



**HAL**  
open science

## **Buffering deleterious polymorphisms in highly constrained parts of HIV-1 envelope by flexible regions**

Romain Gasser, Meriem Hamoudi, Martina Pellicciotta, Zhicheng Zhou, Clara Visdeloup, Philippe Colin, Martine Braibant, Bernard Lagane, Matteo Negroni

► **To cite this version:**

Romain Gasser, Meriem Hamoudi, Martina Pellicciotta, Zhicheng Zhou, Clara Visdeloup, et al.. Buffering deleterious polymorphisms in highly constrained parts of HIV-1 envelope by flexible regions. *Retrovirology*, 2016, 13 (1), pp.50. 10.1186/s12977-016-0285-6 . inserm-01350513

**HAL Id: inserm-01350513**

**<https://inserm.hal.science/inserm-01350513>**

Submitted on 30 Jul 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH

Open Access



# Buffering deleterious polymorphisms in highly constrained parts of HIV-1 envelope by flexible regions

Romain Gasser<sup>1</sup>, Meriem Hamoudi<sup>1,2</sup>, Martina Pellicciotta<sup>1</sup>, Zhicheng Zhou<sup>3</sup>, Clara Visdeloup<sup>4</sup>, Philippe Colin<sup>3</sup>, Martine Braibant<sup>4</sup>, Bernard Lagane<sup>3</sup> and Matteo Negroni<sup>1\*</sup>

## Abstract

**Background:** Covariation is an essential process that leads to coevolution of parts of proteins and genomes. In organisms subject to strong selective pressure, coevolution is central to keep the balance between the opposite requirements of antigenic variation and retention of functionality. Being the viral component most exposed to the external environment, the HIV-1 glycoprotein gp120 constitutes the main target of the immune response. Accordingly its more external portions are characterised by extensive sequence heterogeneity fostering constant antigenic variation.

**Results:** We report that a single polymorphism, present at the level of the viral population in the conserved internal region C2, was sufficient to totally abolish Env functionality when introduced in an exogenous genetic context. The prominent defect of the non-functional protein is a block occurring after recognition of the co-receptor CCR5, likely due to an interference with the subsequent conformational changes that lead to membrane fusion. We also report that the presence of compensatory polymorphisms at the level of the external and hypervariable region V3 fully restored the functionality of the protein. The functional revertant presents different antigenic profiles and sensitivity to the entry inhibitor TAK 779.

**Conclusions:** Our data suggest that variable regions, besides harbouring intrinsic extensive antigenic diversity, can also contribute to sequence diversification in more structurally constrained parts of the gp120 by buffering the deleterious effect of polymorphisms, further increasing the genetic flexibility of the protein and the antigenic repertoire of the viral population.

**Keywords:** HIV, Envelope, Coevolution, Viral entry, Antigenic variation

## Background

Entry of the human immunodeficiency virus type 1 (HIV-1) into the target cell begins with the recognition by the viral envelope (Env) of the receptor CD4, a molecule of the immunoglobulin superfamily [1]. This step is followed by conformational changes of the Env protein [2] that generate its co-receptor binding site, in most cases the member of the G protein-coupled chemokine receptors CCR5, with a tendency to shift to the use of CXCR4

during the late phases of the acquired immunodeficiency syndrome [3, 4]. This binding triggers further structural rearrangements, leading to a conformation competent for promoting fusion of the viral and cellular membranes [5]. Progressive stepwise unmasking of different epitopes that occurs during the process implies that most of the conformations adopted by the protein are created only once the target cell has been recognised and, thereby, in a confined environment where the immune system has limited access. Despite this, recognition of Env by the immune system is manifestly the crucial immunological parameter for the control of viral replication, Env being

\*Correspondence: m.negroni@ibmc-cnrs.unistra.fr

<sup>1</sup> Architecture et Réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC, 15 rue René Descartes, 67084 Strasbourg, France

Full list of author information is available at the end of the article

the only target of neutralizing immune response documented to date [6–9].

The extensive genetic diversity of HIV-1 Env therefore constitutes a main asset for viral escape to immune response. However, genetic diversity in a protein that must undergo such extensive conformational changes constitutes an intrinsic weakness since each genetic polymorphism must be simultaneously compatible with all the various architectural arrangements. Indeed, relentless sequence variation due to misincorporation occurring during reverse transcription [10], to hypermutagenesis by cellular restriction factors [11], and to pervasive recombination occurring throughout the HIV genome [12–14] constantly challenges retention of functionality. And these requirements restrict the sequence space the protein can explore, limiting its antigenic diversity [15, 16]. Extensive coevolution of structurally and functionally related parts of the protein is the solution for the virus to conciliate genetic diversity and maintain of functionality.

In particular, deciphering coevolution networks in the gp120 is important at least from three standpoints. One is to provide insights into the dynamics of protein evolution in a model organism that undergoes physiological extensive sequence diversity. Second, it is important for understanding the tridimensional arrangement of the protein. A number of structures of HIV-1 Env has been indeed obtained under different conformations, mostly pre-CD4 bound [17, 18] and CD4 bound [19]. The remarkable structural flexibility of the protein, though, requires the use either of ligands (as antibodies) or of mutated or truncated versions of the protein to get a complex sufficiently rigid as to allow structural studies [5, 17–19]. In particular, mutations of the variable regions of the protein are frequently employed [20, 21]. All these modifications of the native protein can alter the natural conformation of the protein or stabilize it in conformations that might not be fully representative of those most commonly assumed by the free protein [22]. Furthermore, the arrangements of the protein during the recognition of the coreceptor and the subsequent conformational rearrangements triggered by this interaction still remains to be understood. For this step of the entry process, the only structure available to date is a soluble CD4-gp120 monomer bound to the tyrosine-sulfated antibody 412d that mimics CCR5 binding. This structure has allowed docking the N-terminal portion of CCR5 at the base of the V3 loop [23]. The second extracellular portion of the coreceptor, instead, has been mapped, by mutational analysis, to interact with the apical loop of V3 [24]. A structure of the Env trimer complexed to CD4 and CCR5, which would define the architecture of a prefusion intermediate, is however not yet available.

Finally, understanding coevolution networks can guide the design of antigenic modifications while retaining its functionality, an important issue for vaccine purposes. Properly folded antigens are indeed essential for eliciting neutralizing antibodies and the presence on VLP used for vaccines of a high proportion of non-correctly folded trimers appears as an important factor limiting the efficiency of vaccine approaches [25, 26].

