in viruses, we hope to identify bioinformatics patterns by which they may be identified elsewhere, in prokaryotic and eukaryotic genomes.

Viruses have long been used as searchlights to illuminate unknowns throughout biology. A key paper cited in the conferral of James Rothman's 2013 Nobel Prize for cellular trafficking was founded on discovery of the mechanism by which the vesicular stomatitis virus glycoprotein (VSV-G) is transported and processed (63). Other studies on VSV-G ultimately illuminated a molecular cause of cystic fibrosis (64). The discovery of the major cancer potentiator p53 was actuated by its precipitation by the T antigen of simian virus 40 (65–69). The discovery of the major histocompatibility complex, a critical factor in the success of lifesaving organ transplants, was achieved through analysis of susceptibility to lymphocytic choriomeningitis virus (70–73). This body of work led to Rolf Zinkernagel and Peter Doherty's 1996 Nobel Prize.

Why do viruses so clearly illuminate so much else in biology? The tremendous evolutionary selective pressure under which viruses live forces them to quickly evaluate and exploit the host cells in which they reside: to identify and usurp key host processes, identify and suppress host defenses, etc. Viruses know our cells better than we do and point us in the necessary direction. For structural biology, viruses shine another beacon in how they encode information, optimizing the

forms and functions of their proteins in order to store maximum information in a compact genome. Structures of viruses found in extreme environments also illuminate mechanisms by which protein assemblies achieve extraordinary strength and stability (74, 75), and understanding dynamic capsid structures and viral entry is allowing development of novel drug-delivery strategies (76, 77). Now viral transformers may give us the models and examples we need to find this behavior elsewhere.

Our understanding is most strongly shaped by what we see. The ability to determine atomic structures of proteins has provided biology with clear and compelling roadmaps for interpreting biological function. Yet protein structures are snapshots. Proteins whose structures have been solved by X-ray crystallography are those that are able to form stable intermolecular contacts. Sometimes these contacts can reveal biologically relevant interfaces; at other times, they may force a dynamic protein into only one or two of its possible conformations. Electron microscopy requires no crystals but is subject to the same selective forces in its requirement for pure and homogeneous protein for high resolution. No matter what method is used to solve the structure, choices made in construct design, expression, and purification may influence the predominant protein conformation and hence the structures obtained.

Our intellectual framework is forged by those structures already revealed. We typically expect, when a complete protein structure has been deposited in the PDB, that the structure is more or less solved. The discovery of different, rearranged forms of VP40, lymphotactin, and the other examples outlined above often comes as a surprise. The growing number of these alternative structures, along with biological support for their functions, suggests that more such multistructural or transformer proteins exist in biology and may be hiding in plain sight. How often has there been a mutation that does not make sense in the context of a known structure? Often such mutants are dismissed as being detrimental to the fold of the protein. Sometimes they are. Other times, these point mutations (such as I307R for VP40) may alter the equilibrium of the protein's distinct forms and individual functions. For many proteins there will be one characteristic structure, with the normal amount of breathing or conformationally flexible termini. But for other key proteins, there may be more than one structure: a potential hidden proteome.

The examples described here illustrate how structural rearrangement of transcribed polypeptides can extend the functional reach of the genome that encodes them. The definition and study of the multiple structures and multiple functions of these proteins constitute a rich and still emerging field that will offer valuable insights for structural biology and structural interpretation in general. And viruses in particular may again provide the leading light for biological advancement.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Elliott LH, Kiley MP, McCormick JB. 1985. Descriptive analysis of Ebola virus proteins. *Virology* 147:169–76
2. Sanchez A, Kiley MP, Holloway BP, McCormick JB, Auperin DD. 1993. Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Res.* 29:215–40
3. Jasenosky LD, Neumann G, Lukashevich I, Kawaoka Y. 2001. Ebola virus VP40-induced particle formation and association with the lipid bilayer. *J. Virol.* 75:5205–14
4. Hoenen T, Jung S, Herwig A, Groseth A, Becker S. 2010. Both matrix proteins of Ebola virus contribute to the regulation of viral genome replication and transcription. *Virology* 403:56–66

