Generation and Characterization of an HIV-1 Subtype C Transmitted and Early Founder Virus Consensus Sequence

Mark Killick, Alexio Capovilla, and Maria A. Papathanasopoulos

Abstract

The tight bottleneck during HIV-1 transmission generally results in only a single virus variant being transmitted. Investigation of the HIV-1 envelope glycoprotein (Env) can identify vulnerabilities of transmitting viruses that can be targeted by vaccines designed to elicit protection against global HIV-1. This study generated an HIV-1 subtype C consensus transmitted and early founder virus Env (Env_{FVC}) after detailed sequence analysis of 1,894 *env* genes obtained from 80 acutely infected individuals from South Africa, Malawi, and Zambia. The inferred Env_{FVC} sequence incorporates characteristics of transmitted and early founder viruses and results in the expression of a functional and conformationally intact Env. Overall, the “subtype-based” or “region-based” Env_{FVC} described here can be used in the development of a useful immunogen for novel vaccine design.

A MAJOR CHALLENGE IN THE development of broadly effective HIV-1 vaccines is that of contending with the extensive genetic variability found in circulating viruses. Research into early transmission events and acute HIV-1 infection has revealed a pronounced genetic bottleneck in the transmission of viral quasiespecies from donor to recipient, and in the majority of cases productive infection arises as a result of a single viral variant, termed the founder virus.^{1–6} In some instances, infection by multiple genetic variants is significantly associated with inflammatory genital infection in the newly infected individual.⁴ Interestingly, results from macaques infected with the chimeric simian/human immunodeficiency virus SHIV_{sf162p4} showed a correlation between the dose administered and the number of variants transmitted.⁷ The abovementioned studies describe transmitted and early founder viruses in the context of the HIV-1 envelope glycoprotein (Env), the protein responsible for viral entry and the target of primary anti-HIV-1 immune responses and neutralizing antibodies. This is a significant finding since it indicates that there is a finite window of HIV-1 vulnerability that can be exploited by the appropriate vaccine-elicited immune responses⁵ that provide protection from transmission.

It is unknown whether the transmitted viruses carry an advantage over variants that fail to transmit. Sequence analysis has shown that transmitted variants tend to contain Env with

shorter, more compact variable loop structures and fewer glycosylation sites as compared to viruses present during later stages of chronic infection. The more compact Env structures of the founder virus variants are thought to increase fitness in a transmission setting due to greater exposure of the CD4 binding site, although this increases Env sensitivity to neutralizing antibodies.^{2,3} As infection progresses, escape from neutralizing antibodies is the main selecting force driving sequence diversity and glycosylation in the Env variable loops.⁸ Founder virus populations also have an increased affinity for the $\alpha_4\beta_7$ integrin, which is proposed to lead to the enhancement in the transmission fitness of these variants through improved cell-to-cell transmission or specific infection of lymphocytes that home to gut-associated lymphoid tissue (GALT).^{9,10}

A greater understanding of transmitted viruses may help identify Env vulnerabilities that can be targeted by vaccines designed to elicit protective antibodies against clinically relevant HIV-1. In addition, the extreme genetic bottleneck selects for an HIV-1 genome that encodes more relevant sequences for vaccine design than viruses that are the basis for current vaccines. HIV-1 subtype C has emerged as the most prevalent strain worldwide, and may be associated with a greater propensity for transmission than other HIV-1 group M subtypes.¹¹ Thus, this study aimed to evaluate all the

available HIV-1 subtype C founder virus Env sequences and to generate and characterize a transmitted/early founder virus consensus Env sequence for incorporation into a vaccine immunogen construct.

Sequence data previously published using single genome amplification of HIV-1 Env sequences from patients acutely infected with HIV-1 subtype C were used to generate an HIV-1 subtype C founder virus Env sequence. The patient cohort was composed of 80 individuals from South Africa (40),¹ Malawi (29),¹ and Zambia (11)⁴ all acutely infected with HIV-1 subtype C, as classified according to the Fiebig stage classification system, following heterosexual transmission. Twenty-four of the individuals were in Fiebig stage I/II (5–10 days postinfection), 20 patients were in stage III/IV (13–19 days postinfection), and 36 patients were identified as being in the latter stages of acute infection, Fiebig stages V/VI (88 days onward postinfection). Overall, in 70% of the patients identified among these three cohorts, productive infection occurred as a result of a single transmitted viral variant. For the remaining 30% of patients, productive infection was the result of simultaneous transmission of multiple variants (mean 2.4, range 2–5).

