

Molecular dynamics simulations reveal regions of instability on BG505 SOSIP HIV Envelope glycoprotein

Modestas Filipavicius, Clare Towse, Valerie Daggett

Abstract

HIV envelope (Env) is the sole target of neutralizing antibodies. Soluble and well-ordered Env mimetics are attractive vaccine candidates. Here we use molecular dynamics and *ilmm* to determine unstable regions of BG505 SOSIP trimer. BG505 SOSIP is a trimer of monomers of gp120 and truncated gp41 chains linked together by disulphides. MD and RosettaDesign simulations suggest stabilizing the helical HR2 domain of gp41 which interacts with unstable helix on the adjacent gp41 subunit. The stabilization of intersubunit interactions will generate novel homogenous and stable trimers.

Introduction

HIV envelope glycoproteins (Env) sparsely decorate HIV surface and are the only immune system targets that elicit autologous broadly neutralizing antibodies (bNAbs). However, to create a broadly effective vaccine a heterologous response to HIV Env antigen is needed. To that end, soluble and thermodynamically stable alternatives to Env have been engineered. Achieving higher stability remains a challenge, because the protein is expressed as a trimer of gp160 subunits, and each gp160 consists of noncovalently linked gp41 and gp120 subunits (Gao et al., 2005). The transmembrane and membrane proximal external region domains are not stable in solution, so most constructs have expunged those segments. One common problem Env researchers encounter when purifying the protein is dissociation into monomers or formation of aberrantly linked dimers. The

best model so far has a disulphide bond between gp41 (C605) and gp120 (C501), and an I559P mutation which increases helix breaking, together known as SOSIP modifications (Sanders et al., 2002). SOSIP trimers require furin cleavage at six-histidine site that was added too.

The recent crystal structure resolution of one particular, SOSIP-modified isolate, BG505, has reinvigorated the field to develop HIV vaccine, since BG505 SOSIP trimer elicits Tier II Nabs (Julien et al., 2013). It also recapitulates autologous neutralizing antibody binding found in the sera of patient #BG505 (Sanders et al., 2013).

Multiple approaches have been taken to stabilize BG505 SOSIP trimer into a more stable immunogen. For instance, Richard Wyatt's group introduced a cleavage independent BG505 (Sharma et al., 2015) and HIV subtype B JRFL- and subtype C 1605-derived trimers (Guenaga et al., 2015).

Recent Hydrogen-Deuterium exchange (HDX) experiments elucidated the solution dynamics of Env secondary structures and NAb-binding epitopes (Guttman et al., 2014; Sanders et al., 2015). The variable loops (V1/V2/3), binding sites of most of neutralizing antibodies, cover up the HIV co-receptor (CXCR/CCR5) binding site at prefusion state (Figure 1). These regions are most heavily glycosylated. Upon CD4 binding the protein changes conformations of bridging sheet which in turn disrupts V1/V2 and V3 interactions and exposes coreceptor binding site.

One potential weakness in SOSIP design is C-terminal helical portion of HR2, residues 649-664, which is unusually de-protected from hydrogen to deuterium exchange, implying that its secondary structure is unstable. Most importantly, this is the region that mediates trimeric intersubunit interactions of gp41, by creating contacts between HR2 on one monomer with the N-terminus of gp41 of adjacent monomer. By stabilizing this region we could create a more stable homogeneous trimer.

To this goal we utilized RosettaDesign to run a mutagenesis protocol and introduced multiple mutations on c-terminus of HR2 and around it. We have next pursued a molecular dynamics simulation with *i/mm* towards a fully assembled gp41 trimer with relevant gp120 regions inside it. MD simulation time increases exponentially with increasing number of atoms in the system, hence choosing only the relevant loops from gp120 is of paramount importance. We ran monomeric gp41 first (153 residues), followed by gp41+gp120 (together called gp140, and not gp160 since the MPER and TMD of gp41 are missing). Next, we compared the relative stabilities of gp41 alone and gp140, finding the sweet spot for truncating gp140 down to the full gp41 chain plus the interacting helices from gp120 looped by a polyglycine (together called gp140_trunc). After these analyses we proposed a trimeric gp140_trunc structure (around 600 aa) which could be simulated in the future (WT) as well as its triple mutant (gp140_trunc_mt).

