Review

CellPress

Insights into the trimeric HIV-1 envelope glycoprotein structure

Andrew B. Ward[1,2] and Ian A. Wilson[1,2,3]

[1] Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
[2] International AIDS Vaccine Initiative Neutralizing Antibody Center and Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, Scripps Research Institute, La Jolla, CA 92037, USA
[3] Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

The HIV-1 envelope glycoprotein (Env) trimer is responsible for receptor recognition and viral fusion with CD4$^+$ T cells, and is the sole target for neutralizing antibodies. Thus, understanding its molecular architecture is of significant interest. However, the Env trimer has proved to be a challenging target for 3D structure determination. Recent electron microscopy (EM) and X-ray structures have at last enabled us to decipher the structural complexity and unique features of the Env trimer, and how it is recognized by an ever-expanding arsenal of potent broadly neutralizing antibodies. We describe our current knowledge of the Env trimer structure in the context of exciting recent developments in the identification and characterization of HIV broadly neutralizing antibodies.

Introduction: bringing the trimeric HIV-1 Env structure to fruition.

The trimeric HIV-1 Env glycoprotein is a metastable type I membrane fusion machine that mediates entry of the virus into the target cell. Env is expressed as a gp160 precursor that is proteolytically cleaved by furin into gp120 and gp41 heterodimers. Three such heterodimers assemble into the final trimeric Env spike. The gp120 subunit has a highly-variable surface including five variable loops (V1–V5). By contrast, the gp41 subunits are more conserved in sequence because they house the fusion machinery, which is complex with many moving parts that undergo enormous conformational rearrangements during the fusion process. The gp41 membrane-proximal external region (MPER) connects the gp41 ectodomain to the transmembrane domain (TMD) and the C-terminal domain (CTD). Perhaps the greatest challenge for structure determination (as well as immunological characterization) is that the Env trimer readily dissociates into gp120 and gp41 subunits, making Env a particularly difficult molecule to study using conventional biophysical methods.

Corresponding authors: Ward, A.B. (abward@scripps.edu); Wilson, I.A. (wilson@scripps.edu).
Keywords: HIV-1; envelope glycoprotein trimer; structure; broadly neutralizing antibodies.

0968-0004/
 http://dx.doi.org/10.1016/j.tibs.2014.12.006

Since the original pioneering structure of a monomeric HIV-1 gp120 core determined more than 15 years ago [1], a plethora of gp120 structures have been solved in various forms. Structures of gp120 and its outer domain have been determined with soluble CD4 (sCD4) and co-receptor mimics [1–4], and with different antibodies that bind to the CD4 binding site (CD4bs) or the gp120 outer domain [4–15]. These antibodies, as well as sCD4, have been essential for obtaining structural information because they act as stabilizing agents and crystallization chaperones, although recently some unliganded g120 structures have been determined [16]. All structures of gp120 exhibit a similar core fold, consisting of an inner and an outer domain (OD) connected by a bridging sheet. For successful gp120 crystallization and X-ray structure determination, the functionally important hypervariable loops V1, V2, and V3 at the trimer apex needed to be deleted or severely truncated [17].

Despite the challenges presented by Env, substantial progress has been made recently in obtaining a 3D structure of the HIV Env trimer as well as elucidating Env–antibody and Env–receptor interactions. With a more complete understanding of the Env trimer, a wide variety of previous observations can now be interpreted or put into the appropriate context. The Env trimer structure has also provided a basis for rational vaccine-design efforts aimed at eliciting antibodies against Env [18]. This review is intended to give an overview of recent breakthroughs that have led to elucidation of these soluble Env trimer structures [19–21] and enabled identification of the defining features and characteristics of the pre-fusion gp120 and gp41 subunits, the variable loops, the glycans, and the antigenic surface of this viral fusion machine.

Hitting a moving target: strategies to study Env

Early electron tomography efforts to study the structure of the Env trimer on the viral surface [22–24] were limited in the resolution that they achieved, but they provided a rough outline of the molecular shape of the trimer and allowed docking of gp120 crystal structures to obtain molecular models. More recent tomograms [25,26] at 20–30 Å resolution yielded further details through hybrid or integrative approaches that fitted the crystal structures of gp120 and/or CD4 into the low-resolution EM reconstructions and enabled other portions of the trimer to be modeled for the gp120 region [26], but not for gp41. However,

only limited information regarding the variable loops in gp120 could be gleaned from these low-resolution models.

Many different constructs of soluble, engineered versions of Env have been pursued over the past two decades for structural studies by a large number of groups worldwide, but all save one met with failure as a result of several compounding factors. The Env trimer is not stable and readily dissociates into component subunits. One common strategy was to prevent gp120–gp41 dissociation by removal of the cleavage site on the gp160 precursor and adding trimerizing motifs, such as leucine zippers and foldons, to attempt to stabilize HIV trimers [27,28]. However, while these trimers can be visualized as a single band by gel electrophoresis, we now know that lack of cleavage between gp120 and gp41 and the presence of trimerizing motifs exerted a negative influence on the quaternary structure of the trimer [29,30] such they do not adopt a stable, native-like fold. The heterogeneous nature of these trimers makes them nearly impossible to characterize structurally at even modest resolution, let alone atomic resolution.