Of the published 1,927 *env* sequences isolated from these patients, 1,894 sequences were available and were downloaded as fasta files from the Los Alamos National Laboratory HIV sequence database (www.hiv.lanl.gov) and analyzed further. An average of 22 *env* sequences, with a range of 15–42, were obtained for each patient from 40 South African and 29 Malawian patients (accession numbers FJ443128–FJ444632¹). From 11 Zambian patients, an average of 28 *env* sequences (range of 24–49) was obtained for each patient (accession numbers: EU166413–EU166473, EU166482–EU166517, and EU166544–EU166919⁴). It should be noted that sequence analysis of *env* isolated from patient ZM231F revealed it was not HIV-1 subtype C, and this was excluded in our analysis.⁴

The individual *env* sequences for each patient were initially aligned using Clustal X 2.0.11 (www.clustal.org). Consensus *env* sequences were manually generated for each quasispecies/variant present in each patient, and subsequently translated in GeneDoc 2.6.002 (www.psc.edu/biomed/genedoc) to obtain the Env amino acid sequence. Subsequently, consensus Envs from all 80 patients were aligned in Clustal and viewed in GeneDoc to manually generate a geographic Env consensus sequence from each country. The ProSite search function available in GeneDoc was used to identify potential N-linked glycosylation sites (PNGS), known broadly neutralizing antibodies (bNAb) epitopes/residues, and other relevant functional sites. The South African, Malawian, and Zambian consensus Env sequences were then analyzed to manually generate a final Env consensus sequence based on the information generated above. This is subsequently termed the HIV-1 subtype C consensus founder virus Env, or Env_{FVC}. Phylogenetic and molecular evolutionary analysis was achieved using a multiple alignment of the newly generated Env_{FVC} gp160 with references from HIV-1 subtypes A to K, CRF01_AE, CRF02_AG, consensus C and ancestral C Env sequences, and by generating a neighbor-joining tree using MEGA 4 (www.megasoftware.net).

Overall, an HIV-1 subtype C founder virus consensus Env sequence was successfully generated after detailed analysis of 1,894 *env* sequences derived from the 80 acutely infected

patients (Fig. 1). Phylogenetic tree analysis revealed that Env_{FVC} was most closely related to the inferred ancestral and consensus C Env sequences rather than those from circulating viruses (results not shown). The amino acid length of the variable loops and number of PNGS were calculated. The Env_{FVC} sequence was generated with shorter V1/V2, V4, and V5 loop sequences, as well as one less PNGS site in the V1/V2 loop structure and additional PNGS included in the C2 region and gp41 ectodomain, as compared to the average number found in early transmitted subtype C viruses. These differences did not exclude any relevant known functional epitopes. Importantly, the Env_{FVC} sequence contained the relevant residues for CD4 binding, including the loop D of the constant region 2 (C2) (all numbering according to HXBc2: residues 275–283), the CD4 binding loop in the C3 region (residues 364–374), the bridging sheet ($\beta_{20/21}$ hairpin) (residues 425–430), and β_{23} (residues 455–461) in the C4 region and the β_{24} – α_5 connection in the C5 region (residues 469–475; Fig. 1, red boxes with critical residues shaded red). Residues within these regions form the epitopes of the CD4bs-directed broadly neutralizing antibodies such as IgG1b12, VRC01, PGV04, and 3BNC117.^{12–15} Glycan-dependent epitopes such as those required for the monoclonal antibodies PG9/PG16,¹⁶ CH01-04,¹⁷ and PGT141-145¹⁸ (residues N156 and N160) and the PGT family (PGT 125–128, 130, and 131¹⁸) of bNAbs (residues N301 and N332) were present (Fig. 1, green boxes). Interestingly, if PNGS at positions 160 and 332 are not present during primary HIV-1 infection, they are subsequently introduced as a direct response to neutralizing antibody-mediated pressure. The relevance of PNGS at positions 160 and 332 in the context of immunogen design should be elucidated further. In addition, the linear epitope for the $\alpha_4\beta_7$ integrin binding site (LDI/V—residues 179–181) and MPER directed bNAb 4E10 (NWFDIT—residues 671–677) were intact (Fig. 1, black background, with pink or yellow text, respectively).