As mentioned above, recombination is pervasive during HIV-1 replication and it is a process particularly challenging for the preservation of functionality in the presence of extensive coevolution networks. Indeed, recombination leads, in a single replication cycle, to the generation of multiple polymorphisms, often requiring the rapid introduction of several compensatory mutations to retain functionality. Accordingly, previous studies demonstrated the fragility of various HIV-1 genes to functional inactivation by recombination [16, 27, 28]. Starting from these observations, we recently probed the functionality of Env after swapping the variable regions V1V2 that, together with V3, are the most external portions of the protein in the pre-CD4 bound state. We observed that substituting these hypervariable domains from one primary isolate with that of another isolate was responsible for a dramatic drop in functionality in the chimerical protein. This indicates that also these parts of the protein, which are located in peripheral positions and harbour extensive genetic diversity, are submitted to coevolution constraints [29]. These results underlined the poor robustness of HIV-1 Env despite its well-known sequence diversity.

In the present work, through the initial generation of chimerical proteins, as a model of recombination, we focus on the involvement in coevolution of the constant region C2, a part of the protein central to several steps of the process of viral entry. This region not only is crucial because it constitutes the link between the two main domains of the gp120 (inner and outer domain) and contributes to the formation of essential structures as the bridging sheet, but also because it is located in a central portion of the Env core. Sequence diversification in this region can therefore have dramatic consequences on the global arrangement of the protein. Furthermore, it has been shown recently that C2 participates in inducing neutralising antibodies as 8ANC195 [30–32]. We report here, that genetic variation in a variable region of the protein such as V3 can relieve a dramatic decrease in functionality induced by a single sequence polymorphism in a constant region as C2. This underlines the importance of flexible regions of the gp120 in extending the sequence space that can be explored by this antigen, central for neutralisation of HIV-1.

## Results

### Replacement of the constant region C2 abolishes Env functionality

To study the involvement of the constant region C2 in coevolution with other portions of HIV-1 Env, we have generated four chimeric envelopes starting from primary isolates of different subtypes of HIV-1 group M. The isolates used were chosen on the basis of previous works where we observed a loss of functionality for recombinant and chimeric Env generated starting from these envelopes [28, 29]. In an Env protein from either an isolate of subtype A or B (called “Env A” and “Env B”, respectively) we have replaced the original C2 region with the C2 from an isolate of either group C or group G (whose envelopes will be referred as “Env C” and “Env G”, respectively). The resulting chimeras have been called A C2C and A C2G, for the Env A protein that contains the C2 region either from Env C, or from Env G, respectively (Fig. 1a, b). Similarly, Env B proteins that have been modified to contain the C2 region from Env C or from Env G, are named B C2C and B C2G, respectively (in order to avoid confusion we will always refer to amino acids using the three-letter code and single capital letters will always refer to HIV-1 M subtypes). The rationale for the choice of these specific chimeras was that, since we were searching for a possible perturbation of coevolution networks that would translate into a decrease in functionality, such a decrease would have been more manifest if the insertion of the exogenous sequence was made using as backbones the most functional isolates, in our case isolates A and B [28, 29].

The functionality of the chimerical envelopes has been tested by viral entry assays, as described in “Methods” section. For this test, pNL4-3.Luc.E<sup>-</sup> virions complemented by the individual chimeric envelopes to test have been used. All four chimeras displayed a dramatic drop in functionality with a residual functionality below 1 % in three cases and around 10 % for the chimera A C2C (Fig. 1c). Therefore, amino acids present in the C2 regions of isolates C and G, which were inserted in the A or B envelopes, dramatically perturb the functionality of both proteins.

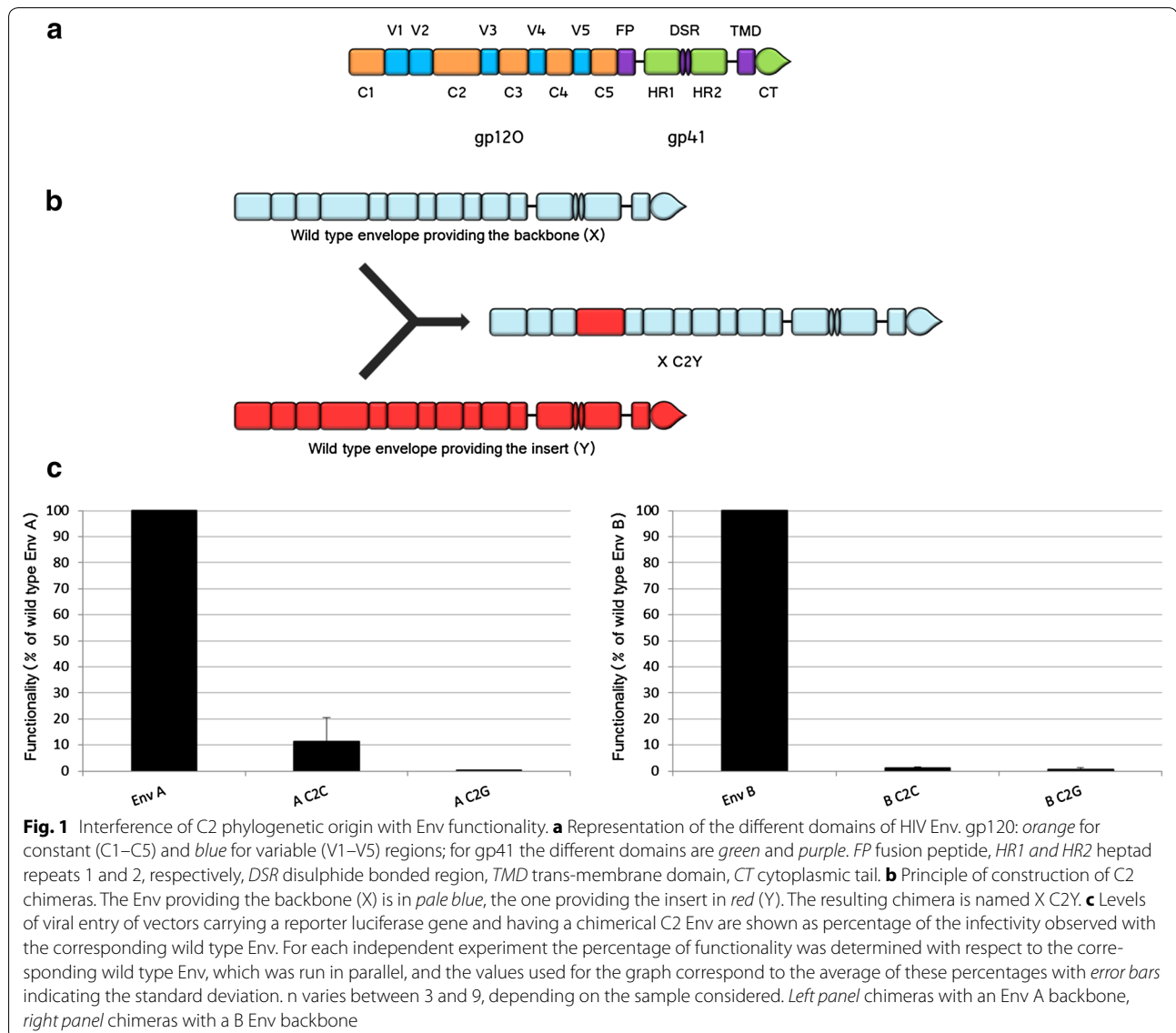
In order to identify the determinants for the drop in functionality caused by the insertion of the exogenous C2, we aligned the C2 region from the four isolates used. Sequence variability was circumscribed to three areas that we named R1, R2 and R3, for region 1, 2, and 3, respectively (Fig. 2a). Clustering of sequence diversity in these three regions is not limited to our isolates but well recaptures the pattern of sequence divergence at the population level, as shown by Shannon entropy

