

gp41: HIV's shy protein

Michael B Zwick, Erica O Saphire & Dennis R Burton

The first X-ray crystal structures of gp41, the protein that mediates fusion of HIV-1 to target cells, were solved in the mid-1990s. The structures provide a foundation for understanding viral entry and the mechanism of action of compounds that block fusion. The first fusion inhibitor has recently entered the clinic, and the hope is that more potent and broadly active compounds, based on molecular design, will follow.

The structure of gp41 was a long-sought and elusive goal for HIV biologists. The difficulties of working with a membrane protein, the strong tendency of gp41 to aggregate, and other hurdles stood in the way. When Chan *et al.*¹ and Weissenhorn *et al.*² finally produced the first structures in 1997, followed shortly by Tan *et al.*³, they began to give molecular form to the apparatus of viral fusion, whose outline had emerged earlier in the decade. The structures are the focus of much work to design new inhibitors of the fusion process, and serve as an outstanding example of how findings in basic science can stimulate exploration of new therapeutics.

This is what fusion looks like: envelope spikes composed of gp41 and gp120 decorate the surface of HIV-1. The surface glycoprotein of the spike, gp120, engages CD4 and CCR5 or CXCR4 coreceptors, primarily on macrophages and CD4⁺ T cells. gp120 undergoes conformational changes to activate gp41, a transmembrane glycoprotein that uses conserved structural elements to mediate the fusion of the host and viral membranes. Genetic information can then flow from the virus to the target cell.

The structure of the spikes is dictated by the virus' need to bind and enter target cells while avoiding a host neutralizing antibody response. In such a response, host antibodies bind to the virus and prevent it from entering the target cell. To manage this problem, HIV-1 has evolved to keep the vulnerable conserved regions of the envelope spike—such as those on gp41 that comprise the fusion machinery—hidden as

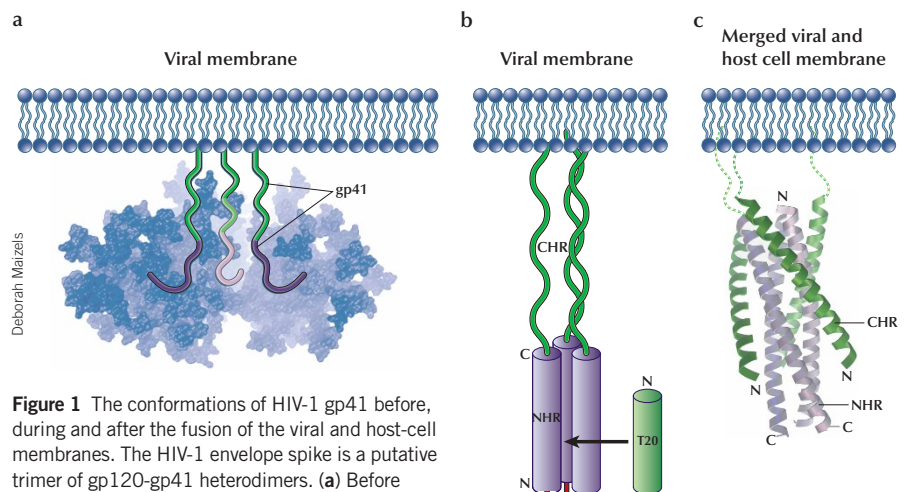


Figure 1 The conformations of HIV-1 gp41 before, during and after the fusion of the viral and host-cell membranes. The HIV-1 envelope spike is a putative trimer of gp120-gp41 heterodimers. **(a)** Before fusion. In the native, untriggered conformation, the ectodomain (extraviral portion) of gp41 is largely cloaked by gp120 (transparent here). The structural details of this native configuration are not well elucidated. **(b)** Fusion intermediate(s). Upon binding CD4 and coreceptor, gp120 undergoes conformational changes that expose gp41 and activate the fusion machinery, notably the N-HR regions and the fusion peptide (FP). gp41 is probably in a somewhat extended arrangement in which the N-HR region of three gp41 molecules form an α -helical bundle, positioning the fusion peptides for insertion into the host-cell membrane. The C-HR region of gp41 is drawn here without defined secondary structure, but some models show it in an α -helical conformation. gp41-targeting fusion inhibitors such as T20 act on the fusion-intermediate conformation(s) of gp41 to inhibit six-helix bundle formation. **(c)** After fusion. Fusion seems to coincide with chain reversal, or 'jackknifing', such that the antiparallel N-HR and C-HR regions form a six-helix bundle, represented here by the structure of Weissenhorn *et al.*².

much as possible. This serves two purposes: it restricts the ability of conserved epitopes to elicit antibodies (low immunogenicity), and it helps prevent antibodies that are elicited from binding to conserved epitopes (low antigenicity). Neutralizing antibodies against conserved epitopes are a particular threat to HIV-1, as escape *via* mutation is more difficult than for variable regions of the envelope spikes.