5. Valmas C, Basler CF. 2011. Marburg virus VP40 antagonizes interferon signaling in a species-specific manner. *J. Virol.* 85:4309–17
6. Dessen A, Volchkov V, Dolnik O, Klenk HD, Weissenhorn W. 2000. Crystal structure of the matrix protein VP40 from Ebola virus. *EMBO J.* 19:4228–36
7. Gomis-Ruth FX, Dessen A, Timmins J, Bracher A, Kolesnikowa L, et al. 2003. The matrix protein VP40 from Ebola virus octamerizes into pore-like structures with specific RNA binding properties. *Structure* 11:423–33
8. Timmins J, Schoehn G, Kohlhaas C, Klenk HD, Ruigrok RW, Weissenhorn W. 2003. Oligomerization and polymerization of the filovirus matrix protein VP40. *Virology* 312:359–68
9. Bornholdt ZA, Noda T, Abelson DM, Halfmann P, Wood MR, et al. 2013. Structural rearrangement of Ebola virus VP40 begets multiple functions in the virus life cycle. *Cell* 154:763–74
10. Ruigrok RW, Schoehn G, Dessen A, Forest E, Volchkov V, et al. 2000. Structural characterization and membrane binding properties of the matrix protein VP40 of Ebola virus. *J. Mol. Biol.* 300:103–12
11. Scianimanico S, Schoehn G, Timmins J, Ruigrok RH, Klenk HD, Weissenhorn W. 2000. Membrane association induces a conformational change in the Ebola virus matrix protein. *EMBO J.* 19:6732–41
12. Zschornig O, Arnold K, Ohki S. 1993. Effect of glycosaminoglycans and PEG on fusion of Sendai virus with phosphatidylserine vesicles. *Biochim. Biophys. Acta* 1148:1–6
13. Timmins J, Scianimanico S, Schoehn G, Weissenhorn W. 2001. Vesicular release of Ebola virus matrix protein VP40. *Virology* 283:1–6
14. Noda T, Sagara H, Suzuki E, Takada A, Kida H, Kawaoka Y. 2002. Ebola virus VP40 drives the formation of virus-like filamentous particles along with GP. *J. Virol.* 76:4855–65
15. Hoenen T, Biedenkopf N, Zielecki F, Jung S, Groseth A, et al. 2010. Oligomerization of Ebola virus VP40 is essential for particle morphogenesis and regulation of viral transcription. *J. Virol.* 84:7053–63
16. Nanbo A, Watanabe S, Halfmann P, Kawaoka Y. 2013. The spatio-temporal distribution dynamics of Ebola virus proteins and RNA in infected cells. *Sci. Rep.* 3:1206
17. Oda S, Noda T, Wijesinghe KJ, Halfmann P, Bornholdt ZA, et al. 2015. Crystal structure of Marburg virus VP40 reveals a broad, basic patch for matrix assembly and a requirement of the N-terminal domain for immunosuppression. *J. Virol.* 90:1839–48
18. Valmas C, Grosch MN, Schumann M, Olejnik J, Martinez O, et al. 2010. Marburg virus evades interferon responses by a mechanism distinct from Ebola virus. *PLOS Pathog.* 6:e1000721
19. Murzin AG. 2008. Metamorphic proteins. *Science* 320:1725–26
20. Jaffe EK, Lawrence SH. 2012. The morpheein model of allostery: evaluating proteins as potential morpheeins. *Methods Mol. Biol.* 796:217–31
21. Jeffery CJ. 2004. Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins. *Curr. Opin. Struct. Biol.* 14:663–68
22. Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, et al. 2008. The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* 322:1211–17
23. Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, et al. 2008. A specific cholesterol binding site is established by the 2.8 Å structure of the human β2-adrenergic receptor. *Structure* 16:897–905
24. Wynne SA, Crowther RA, Leslie AG. 1999. The crystal structure of the human hepatitis B virus capsid. *Mol. Cell* 3:771–80
25. Packianathan C, Katen SP, Dann CE III, Zlotnick A. 2010. Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. *J. Virol.* 84:1607–15
26. Zlotnick A, Tan Z, Selzer L. 2013. One protein, at least three structures, and many functions. *Structure* 21:6–8
27. DiMattia MA, Watts NR, Stahl SJ, Grimes JM, Steven AC, et al. 2013. Antigenic switching of hepatitis B virus by alternative dimerization of the capsid protein. *Structure* 21:133–42
28. Elgouhari HM, Abu-Rajab Tamimi TI, Carey WD. 2008. Hepatitis B virus infection: understanding its epidemiology, course, and diagnosis. *Clevel. Clin. J. Med.* 75:881–89
29. Lopez-Pelegrin M, Ksiazek M, Karim AY, Guevara T, Arolas JL, et al. 2015. A novel mechanism of latency in matrix metalloproteinases. *J. Biol. Chem.* 290:4728–40
30. Chang YG, Cohen SE, Phong C, Myers WK, Kim YI, et al. 2015. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* 349:324–28