analysis along Env performed on isolates issued from the database for each of the four subtypes used in the study (Fig. 2b–e). This underscores the pertinence at the level of the viral population of the isolates employed in the present study.

### Identification of the region responsible for the loss of functionality in the C2 chimeras

The loss of functionality observed in the chimeras is necessarily due to the replacement of amino acids in the C2 region present in the original A or B sequences by those present in isolates C or G. Since the differences among these isolates were confined to regions R1, R2, and R3, the amino acids that cause the drop in functionality of the chimeras must be located either in one of these regions or in a combination of them. To assess their relative contribution to the drop in functionality observed, we focused on the A C2G and B C2G chimeras (since the insertion of the C2 sequence from isolate G triggered a stronger drop in functionality than that of C2 from isolate C, Fig. 1c) and we investigated at which extent inserting R1, R2, or R3 from isolate A or B would restore the functionality of the chimera (Fig. 3, panels a and b, respectively). Reinsertion of the original sequence R1 restored functionality at around 13 % for Env A (Fig. 3a) and 70 % for Env B (Fig. 3b). Both increases are significant ( $p < 0.05$  and  $p < 0.001$ , respectively), indicating that these regions contain important determinants for the loss of functionality observed in the A C2G and B C2G chimeras. Replacement of R2, instead, did not lead to detectable recovery neither for the A nor for the B backbones ( $p$  values, 0.384 and 0.152, respectively). Finally, the replacement of R3 led to a modest recovery (5 % of functionality,  $p < 0.05$ ) for isolate A while no effect was observed for isolate B ( $p = 0.488$ ).

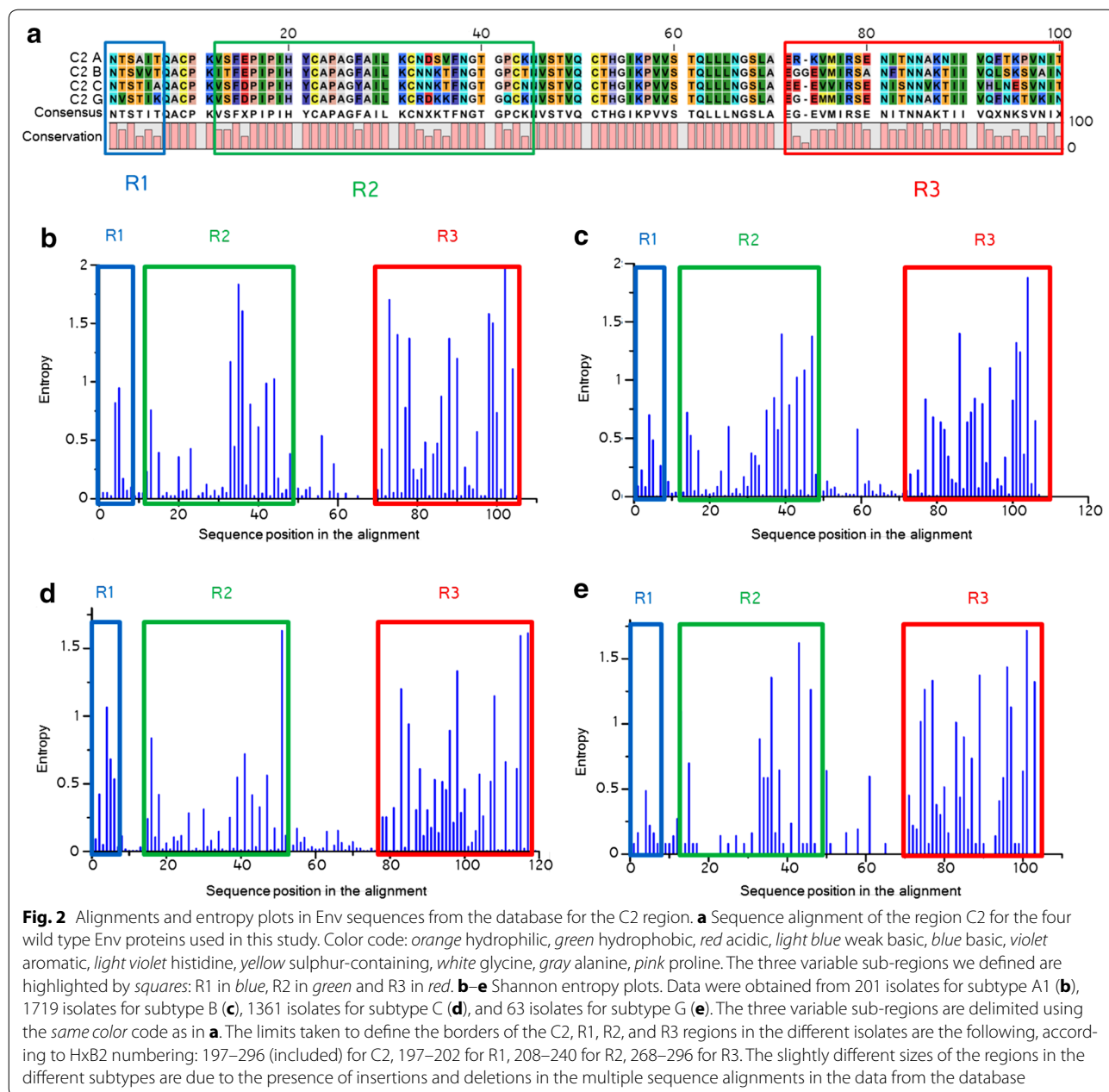
Altogether, the clearest picture emerges from the results obtained with the B Env for which the R1 sequence alone restores the functionality of the originally defective B C2G chimera to 70 % of the functionality of wt B Env. This indirectly suggests that this region harbours most determinants for the drop of functionality of B C2G with respect to wt B Env. To confirm this interpretation, we generated a wt Env B where we replaced uniquely the R1 region by that of the G Env (B R1G chimera, Fig. 3c). The presence of the exogenous R1 sequence alone was sufficient to drop functionality to levels comparable to those obtained for the replacement of the whole C2 region (Fig. 3c). This indicates that the sequence polymorphisms differentiating B and G envelopes in R1 are sufficient to account for an almost complete drop in functionality of the B Env.