In the mature, 'untriggered' spike, much of the surface of gp41 is masked by gp120. After receptor engagement, during which gp41 is perhaps most vulnerable, access to the fusion-intermediate state of gp41 is restricted by the

closely opposing membranes. The fusion process itself takes only minutes, which significantly limits access to the transiently exposed epitopes on gp41. In contrast, the postfusion conformation of gp41 is a stable repeating unit that is highly immunogenic. This postfusion conformation appears to be so radically different from the fusion-intermediate conformations of gp41 that elicited antibodies are unable to bind to the intermediate conformations and therefore are unable to inhibit fusion.

This description was given a molecular underpinning in 1997, with the X-ray crystal structures of the putative postfusion form of

Michael B. Zwick, Erica O. Saphire and Dennis R. Burton are in the Departments of Immunology and Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA.
e-mail: burton@scripps.edu

HIV-1 gp41 (refs. 1–3). Before then, it was known that the N- and C-terminal portions of the external domain of gp41 contain heptad repeat sequences. Peptides corresponding to the N-terminal (N-HR) and C-terminal (C-HR) heptad repeat regions of gp41, notably T21 (DP107) and T20 (DP178), respectively, could inhibit HIV-1 entry into cells^{4–6}. T20, rebranded as enfuvirtide, was approved by the US Food and Drug Administration for patients last fall.

To overcome the technical hurdles of crystallizing gp41, it was necessary to truncate the protein to make it stable and soluble. Each of the three laboratories crystallized a proteolytically stable core of the extraviral domain of gp41; (devoid of the N-terminal hydrophobic fusion peptide, a disulfide-loop region linking the N-HR and C-HR regions, and the membrane-proximal region).

Although each of the groups crystallized fragments of different sizes, the resulting structures were nearly identical. gp41 was found to have a trimeric core consisting of parallel α -helices corresponding to the N-HR region of gp41. Packed around this trimeric core were three, slightly oblique C-HR α -helices, arranged antiparallel to the inner core to create a trimer of heterodimers, or a six-helix bundle (Fig. 1).

Notably, this arrangement had also been observed in the pH-induced conformation of the influenza hemagglutinin molecule⁷ and the transmembrane subunit of Moloney murine leukemia virus⁸. The structural motif has since been found in Ebola virus, Newcastle disease virus, respiratory syncytial virus and the coronavirus linked to severe acute respiratory syndrome^{9,10}. The motif is a striking example of convergent evolution, and has revealed a vulnerability to be exploited in drug design.

The gp41 structures provide an understanding of how T20 and other gp41-targeted fusion inhibitors work, and shed light on the difficulties of designing vaccines that elicit neutralizing antibodies to gp41.

The T20 peptide corresponds to the C-HR region of gp41 and inhibits HIV-1 entry by packing against the outer grooves on the trimer of α -helices, thereby preventing the formation of six-helix bundles¹¹. Resistance to T20 has been documented, typically involving one or two substitutions along the outer grooves in the N-HR region of gp41.

The relatively new drug T1249, a peptide analog of T20, is an effective inhibitor of T20-resistant isolates of HIV-1 and is in clinical trials (<http://www.trimeris.com/pipeline/clinical.html>). Resistance to T1249 is anticipated but has yet to be documented. Peptides corresponding to the N-HR region of gp41 inhibit six-helix bundle formation by targeting the C-HR region. Monomeric N-HR peptides such as T21 are

prone to aggregation, making them much less potent than T20; potency is increased substantially with soluble, trimeric N-HR constructs^{12,13}.

In principle, antibodies that could bind to the N-HR inner helical core after receptor activation should also block entry of HIV-1, similar to the fusion inhibitors. In terms of vaccine design, however, eliciting neutralizing antibodies to this target may be difficult because of spatial and temporal limitations during the fusion process. The broadly neutralizing antibodies defined to date seem to recognize gp41 epitopes that are close to the viral membrane but accessible during fusion^{14,15}. Vaccinologists await structures for the fusion intermediates.

The gp41 structures have stimulated progress in the design and selection of fusion inhibitors. In one study, an engineered trimer of the inner N-HR helical peptides was used to select peptides from a random phage display library. These peptides, selected as proteolytically stable mirror-image or D-peptides, potentially inhibit HIV-1 entry into cells¹³. In another study, investigators constructed a five-helix bundle that lacks one of the outer C-HR peptides. The five-helix bundle potently inhibited HIV-1 entry, probably by binding to a single C-HR region on HIV-1 gp41 and preventing six-helix bundle formation¹⁶. In a particularly inspired demonstration of structure-based design, the solvent-exposed residues on one face of the α -helix of a gp41 C-HR peptide were substituted onto the solvent-exposed face of another helical scaffold, GCN-4. Indeed, the resulting hybrid peptide blocked HIV-1 entry into cells¹⁷.