Methods

Homology modelling with Modeller

The crystal structure of unliganded BG505 SOSIP (Do Kwon et al., 2015) has two chains corresponding to gp120 (chain G, 464 out of 487 residues resolved) and gp41 (chain B, 127 out of 157 residues resolved). We modelled-in missing residues with Modeller 9.15 with EasyModeller 4 GUI, using 10 decoys. We chose the best loop model based on low DOPE score and agreement with Cryo-EM (Lyumkis et al., 2013). All models were prepared in 4ZMJ protomeric, but not biologic assembly, since we simulated various truncations of gp140 protomer. After preparing the final gp41 construct with relevant gp120 loops, we modeled in missing residues on the biological assembly and excised the relevant gp41+gp120_trunc structure.

RosettaDesign

We used RosettaDesign (with Rosetta's Python-based interface, PyRosetta) to find most favorable single aa mutations on HR2 (residues 649-664) and an opposing helix from another monomer (525-540) (Figure 2). The full structure of 4ZMJ was inputted, but we only allowed the *mover* protocols in gp41 region. The resfile specified the use of all amino acids (except cysteine) and rotamer repacking (ALLAA). To generate reference state, we did the fixed backbone repacking and minimization. Minimization consisted of small, shear and min moves. Next, we took 20 lowest scoring mutants and tested which ones in duplicates or triplicates were most stabilizing.

Molecular Dynamics with illm

Simulations were done with *in lucem* molecular mechanics (*illm*) (Beck et al., 2008). Design pdb files were edited to account for disulphide bonds (SSBOND), gp140 had 11 in total, then minimized using 1000 steps of Steepest Descent minimization protocol, hydrogens were restored and a water box that surrounds protein by at least 10 Å was added (Levitt et al., 1995). Temperature was set to 310 K (physiological T), water density was adjusted to 0.985506 g/ml and water was added at most within 1.8 Å from the protein surface. Next, the solvated protein went through a 1000-step SD minimization at a given T, plus 1 ps of raising T by 40K.

A restart file was created with velocity, position and acceleration terms for all simulation atoms. Simulations were run for 10 ns for gp41 and gp140_trunc, and 13 ns for gp140. We used the standard 2 fs timestep, structures were saved every 1 ps for later analysis.

MD simulation analysis

Overall and individual residue RMSD and RMSF were calculated every 2 fs, with the MD simulation starting structure taken as a reference. 3D structure snapshots were recorded to individual PDB files every 100 ps. Secondary structure interactions and motifs were classified as loops, 3/10, alpha and pi helices, and alpha, beta, sheets or salt bridges, and visualized with DSSP package as outlined in McCully et al., 2013). Contact analysis (VCONT) revealed the contacts between residue side chains, which were classified as non-native, if they did not appear in the original structure. Asymmetric contacts plot has non-native contacts on the top left and native on bottom right, colored according to fraction of time in contact. Total contacts plot reveals native contacts before the simulation.

Gp140_trunc design

Gp140_trunc was designed after analyzing gp140 and gp41 MD data first. We used full gp41 sequence (chain B in 4ZMJ). We first cut gp120 sequence (chain G) at residue E49, resulting in A31-E49 chain, which corresponds to the very beginning of gp120's N-terminus. Next, A48 and E49 were mutated to Glycines to make a polyglycine linker with another part of gp120. Third, we excised residues T50-V488, and kept V489-R513 region, which corresponds to C-terminus of gp120 (Figure 3). V489A was introduced to decrease beta-sheet formation next to the linker.

Results

RosettaDesign revealed that M535 mutations on gp41's HR2 region were most stabilizing, especially to Ile and Val. However, when M535 was allowed to mutate to any aa and interact with the opposing chain from another subunit, then M535W – N656Y double mutant had lowest energy by -20, which is explained by Pi-stacking interaction between the two aromatic rings and hydrogen

115 bonds between Trp nitrogen and Tyr alcohol (Figure 2 D, E). Also, the third mutation, K655Y,
116 puts a large hydrophobic group into hydrophobic core.

117 RMSD analysis showed that gp41's Ca have moved the most from the minimized starting structure
118 (Figure 4). It had the highest RMSD value among the three constructs, around 7 Å. Meanwhile
119 gp140, a 634 aa structure, was the most stable, its RMSD for the last 2 ns was on average 4.5 Å.
120 Low RMSD is not surprising, since gp140 constitutes around 50% of typical BG505 SOSIP
121 transfection product, prompting the researchers to come up with new methods of purifying the
122 trimeric, but not monomeric protein. For example, antibody PGT145 binds at the trimer apex
123 quaternary structure and therefore is specific for the trimer only.

124 Gp140_trunc construct that differs from gp41 only by 44 amino acids was stable for the first 6 ns,
125 but then steeply became more disordered. This is due to the inserted gp120 loops having floppy
126 ends.