### The polymorphism at position 202 induces the loss of functionality

Despite the dramatic difference at the level of functionality, wt B and B R1G Env differ only for the four amino acids shown in Fig. 4a. To evaluate their specific involvement, we checked whether the replacement in B R1G Env of each individual amino acid by those present in the B Env would restore functionality (Fig. 4b). The replacement from Lys to the original Thr present in wt B Env at position 202 was sufficient to fully restore functionality of B R1G bringing it to levels no longer significantly different from those of wt B Env ( $p = 0.584$ ). Replacement at the other three positions, instead, did not significantly alter functionality with respect to the starting chimera ( $p$  values: 0.328, 0.766 and 0.082 for positions 198, 200, and 201, respectively). At the level of the viral population, two

of the four positions that vary among the four subtypes considered in R1 (positions 200 and 201) are poorly conserved, particularly in subtypes A and C (Fig. 4c), reflecting genetic flexibility of these sites. This can account for the lack of restoration of functionality observed when replacing residues in these two positions. The other two positions (198 and 202), in contrast, presented a high subtype-specific conservation, with subtype G deviating from the other subtypes (Val in place of Thr for position 198 and Lys instead of Thr for position 202). This suggests that these amino acids could have coevolved with other subtype-specific residues elsewhere in the protein. Concerning residue 202, the restoration of functionality observed by inserting in this position the amino acid of the same phylogenetic origin than the regions other than C2 suggests that its coevolving partner is located outside



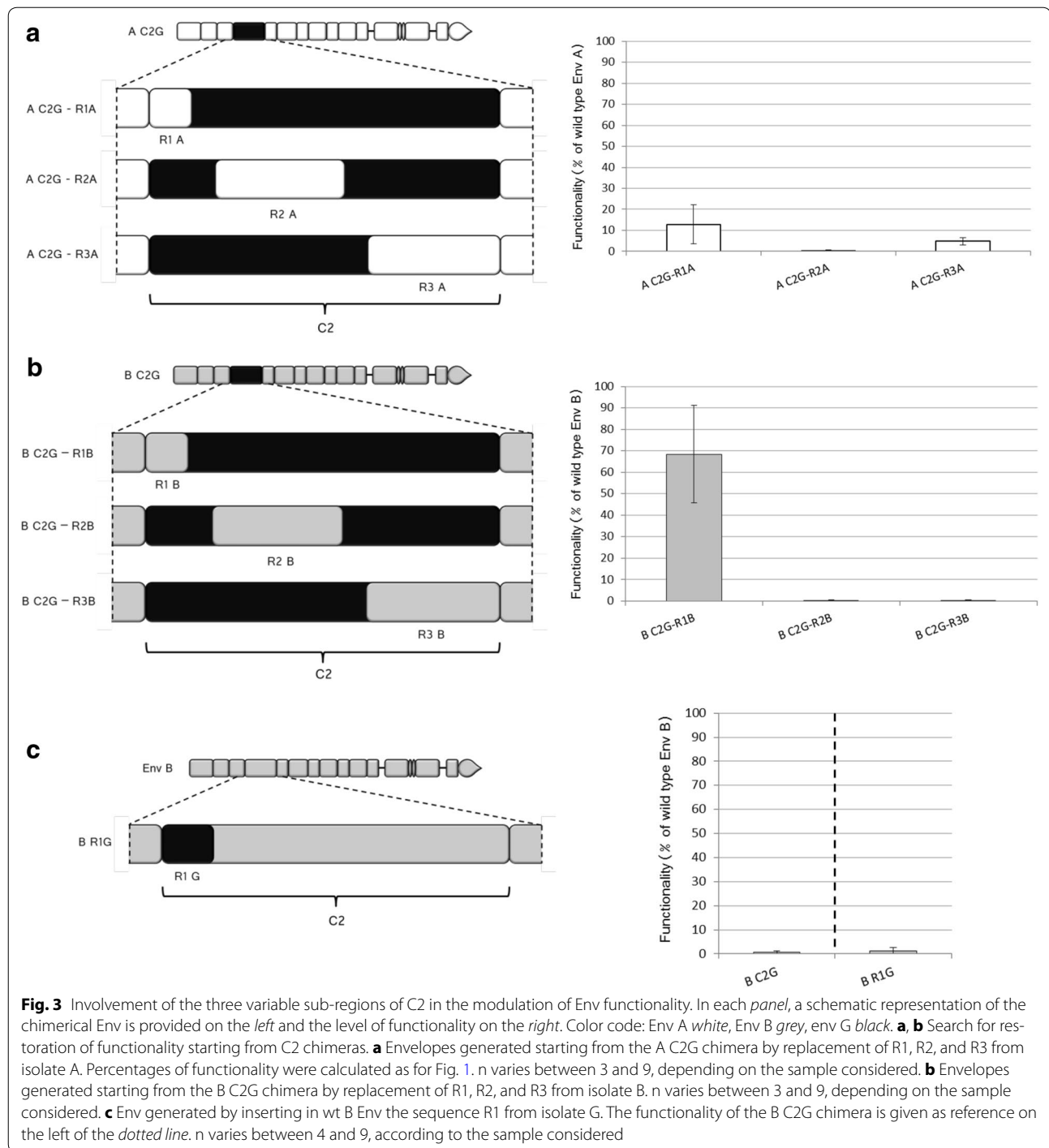
C2. In contrast, the fact that replacement at position 198 by the amino acid of the same origin than the regions outside C2 failed to restore functionality suggests that its coevolving partner might instead be located within C2.

To confirm the importance of position 202, we then made the “mirror” replacement by substituting, in wt B Env, uniquely the amino acid at position 202 by the one present in isolate G (Lys instead of Thr), generating B 202G Env. The functionality of this Env was nearly undetectable (Fig. 4d) confirming the central role of this polymorphism for modulating Env functionality. Since position 202 of isolate G is incompatible with the rest of

the Env B backbone it would also be expected that the reversal situation (the presence at position 202 of the Thr from Env B instead of the Lys of Env G) would perturb the functionality of Env G. Indeed, a significant loss of functionality ( $p < 0.01$ ) was observed also in this case (Fig. 4d), although significantly less severe ( $p < 0.05$ ) than in the case residue 202 from Env G inserted in Env B.