There are major problems with peptide therapeutics, including the cost and difficulty of manufacturing them. T20 currently costs \$20,000 per year per patient, and the supply has yet to satisfy the demand. In an initial effort to develop nonpeptide inhibitors of HIV-1 entry using the crystal structures of gp41, molecular docking algorithms have been used to search small molecule libraries *in silico* for compounds that might bind in a deep hydrophobic cavity in the inner N-HR peptide core. Some of the best-ranking compounds inhibited HIV-1 infectivity with modest potency *in vivo*¹⁸.

Another group used a short C-HR peptide as a scaffold for attaching a nonpeptide combinatorial library onto its N terminus. The library was screened against an engineered trimer of N-HR peptides, and a hybrid ligand, composed of a synthetic moiety linked to the C-HR peptide, was selected that was able to inhibit HIV-1 envelope-mediated cell-cell fusion¹⁹. Further optimization of the nonpeptide moiety is envisioned as a route to a stand-alone, small-molecule fusion inhibitor.

In another approach, a small molecule was designed to mimic a hydrophobic surface along one face of a heptad repeat α -helix; the pro-teomimetic was able to partially disrupt the six-helix bundle and inhibit HIV-1 mediated fusion²⁰.

As the hidden secrets of gp41 continue to be revealed, new leads for HIV-1 therapies and vaccines will emerge. Interesting new targets for intervention may be uncovered by obtaining detailed structural information on the native and fusion-intermediate conformations of gp41. It will also be useful to elucidate the influence on gp41 structure and function of the segments that were omitted in the gp41 crystal structures.

Small-molecule drug design has been extremely effective in inhibiting the HIV-1 protease and reverse transcriptase, but it remains to be seen how well it will work against the envelope complex. Since drug resistance is so prevalent for HIV-1 and has already been documented for T20, drug cocktails may be the best solution. Indeed, the range of potential targets on gp41 and a growing arsenal of other entry inhibitors²¹ make synergy with existing and future drug combinations an exciting possibility.

This material is part of Nature Medicine's 10 year anniversary series. For more content related to these special focus issues, please see http://www.nature.com/nm/special_focus/anniversary/index.html

- Chan, D.C., Fass, D., Berger, J.M. & Kim, P.S. *Cell* **89**, 263–273 (1997).
- Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J. & Wiley, D.C. *Nature* **387**, 426–430 (1997).
- Tan, K., Liu, J., Wang, J., Shen, S. & Lu, M. *Proc. Natl. Acad. Sci. USA* **94**, 12303–12308 (1997).
- Jiang, S., Lin, K., Strick, N. & Neurath, A.R. *Nature* **365**, 113 (1993).
- Wild, C. *et al. Proc. Natl. Acad. Sci. USA* **91**, 12676–12680 (1994).
- Wild, C.T., Shugars, D.C., Greenwell, T.K., McDanal, C.B. & Matthews, T.J. *Proc. Natl. Acad. Sci. USA* **91**, 9770–9774 (1994).
- Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. *Nature* **371**, 37–43 (1994).
- Fass, D., Harrison, S.C. & Kim, P.S. *Nat. Struct. Biol.* **3**, 465–469 (1996).
- Colman, P.M. & Lawrence, M.C. *Nat. Rev. Mol. Cell Biol.* **4**, 309–319 (2003).
- Bosch, B.J., van der Zee, R., de Haan, C.A., & Rottier, P.J. *J. Virol.* **77**, 8801–8811 (2003).
- Melikyan, G.B. *et al. J. Cell Biol.* **151**, 413–423 (2000).
- Bewley, C.A., Louis, J.M., Ghirlando, R. & Clore, G.M. *J. Biol. Chem.* **277**, 14238–14245 (2002).
- Eckert, D.M., Malashkevich, V.N., Hong, L.H., Carr, P.A. & Kim, P.S. *Cell* **99**, 103–115 (1999).
- Binley, J.M. *et al. J. Virol.* **77**, 5678–5684 (2003).
- Zwick, M.B. *et al. J. Virol.* **75**, 10892–10905 (2001).
- Root, M.J., Kay, M.S. & Kim, P.S. *Science* **291**, 884–888 (2001).
- Sia, S.K. & Kim, P.S. *Proc. Natl. Acad. Sci. USA* **100**, 9756–9761 (2003).
- Debnath, A.K., Radigan, L. & Jiang, S. *J. Med. Chem.* **42**, 3203–3209 (1999).
- Ferrer, M. *et al. Nat. Struct. Biol.* **6**, 953–960 (1999).
- Ernst, J.T. *et al. Angew. Chem. Int. Ed. Engl.* **41**, 278–281 (2002).
- Moore, J.P. & Doms, R.W. *Proc. Natl. Acad. Sci. USA* **100**, 10598–10602 (2003).