127 RMSF analysis indicated that truncated gp140 moves less than gp140 and gp41 alone (Figure 5).
128 This means that the truncated model accurately represents the “native” gp140 movements.
129 However, there are significantly higher motions at W(571)GIKQARV sequence which constitutes
130 the N-terminus of HR1 (top of HR1 helix, right after the green loop, Figure 5).

131 Next, RMSF data shows higher motions in gp140 compared to gp41 alone (Figure 6). Here 1st
132 residue on gp41 simulation corresponds to 482nd in gp140. Gp140 was most unstable in 527-538
133 region, which corresponds to homology modelled-in missing loop that is at the interface of trimeric
134 assembly. Also, the beginning of gp41 sequence had a flexible region in 482-634, with an average
135 RMSF of 2.5 Å. This region is right after six-Arg furin cleavage site, which explains high
136 movements right before and after that site.

Secondary structure analysis has revealed that gp41 alone has gained alpha-helical structure at residues 19-25 (MGAASMT), which correspond to residues 529-536 in 4ZMJ, while gp140 and gp140_trunc have lost all secondary structure there (Figure 7). This finding is critical for our future design, since the second part of the helix mediates inter-subunit assembly via interactions with HR2, and M535W was chosen by Rosetta as the key stabilizing mutation here. Furthermore, MD simulation proves that AASM(535)TLTVQARNL sequence does not have a stable helical secondary structure (Figure 9). Contrary to what the crystal structure 4ZMJ is showing, the single helix breaks into two helices via T536. Therefore stabilizing M535 is of paramount importance. Unique for gp140_trunc was a permanent salt bridge between E23-R118 (460-555 in 4ZMJ) and a transient K21 and Q110 (458-547) bridge that happens in around 50% of simulation time (Figure 8, 11). This interaction results in kinking of HR1 helix, which is not observed in the other models. HR1 is pushed upwards compared to gp140, which makes sense, since it has no clashes with its upstairs neighbor gp120 after the latter structure was trimmed away. Hence, in the next design iteration we might mutate out K458 to abrogate the transient salt bridge.

Discussion

MD simulation has recapitulated our HDX experimental findings that the C-terminus of the HR2 has a disordered secondary structure. Although here we simulated a monomeric structure, our experimental data for trimeric BG505 SOSIP shows the same intrinsic flexibility of the helix between the oligomeric arrangements. The HR2 helix could play a role in stabilizing inter-subunit interactions between adjacent gp41 domains, hence making the trimer more stable and a better potential antigen. Similarly, the mutations that we are proposing here, should not reduce the antigenicity profile of the Env trimer, since none of the known bNAbs bind at gp41 interface. Yet,

160 at the same time, the region is extremely conserved among different isolates, which makes us
161 predict the elicitation of bNAbs against the newly stabilized region.

162 M535 mutation was implicated in the literature too. BG505 M535, N543 and K567 residues have
163 been reported to contribute to the trimerization efficiency of soluble gp140 and the thermal
164 stability of Env trimers on virus particles (Dey et al., 2007; Leaman and Zwick, 2013). Yet none
165 of the recent SOSIP designs have attempted to utilize these observations.

166 The next step in this project would be preparing a homology model of the whole gp140 trimer, in
167 order to get the correct restoration of intrinsically disordered
168 IVQQQ(552)SNLLRAPEEAQQHLLKL in gp41. This cannot be done in Chimera, since we'll be
169 using six chains (B and G trimers), so Modeller will be used instead. We might choose to remove
170 six-Arg furin cleavage site too. Next, we will cut the gp120 sequence in the same way we did with
171 gp140_trunc. Finally, we will simulations both the WT and triple mutant gp41 trimers.

172 Furthermore, even if we did not have enough computational resources to roll out the whole project,
173 a large gp140 simulation revealed interesting features about Envs in general, since the gp140
174 monomer has an almost identical Xtal structure and antibody binding profile as the trimer (in fact,
175 BG505 SOSIP transfection preps typically produce 30% monomer). For instance, the loops that
176 were not resolved by crystallography were moving the most, which is not surprising since they
177 have sequences characteristic of intrinsically disordered proteins. Good examples are the
178 TSVQ(402)GSNST and especially IVQQQ(552)SNLLRAPEEAQQHLLKL, peptides dominated
179 by charged and polar residues. These sequences play a key role in stabilizing the fusion protein
180 when Env opens up towards the target membrane. Interestingly, the latter sequence was recently
181 shown by Andrew Ward (Lee et al., 2016) to form an alpha-helix in HIV clade B isolate, JRFL,
182 upon binding to a neutralizing antibody PGT151 (Figure 10). It would be exciting to further