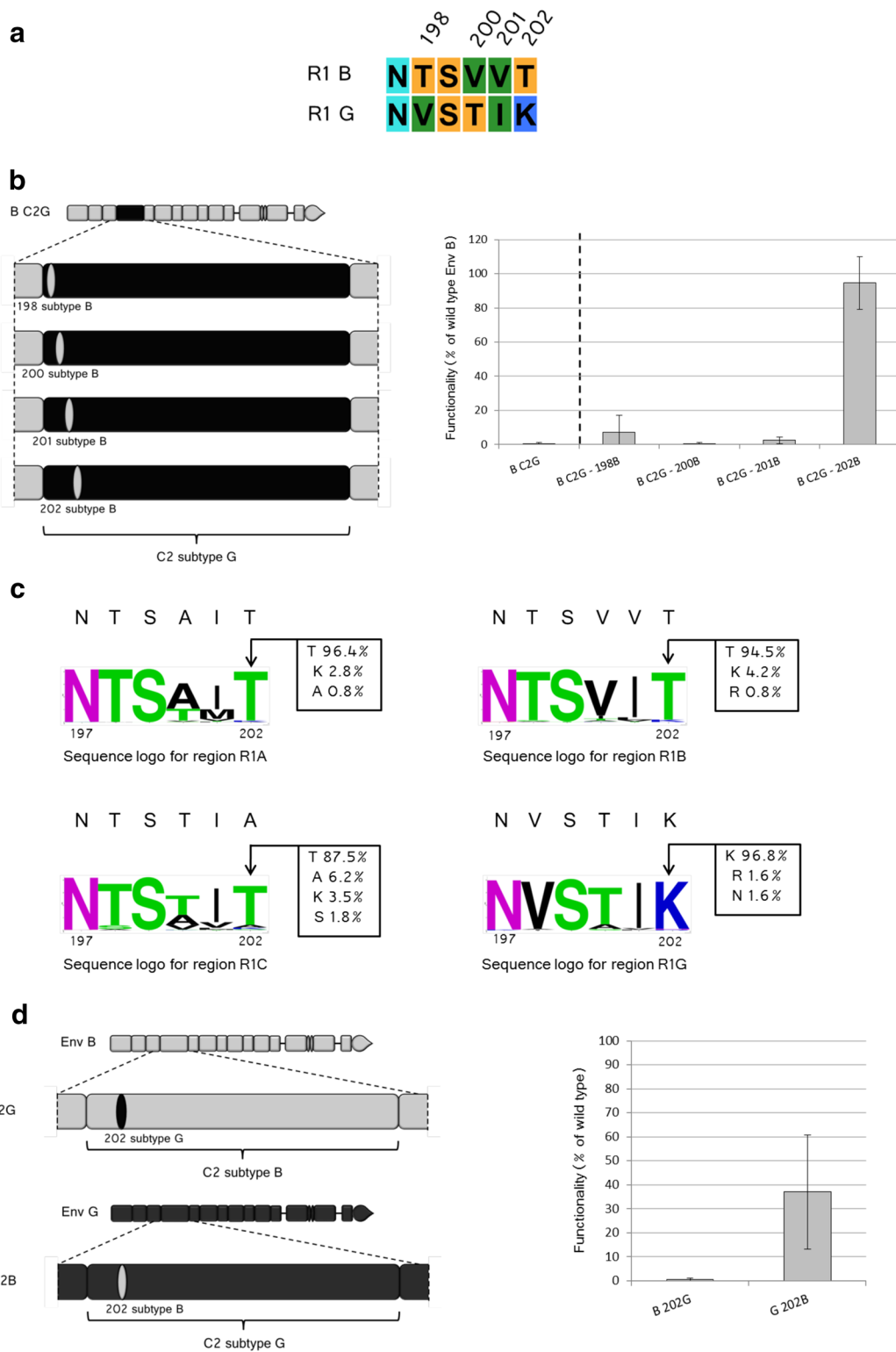
#### The coevolving partners of residue 202: identification of the region

Given the dramatic impact of the polymorphism at position 202 on the functionality of Env, we focused on this



position to characterize the involvement of C2 in coevolution networks through the search of amino acids that can regulate the functionality of Env through covariation with position 202. For this, we focused primarily on the case where the effect was the most marked: Env B containing the polymorphism of Env G at position 202. We therefore systematically replaced in that envelope various

regions of isolate B by the corresponding regions of the G isolate, as indicated in Fig. 5a. The rationale was that if amino acids that had coevolved with position 202 were harboured in the region that was replaced functionality should have been restored. With the exception of V3, the replacement of all the other regions did not affect the functionality that remained at nearly undetectable levels.





(See figure on previous page.)

**Fig. 4** Identification of the amino acid responsible for the loss of functionality in R1 chimeras. **a** Sequence alignment between regions R1 of the Env B and G showing the four amino acids differing between the two sequences: positions 198, 200, 201 and 202 (numbering according to HxB2). Color code is the same as for Fig. 2a. **b** *Left* Schematic representation of the B C2G chimeras carrying the substitutions of the individual amino acids indicated. Color code: Env B grey, Env G black. *Right* Functionality of the Env depicted on the *left*. The functionality of B C2G is given as reference on the *left* of the dotted line. *n* varies between 4 and 9, according to the sample considered. **c** Sequence logo (WebLogo, <http://weblogo.berkeley.edu/logo.cgi>) for R1 sub-regions of different subtypes are shown, subtype A at *top left*, subtype B *top right*, subtype C *bottom left* and subtype G *bottom right*; the relative prevalence at the position equivalent to position 202 is indicated. The sequence of the isolates used in our study is given on the *top* of each sequence logo. Color code: purple weakly basic residues, green polar residues, black hydrophobic residues, blue basic residues. **d** Schematic representation (*left*) and level of functionality (*right*) of Env B 202G and of Env G 202B. Percentages of functionality were calculated as for Fig. 1. *n* varies between 6 and 9, depending on the sample considered

Replacing V3, instead, restored significantly ( $p < 0.01$ ) the functionality of the protein to approximately 65 % that of wt B Env (Env B 202G-V3G, Fig. 5a). Therefore, although other portions than V3 probably participate to the network as suggested by the incomplete recovery of functionality of B 202G-V3G Env, V3 is largely the main component.