During transmission, it is widely accepted that there is also strong selection for viral variants using the CCR5 coreceptor.¹⁹ As expected, the Env_{FVC} V3 loop was predicted to use the CCR5 for cell entry, using both the C-PSSM_{sinsi} (<http://indra.mullins.microbiol.washington.edu/webpssm/>) and geno2pheno [coreceptor] (<http://coreceptor.bioinf.mpi-inf.mpg.de/>) algorithms.

Lastly, the native HIV-1 subtype C consensus founder virus gp160 nucleotide sequence (normal HIV codon use) was generated at GeneArt (Life Technologies), cloned into the mammalian expression vector pcDNA3.1 (Invitrogen, Life Technologies), and used to create an Env_{FVC} pseudotyped virus as per standard protocols. *In vitro* phenotypic inhibition assays using 200TCID₅₀ of the Env_{FVC} pseudovirus revealed an IC₅₀ of 0.05 μ g/ml for the entry inhibitor enfuvirtide, well within previously reported ranges.²⁰ Thus, the generation of an infectious pseudovirus indicates that the inferred HIV-1 subtype C consensus founder virus Env sequence codes for a functional Env_{FVC}. Codon optimized (humanized) gp120 and gp140 constructs were also generated at GeneArt (Life Technologies) and successfully used to express and purify functional and conformationally intact Env_{FVC} monomers and trimers, respectively.

Overall, our analysis confirmed that an HIV-1 subtype C consensus founder virus Env sequence with characteristics of transmitted and early founder viruses can be used to express a

FIG. 1. Alignment of the envelope glycoprotein (Env) amino acid consensus sequences derived from HIV-1 subtype C acutely infected individuals from Malawi (MW_consens), South Africa (ZA_consens), and Zambia (ZM_consens) used to infer a subtype C founder virus Env consensus (Founder_Vi) sequence. The variable loop structures, signal peptide and Env precursor sites are all indicated above the sequence alignment. Potential N-linked glycosylation sites (PNGs) are shaded in blue, $\alpha_4\beta_7$ integrin binding site tripeptide motifs (LDI/V) are indicated in pink, the tetrapeptide crown of the V3 loop structure (GPGQ/R) is shown in green, and the linear epitopes of the two broadly neutralizing antibodies (bNAb) targeting the gp41 MPER region 2F5 (ELDKWA) and 4E10 (NWFDIT) are shown with a black background and are highlighted green and yellow, respectively. Regions boxed in red containing the residues important for CD4 binding and CD4 binding site-directed bNAb epitopes, critical residues are shaded. Glycan-dependent epitopes targeted by bNAb PG9/PG16 and PGT family antibodies are boxed in green. The (•) indicates identity to the consensus sequence and (-) indicates insertions/deletions at a particular position.

functional and conformationally intact Env. Previous research has shown that using a consensus Env gene approach improves the breadth of antibody and cell-mediated responses, although the level of improvement is still somewhat limited.^{21–23} The “subtype-based” or “region-based” Env_{FVC} described here can be used in the development of a useful immunogen for novel vaccine design.

Sequence Data

The full length HIV-1 subtype C founder virus consensus gp160 nucleotide sequence was submitted to GenBank using Sequin V13.05 (www.ncbi.nlm.nih.gov) and is available under accession number KM066116. The alignments are available upon request.

Acknowledgments

Research funding from the South African HIV/AIDS Research and Innovation Platform (SHARP) is gratefully acknowledged.

Author Disclosure Statement

No competing financial interests exist.