31. Prusiner SB. 1982. Novel proteinaceous infectious particles cause scrapie. *Science* 216:136–44
32. Holmes BB, Diamond MI. 2012. Cellular mechanisms of protein aggregate propagation. *Curr. Opin. Neurol.* 25:721–26
33. Prusiner SB. 2012. A unifying role for prions in neurodegenerative diseases. *Science* 336:1511–13
34. Jucker M, Walker LC. 2013. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* 501:45–51
35. Suzuki G, Shimazu N, Tanaka M. 2012. A yeast prion, Mod5, promotes acquired drug resistance and cell survival under environmental stress. *Science* 336:355–59
36. Kuloglu ES, McCaslin DR, Markley JL, Volkman BF. 2002. Structural rearrangement of human lymphotactin, a C chemokine, under physiological solution conditions. *J. Biol. Chem.* 277:17863–70
37. Kuloglu ES, McCaslin DR, Kitabwalla M, Pauza CD, Markley JL, Volkman BF. 2001. Monomeric solution structure of the prototypical 'C' chemokine lymphotactin. *Biochemistry* 40:12486–96
38. Kelner GS, Kennedy J, Bacon KB, Kleyensteuber S, Largaespada DA, et al. 1994. Lymphotactin: a cytokine that represents a new class of chemokine. *Science* 266:1395–99
39. Tuinstra RL, Peterson FC, Elgin ES, Pelzek AJ, Volkman BF. 2007. An engineered second disulfide bond restricts lymphotactin/XCL1 to a chemokine-like conformation with XCR1 agonist activity. *Biochemistry* 46:2564–73
40. Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, Volkman BF. 2008. Interconversion between two unrelated protein folds in the lymphotactin native state. *PNAS* 105:5057–62
41. Tyler RC, Murray NJ, Peterson FC, Volkman BF. 2011. Native-state interconversion of a metamorphic protein requires global unfolding. *Biochemistry* 50:7077–79
42. Bornholdt ZA, Prasad BV. 2008. X-ray structure of NS1 from a highly pathogenic H5N1 influenza virus. *Nature* 456:985–88
43. Carrillo B, Choi JM, Bornholdt ZA, Sankaran B, Rice AP, Prasad BV. 2014. The influenza A virus protein NS1 displays structural polymorphism. *J. Virol.* 88:4113–22
44. Bornholdt ZA, Prasad BV. 2006. X-ray structure of influenza virus NS1 effector domain. *Nat. Struct. Mol. Biol.* 13:559–60
45. Hale BG, Barclay WS, Randall RE, Russell RJ. 2008. Structure of an avian influenza A virus NS1 protein effector domain. *Virology* 378:1–5
46. Kerry PS, Ayllon J, Taylor MA, Hass C, Lewis A, et al. 2011. A transient homotypic interaction model for the influenza A virus NS1 protein effector domain. *PLOS ONE* 6:e17946
47. Kerry PS, Long E, Taylor MA, Russell RJ. 2011. Conservation of a crystallographic interface suggests a role for β-sheet augmentation in influenza virus NS1 multifunctionality. *Acta Crystallogr. F* 67:858–61
48. Xia S, Monzingo AF, Robertus JD. 2009. Structure of NS1A effector domain from the influenza A/Udorn/72 virus. *Acta Crystallogr. D* 65:11–17
49. Aramini JM, Hamilton K, Ma LC, Swapna GV, Leonard PG, et al. 2014. ^{19}F NMR reveals multiple conformations at the dimer interface of the nonstructural protein 1 effector domain from influenza A virus. *Structure* 22:515–25
50. Aramini JM, Ma LC, Zhou L, Schauder CM, Hamilton K, et al. 2011. Dimer interface of the effector domain of non-structural protein 1 from influenza A virus: an interface with multiple functions. *J. Biol. Chem.* 286:26050–60
51. Xia S, Robertus JD. 2010. X-ray structures of NS1 effector domain mutants. *Arch. Biochem. Biophys.* 494:198–204
52. Hale BG. 2014. Conformational plasticity of the influenza A virus NS1 protein. *J. Gen. Virol.* 95:2099–105
53. Das K, Ma LC, Xiao R, Radvansky B, Aramini J, et al. 2008. Structural basis for suppression of a host antiviral response by influenza A virus. *PNAS* 105:13093–98
54. Noah DL, Twu KY, Krug RM. 2003. Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3′ end processing of cellular pre-mRNAS. *Virology* 307:386–95
55. Hale BG, Batty IH, Downes CP, Randall RE. 2008. Binding of influenza A virus NS1 protein to the inter-SH2 domain of p85 suggests a novel mechanism for phosphoinositide 3-kinase activation. *J. Biol. Chem.* 283:1372–80