In order to investigate whether also in the case of G 202B Env functionality could be restored by insertion of a V3 region concordant with residue 202, we generated G 202B-V3B, which resulted to be non-functional (Fig. 5b). This result could be due to a lack of restoration of coevolution networks but it could also be explained if the replacement of the V3 region of Env G with the region V3B was not tolerated per se. To address this issue, we constructed the G V3B chimera (Fig. 5c). As shown in the figure, this chimera was not functional supporting the second hypothesis. As a control we then constructed the B V3G chimera that was fully functional (Fig. 5c). These results suggest why restoration of functionality could be observed in the case of B 202G-V3G but not for G 202B-V3B. Worth being noted, B V3G displayed a functionality of 140 % with respect to wt B Env (Fig. 5c), so the increase in functionality of B 202G-V3G with respect to B 202G could be partly due to this effect. However the amplitude of the effect on B 202G (from undetectable levels to more than 60 %) appears non-compatible with such an explanation. Finally, also data presented in the next chapter do not support this hypothesis.

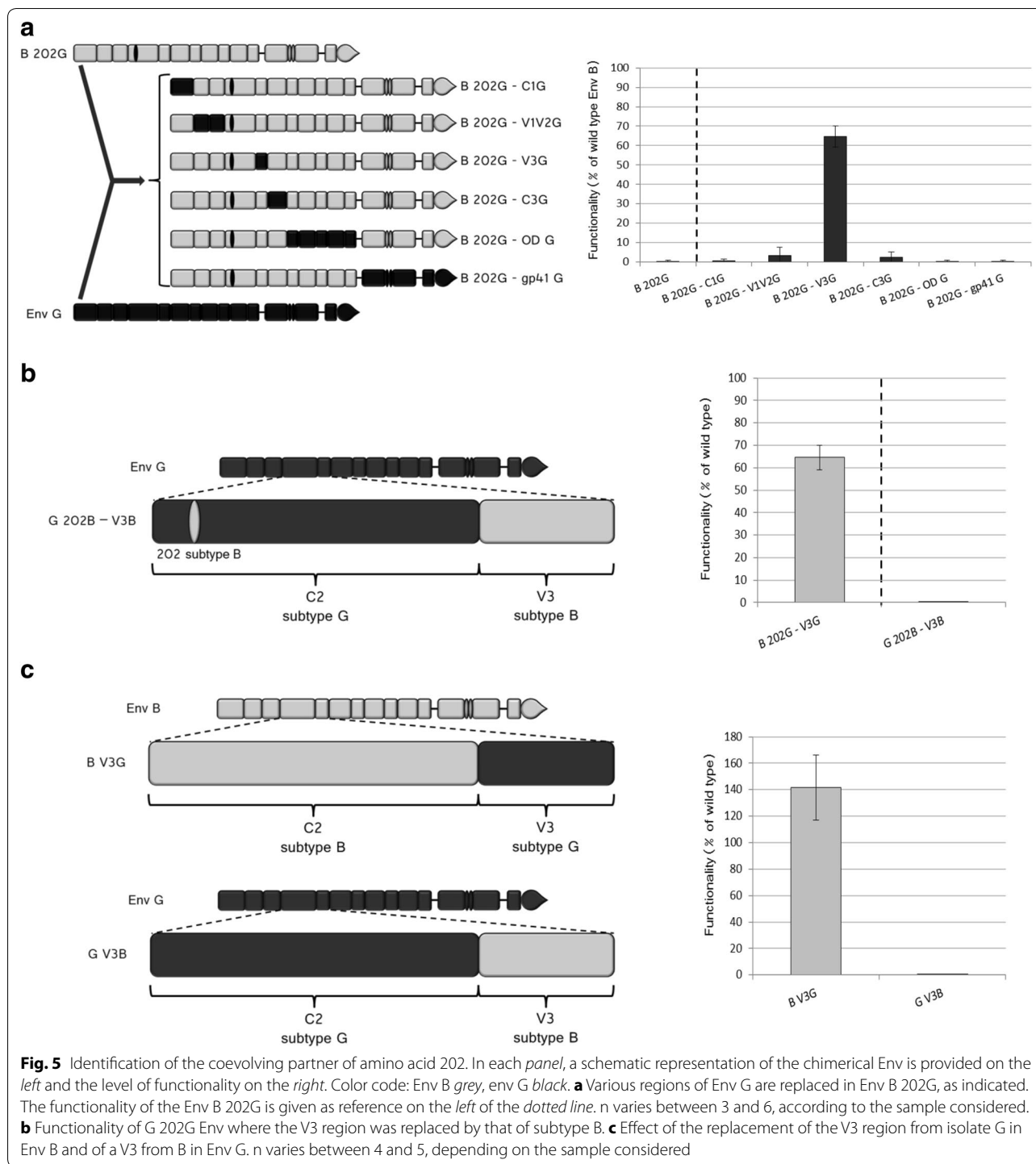
#### The coevolving partners of residue 202: identification of the residues

The need in V3 for amino acids specifically from the G isolate in order to observe the reversion to functionality was confirmed by the replacement in B 202G of V3 by V3 from isolate C (B 202G-V3C Env, Fig. 6a) that displayed a functionality only of 18 % with respect to wt B Env. The partial recovery of functionality with respect to B 202G (18 vs <0.1 %) is likely due to the smaller number of amino acids diverging between isolates G and C than between G and B in V3 (Fig. 6b). Also in this case,

we verified the tolerance of the insertion of the exogenous V3 region (in this case V3C) in wt B Env (chimera B V3C). As it was the case for B V3G, an increased functionality with respect to wt B Env was observed (130 % in this case, Fig. 6b). If the improvement of functionality in B 202G-V3G was due to the beneficial insertion of the V3 region from G, it would be expected to observe a similar increase also for B 202G-V3C. Since in this latter case the effect was instead modest (18 % of functionality with respect to 60 %) we conclude that the effect is rather due to the degree of sequence divergence between the different V3 regions.

The lower sequence divergence between G and C in V3 was then exploited for the identification of the specific amino acids of V3 from isolate G that allow functional preservation of the Env that carries the 202G polymorphism. Indeed, the difference between V3C and V3G is limited to five positions, in contrast to the ten amino acids that differentiate G and B in V3. The identification of the specific amino acids would have been extremely difficult starting from B 202G Env, particularly considering the possibility that multiple residues might be involved simultaneously, making the number of combinations to test unrealistic. For this reason we searched the amino acids involved in the recover of functionality in B 202G-V3G starting from B 202-V3C Env. This was achieved by replacing, in B 202-V3C, individually each one of the five amino acids differing from B 202G-V3G (Fig. 7a).