References

1. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, *et al.*: Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-Poisson distribution of transmitted variants. *J Virol* 2009;83:3556–3567.
2. Chohan B, Lang D, Sagar M, Korber B, Lavreys L, Richardson B, and Overbaugh J: Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. *J Virol* 2005;79:6528–6531.
3. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, Denham SA, *et al.*: Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 2004;303:2019–2022.
4. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, *et al.*: Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog* 2009;5:e1000274.
5. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, *et al.*: Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA* 2008;105:7552–7557.
6. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, *et al.*: Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J Virol* 2008;82:3952–3970.
7. Varela M, Landskron L, Lai RP, McKinley TJ, Bogers WM, Verschoor EJ, *et al.*: Molecular evolution analysis of the human immunodeficiency virus type 1 envelope in simian/human immunodeficiency virus-infected macaques: Implications for challenge dose selection. *J Virol* 2011;85:10332–10345.
8. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, *et al.*: Antibody neutralization and escape by HIV-1. *Nature* 2003;422:307–312.
9. Cicala C, Martinelli E, McNally JP, Goode DL, Gopaul R, Hiatt J, *et al.*: The integrin $\alpha 4\beta 7$ forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1. *Proc Natl Acad Sci USA* 2009;106:20877–20882.
10. Nawaz F, Cicala C, Van Ryk D, Block KE, Jelacic K, McNally JP, *et al.*: The genotype of early-transmitting HIV gp120s promotes $\alpha (4) \beta (7)$ -reactivity, revealing $\alpha (4) \beta (7)$ +CD4+ T cells as key targets in mucosal transmission. *PLoS Pathog* 2011;7:e1001301.
11. Arien KK, Abraha A, Quinones-Mateu ME, Kestens L, Vanham G, and Arts EJ: The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *J Virol* 2005;79:8979–8990.
12. Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TYK, *et al.*: Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 2011;333:1633–1637.
13. Zhou T, Georgiev I, Wu X, Yang ZY, Dai K, Finzi AS, *et al.*: Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* 2010;329:811–817.
14. Wu X, Yang ZY, Li Y, Hogenkorp CM, Schief WR, Seaman MS, *et al.*: Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 2010;329:856–861.
15. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, *et al.*: Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* 2011;333:1593–1602.
16. Walker LM, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL, *et al.*: Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 2009;326:285–289.
17. Bonsignori M, Hwang KK, Chen X, Tsao CY, Morris L, Gray E, *et al.*: Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *J Virol* 2011;85:9998–10009.
18. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien J-P, *et al.*: Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature*. 2011;477:466–470.
19. Connor RI, Sheridan KE, Ceradini D, Choe S, and Landau NR: Change in coreceptor use correlates with disease progression in HIV-1 infected individuals. *J Exp Med* 1997;184:621–628.
20. Derdeyn CA, Decker JM, Sfakianos JN, Zhang Z, O'Brien WA, Ratner L, *et al.*: Sensitivity of human immunodeficiency virus type 1 to fusion inhibitors targeted to the gp41 first heptad repeat involves distinct regions of gp41 and is consistently modulated by gp120 interactions with the coreceptor. *J Virol* 2001;75:8605–8614.
21. Gao F, Weaver EA, Lu Z, Li Y, Liao HX, Ma B, *et al.*: Antigenicity and immunogenicity of a synthetic human

- immunodeficiency virus type 1 group M consensus envelope glycoprotein. *J Virol* 2005;79:1154–1163.
22. Kothe DL, Decker JM, Li Y, Weng Z, Bibollet-Ruche F, Zammit KP, *et al.*: Antigenicity and immunogenicity of HIV-1 consensus subtype B envelope glycoproteins. *Virology* 2007;360:218–234.
23. Kothe DL, Li Y, Decker JM, Bibollet-Ruche F, Zammit KP, Salazar MG, *et al.*: Ancestral and consensus envelope immunogens for HIV-1 subtype C. *Virology* 2006;352:438–449.

Address correspondence to:
Maria A. Papathanasopoulos
HIV Pathogenesis Research Laboratory
Department of Molecular Medicine and Haematology
University of the Witwatersrand
7 York Road
Parktown, 2193, Johannesburg
South Africa
E-mail: maria.papathanasopoulos@wits.ac.za