A very early approach for engineering a soluble stable version of an Env trimer was through incorporation of a disulfide bond (SOS) between gp120 and gp41, together with a mutation of a residue in the N-terminal heptad repeat region (HR1) of gp41 from isoleucine to proline (IP) to keep gp120 and gp41 together and enhance trimerization (together known as SOSIP) [31,32]. A more susceptible cleavage site (Arg_6 or R6) was also engineered between gp120 and gp41 to facilitate cleavage by furin to produce a soluble, cleaved trimer. While encouraging, the first inceptions of this construct [31,32] that incorporated the Env sequence from an infected patient known as JR-FL (designed in the laboratory of John Moore) did not produce material that led to Env trimer crystals. More recently, screening of different genotypes with SOSIP mutations [33], truncation of the hydrophobic membrane proximal external region (MPER) at the C terminus [34,35], and optimizing expression and purification protocols have resulted in a high-quality Env trimer that is eminently suitable for high-resolution structural studies. The current version of the trimer is on a clade A BG505 background and was selected from a panel of over 50 constructs with sequences from different strains and clades. Structural evaluation by negative-stain EM showed that the BG505.664 protein (truncated at residue 664) formed compact homogeneous trimers with a ~100% yield after purification [33]. Importantly, this trimer retained the antigenic and structural properties of native Env [33,34], which eluded other trimeric Env constructs. Retrospectively, we know now from EM studies that these earlier versions did not produce a high percentage of properly folded trimers [29,30].

The present soluble SOSIP trimers do not have the MPER, transmembrane, and C-terminal domains of full-length Env (Figure 1A). Nevertheless, structural comparisons of these soluble Env trimers with tomographic reconstructions of the Env trimer on the virus [26] and, more recently, a detergent solubilized full-length Env using single particle EM [36], revealed remarkable similarity at the resolution of these studies (~15–35 Å). Probing the structures further using newly discovered broadly neutralizing antibodies (bnAbs) that recognize only cleaved, native trimers [36,37], and that interact with both gp120 and gp41 subunits, illustrates that these SOSIP trimers recapitulate the highly-complex and extended epitopes found on native Env. These studies also revealed the structural similarity of Env across diverse Env sequences corresponding so far to three different subtypes.

SOSIP Env trimer structure

The SOSIP trimer structure (Figure 1A,B) [19,20,38] exhibits an overall architecture similar to that of influenza hemagglutinin (HA), the prototypic type I membrane-fusion protein. The conserved fusion machinery (HA2 for HA; gp41 for Env) is at the base or stem and is located proximal to the membrane, while the more variable head domain (HA1 for HA; gp120 for Env), which contains the receptor-binding site, is distal to the membrane and sits atop the fusion machinery. The interaction of the gp120 monomers with gp41 helps to constrain the gp41 subunits in a pre-fusion conformation, and prevent progression into the post-fusion, extended six-helix bundle conformation (PDB 1ENV) [39]. In the gp41 structure, the two main helices, HR1 and HR2, lie parallel and more tangential to the trimer axis, respectively (Figure 1C). The three HR1 helices form a coiled coil along the trimer axis, as in the central helix of influenza hemagglutinin (see Figure 3 of Julien *et al.* [19]). The gp120 variable loops V1, V2, and V3 comprise the trimer apex and make stabilizing inter-protomer contacts, with V3 being sequestered beneath V1–V2 (Figure 1A–C) and behind a glycan at position N197 on the adjacent gp120 protomer. V4 and V5 project outward from the surface of gp120 and do not interact with the other variable loops. This architecture was conserved both within the crystal lattice necessary for X-ray crystallography and in the vitreous buffer milieu used for single-particle cryoEM studies. A recent structure of the SOSIP trimer at higher resolution (~3.5 Å) enabled a more complete gp41 structure to be built, and described fascinating new features of the gp41 domain with implications for membrane fusion [21].

Env structures using complementary techniques

Crystallographic studies typically require removal of glycans to attain better homogeneity of the protein preparation; therefore, many structures offer an incomplete view of the surface of Env. Most bnAb epitopes, however, contain one or more glycans, therefore presenting an apparent conundrum for structural studies. Creative solutions have been devised to overcome this limitation by first making complexes of gp120 with bnAbs and then digesting away accessible glycans not bound or protected by the antibodies, therefore making the complex more amenable to crystallization [9]. Glycans bound by the bnAb thereby remain protected from glycosidases, thus allowing the epitopes of gp120 within the antibody footprint to be visualized in a near-native glycosylated state. Most studies, however, require production in cell lines that produce uniform high-mannose glycans and lack the ability to produce complex glycans, and therefore subtle differences may arise in the observed antibody–epitope interactions if