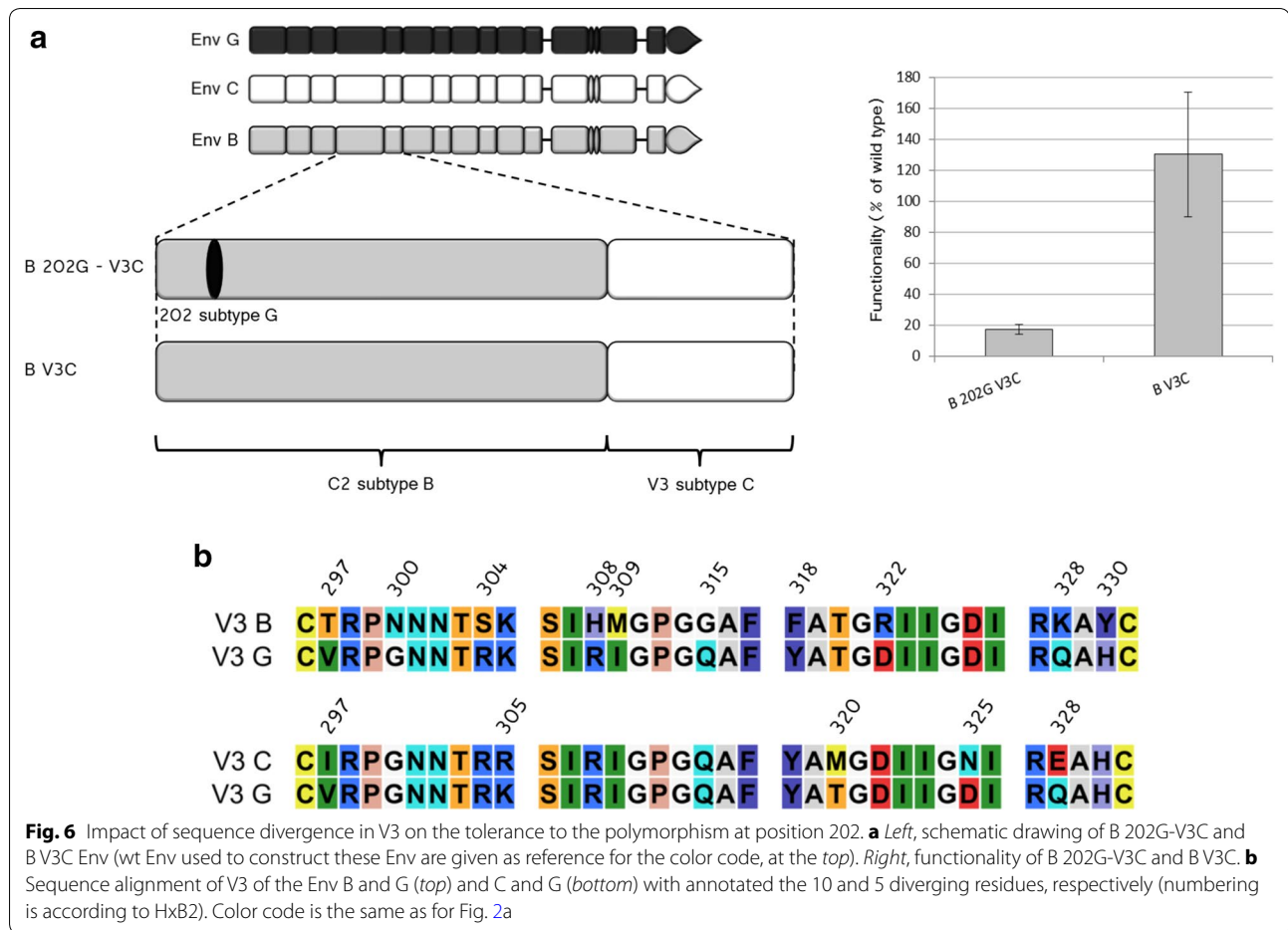
As shown in Fig. 7b, the individual replacements gave contrasting results. A complete loss of functionality was observed for the replacement of the methionine at position 320 by a threonine and of an asparagine by an aspartic acid at position 325, suggesting the participation of Met 320 and of Asn 325 in coevolution networks involving other residues within V3. A significant increase in functionality, indicative of coevolution between V3 and position 202, was instead observed for the replacement of residues 305 and 328 (lysine instead of arginine and glutamine instead of aspartic acid, respectively). Functionality remained however lower than that of the original wt



B Env. When the two mutations were combined (construct 6), though, the functionality of the protein was completely restored (Fig. 7b, no significant difference with respect to wt B Env,  $p = 0.24$ ). This result identifies a network of residues (positions 202–305–328) that through covariation can conciliate the presence of a Lys

at position 202 and a level of functionality as that of the wt Env.

Since substantial differences exist in V3 between the sequences of isolates B and C as shown in Fig. 6b, it was likely that the presence of a Lys at position 305 and a Gln at position 328, sufficient to restore functionality in



B 202G-V3C, would not restore functionality in B 202G. In addition, the possibility that these two positions were not crucial for Env B was also supported by the observation that the same amino acid is found in isolates B and G at position 305 (a lysine). To verify this, we replaced the Lys (originally present in Env B at position 328) by the Gln of isolate G. No restoration of functionality was observed for B 202/328G (functionality lower than 0.01 % that of wt B,  $n = 3$ ) in this case (Fig. 7c). It appears therefore that, depending on the V3 sequence, the residues allowing the restoration of functionality are not the same, reflecting functionally significant differences in the structural arrangement of this region.