183 explore this disordered region using MD simulations. We also witnessed the “IP” element of
184 SOSIP mutations. The non-resolved loop of HR1 after being homology modeled into crystal
185 structure, has helical tendencies, but SOSIP’s I559P mutation prevents that.
186 Finally, it seems that the most flexible regions of gp140 correlate well with HDX protection “heat
187 maps”. It would be interesting to color each residue’s RMSF by spectrum from blue to red, and
188 compare that to maps produced by HDX (Figure 12). All together, the experimental data from
189 HDX and *i/lm* MD simulations suggest a novel Env trimer design strategy - stabilizing the helical
190 HR2 domain of gp41 which interacts with unstable helix on the adjacent gp41 subunit. The
191 stabilization of intersubunit interactions will generate homogeneous and highly stable trimers.
192

193 **References:**

- 194 Beck DA, Daggett V. Methods for molecular dynamics simulations of protein folding / unfolding
195 in solution. *Methods*, 34: 112-120, 2004.
- 196 Beck, D., McCully, M., Alonso, D. & Daggett, V., 2000-2016. in *lucem molecular mechanics*
197 software, University of Washington
- 198 Dey AK, David KB, Klasse PJ, Moore JP (2007) Specific amino acids in the N-terminus of the
199 gp41 ectodomain contribute to the stabilization of a soluble, cleaved gp140 envelope glycoprotein
200 from human immunodeficiency virus type 1. *Virology* 360: 199–208.
- 201 Gao, F., Weaver, E. A., Lu, Z., Li, Y., Liao, H. X., Ma, B., ... & Decker, J. M. (2005).
202 Antigenicity and immunogenicity of a synthetic human immunodeficiency virus type 1 group m
203 consensus envelope glycoprotein. *Journal of virology*, 79(2), 1154-1163
- 204 Guenaga, J., de Val, N., Tran, K., Feng, Y., Satchwell, K., Ward, A. B., & Wyatt, R. T. (2015).
205 Well-ordered trimeric HIV-1 subtype B and C soluble spike mimetics generated by negative
206 selection display native-like properties. *PLoS Pathog*, 11(1), e1004570
- 207 Guttman, M., Garcia, N. K., Cupo, A., Matsui, T., Julien, J. P., Sanders, R. W., ... & Lee, K. K.
208 (2014). CD4-induced activation in a soluble HIV-1 Env trimer. *Structure*, 22(7), 974-984.
- 209 Jeff Gray (JHU), PyRosetta Lecture series
- 210 Julien, J. P., Cupo, A., Sok, D., Stanfield, R. L., Lyumkis, D., Deller, M. C., ... & Ward, A. B.
211 (2013). Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science*, 342(6165), 1477-
212 1483.
- 213 Leaman DP, Zwick MB (2013) Increased functional stability and homogeneity of viral envelope
214 spikes through directed evolution. *PLoS Pathog* 29: e1003184
- 215 Lee, J. H., Ozorowski, G., & Ward, A. B. (2016). Cryo-EM structure of a native, fully
216 glycosylated, cleaved HIV-1 envelope trimer. *Science*, 351(6277), 1043-1048.
- 217 Levitt M, Hirshberg M, Sharon R, Daggett V. Potential-Energy Function and Parameters for
218 Simulations of the Molecular-Dynamics of Proteins and Nucleic-Acids in Solution. *Computer*
219 *Physics Communications*, 91: 215-231, 1995.
- 220 Sanders, R. W., Van Gils, M. J., Derking, R., Sok, D., Ketas, T. J., Burger, J. A., ... & Arendt, H.
221 (2015). HIV-1 neutralizing antibodies induced by native-like envelope trimers. *Science*,
222 349(6244), aac4223.
- 223 Sanders, R. W., Vesanen, M., Schuelke, N., Master, A., Schiffner, L., Kalyanaraman, R., ... &
224 Lu, M. (2002). Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein
225 complex of human immunodeficiency virus type 1. *Journal of virology*, 76(17), 8875-8889
- 226 Sharma, S. K., de Val, N., Bale, S., Guenaga, J., Tran, K., Feng, Y., ... & Wyatt, R. T. (2015).
227 Cleavage-independent HIV-1 Env trimers engineered as soluble native spike mimetics for
228 vaccine design. *Cell reports*, 11(4), 539-550
229