56. Neumann P, Lieber D, Meyer S, Dautel P, Kerth A, et al. 2009. Crystal structure of the Borna disease virus matrix protein (BDV-M) reveals ssRNA binding properties. *PNAS* 106:3710–15
57. Battisti AJ, Meng G, Winkler DC, McGinnes LW, Plevka P, et al. 2012. Structure and assembly of a paramyxovirus matrix protein. *PNAS* 109:13996–4000
58. Leyrat C, Renner M, Harlos K, Huiskonen JT, Grimes JM. 2014. Structure and self-assembly of the calcium binding matrix protein of human metapneumovirus. *Structure* 22:136–48
59. Gaudier M, Gaudin Y, Knossow M. 2002. Crystal structure of vesicular stomatitis virus matrix protein. *EMBO J.* 21:2886–92
60. Graham SC, Assenberg R, Delmas O, Verma A, Gholami A, et al. 2008. Rhabdovirus matrix protein structures reveal a novel mode of self-association. *PLOS Pathog.* 4:e1000251
61. Money VA, McPhee HK, Mosely JA, Sanderson JM, Yeo RP. 2009. Surface features of a *Mononegavirales* matrix protein indicate sites of membrane interaction. *PNAS* 106:4441–46
62. Forster A, Maertens GN, Farrell PJ, Bajorek M. 2015. Dimerization of matrix protein is required for budding of respiratory syncytial virus. *J. Virol.* 89:4624–35
63. Balch WE, Dunphy WG, Braell WA, Rothman JE. 1984. Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine. *Cell* 39:405–16
64. Nishimura N, Balch WE. 1997. A di-acidic signal required for selective export from the endoplasmic reticulum. *Science* 277:556–58
65. Lane DP, Crawford LV. 1979. T antigen is bound to a host protein in SV40-transformed cells. *Nature* 278:261–63
66. Melero JA, Stitt DT, Mangel WF, Carroll RB. 1979. Identification of new polypeptide species (48–55K) immunoprecipitable by antiserum to purified large T antigen and present in SV40-infected and -transformed cells. *Virology* 93:466–80
67. Linzer DI, Levine AJ. 1979. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* 17:43–52
68. Kress M, May E, Cassingena R, May P. 1979. Simian virus 40-transformed cells express new species of proteins precipitable by anti-simian virus 40 tumor serum. *J. Virol.* 31:472–83
69. Chang C, Martin RG, Livingston DM, Luborsky SW, Hu CP, Mora PT. 1979. Relationship between T-antigen and tumor-specific transplantation antigen in simian virus 40-transformed cells. *J. Virol.* 29:69–75
70. Zinkernagel RM, Doherty PC. 1973. Cytotoxic thymus-derived lymphocytes in cerebrospinal fluid of mice with lymphocytic choriomeningitis. *J. Exp. Med.* 138:1266–69
71. Oldstone MB, Dixon FJ, Mitchell GF, McDevitt HO. 1973. Histocompatibility-linked genetic control of disease susceptibility: murine lymphocytic choriomeningitis virus infection. *J. Exp. Med.* 137:1201–12
72. Doherty PC, Zinkernagel RM. 1975. *H-2* compatibility is required for T-cell-mediated lysis of target cells infected with lymphocytic choriomeningitis virus. *J. Exp. Med.* 141:502–7
73. Zinkernagel RM, Dunlop MB, Blanden RV, Doherty PC, Shreffler DC. 1976. *H-2* compatibility requirement for virus-specific T-cell-mediated cytolysis: evaluation of the role of *H-2I* region and non-*H-2* genes in regulating immune response. *J. Exp. Med.* 144:519–32
74. Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, et al. 2001. Viruses from extreme thermal environments. *PNAS* 98:13341–45
75. DiMaio F, Yu X, Rensen E, Krupovic M, Prangishvili D, Egelman EH. 2015. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* 348:914–17
76. Speir JA, Munshi S, Wang G, Baker TS, Johnson JE. 1995. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* 3:63–78
77. Singh P, Gonzalez MJ, Manchester M. 2006. Viruses and their use in nanotechnology. *Drug Dev. Res.* 67:23–41
78. Dessen A, Forest E, Volchkov V, Dolnik O, Klenk HD, Weissenhorn W. 2000. Crystallization and preliminary X-ray analysis of the matrix protein from Ebola virus. *Acta Crystallogr. D* 56:758–60