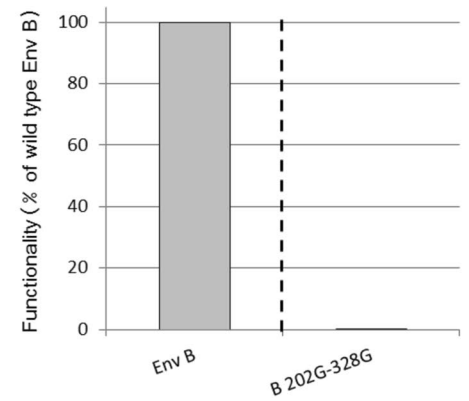
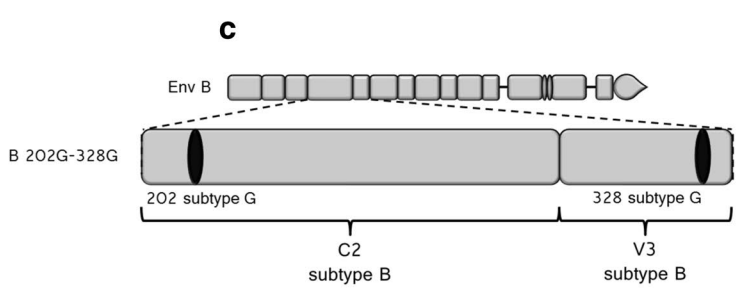
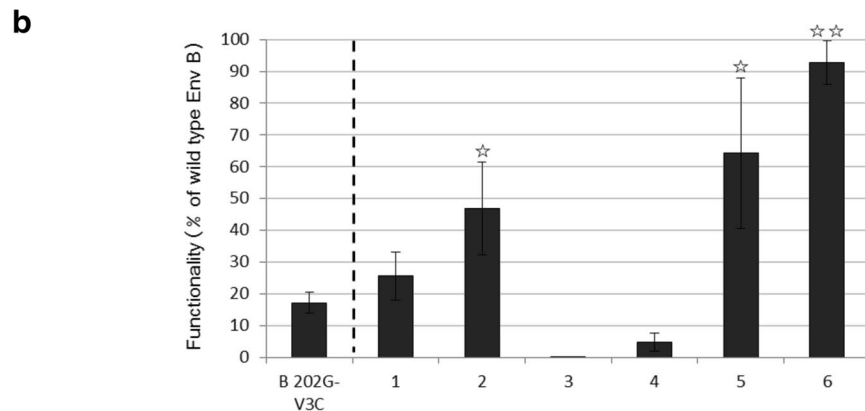
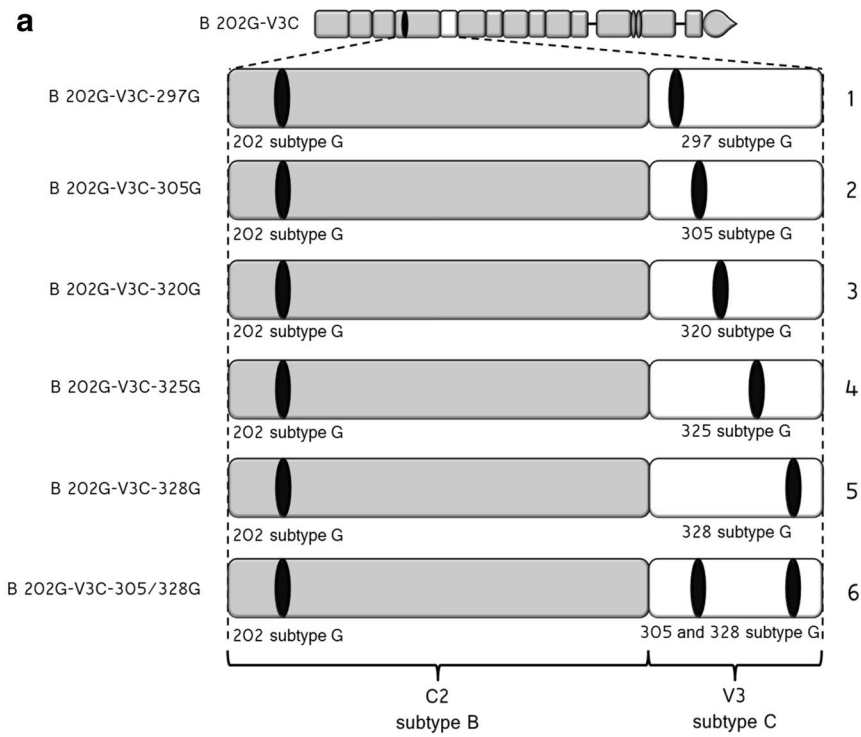
#### The 202 polymorphism does not affect gp120 stability on the viral particle

In order to identify at which level the perturbation of the coevolution network induced by the polymorphism at position 202 results in the abolishment of the functionality observed in B 202G we first verified the presence of the envelope in the viral particles, following a double approach: an ELISA test and a western blot analysis. For the ELISA test (also see “Methods” section) the plaque was coated with the antibody D7324, a polyclonal antibody that recognises an epitope spanning the gp120 and the gp41 subunit [33], a region that should be recognised in a similar manner in our chimeras since it is

(See figure on next page.)

**Fig. 7** Identification of the G-specific amino acids required to restore the functionality in B 202G-V3C Env. **a** Schematic representation of the Env constructed by inserting in the V3C sequence, amino acids from the G isolate, according to the scheme. Color code: Env B grey, Env G black.

**b** Functionality of the Env drawn in **a** (numbering of the samples is according to the numbering provided on the right of the drawings, in **a**). The value of infectivity for B 202G-V3C Env is given as reference on the left of the dotted line and the stars indicate that the functionality of the various Env increases significantly with respect to this reference. One star:  $p < 0.05$ ; two stars,  $p < 0.01$ . Percentages of functionality were calculated as for Fig. 1.  $n$  varies between 3 and 5, depending on the sample considered. **c** Schematic drawing (left) and functionality (right) of B 202G/328G (the functionality of B 202G is given as reference on the left of the dotted line)



located away from the portions of proteins that differ between wt B Env and Env B 202G. D7324 was used to capture viruses from the supernatant of cells transfected to produce viral particles exactly under the same conditions as for the viral particles used for the entry tests described above. After washing the plates, the viral titer was then evaluated by measuring the amount of p24 antigen. As shown in Fig. 8a, similar amounts of protein were detected for the two samples. The same antibody was then used for the western blot analysis and, confirming the results of the ELISA test, it indicated the presence of similar amounts of Env in the two samples (Fig. 8b). Defects in the recruitment of B 202G on the viral particle are therefore not responsible for the loss of infectivity observed with B 202G Env.

Env trimer formation is essentially driven by the formation of the six-helix bundle by gp41. In our case, the gp41 component of the different Env is constant, suggesting an equivalent formation of trimers at the surface of the viral particle. However, maintain of a native, pre-CD4 bound, conformation is influenced by the nature of gp120, and this has an impact on the stability of the architecture of the whole trimer. To infer the proportion of Env in a native trimeric form formed by wt B and B 202G we made an ELISA test as the one described above, using the monoclonal antibody PGT145 that recognises a quaternary epitope present only in the native trimeric form of Env [34] This antibody was chosen because, as shown below (Table 1), it led to a similar neutralisation profile for all the chimeras tested, suggesting a similar efficiency of recognition of the various envelopes and therefore also of wt B and B 202G. This antibody appeared therefore the most promising for detecting the presence of native trimers at the surface of the viral particles. As indicated in Fig. 8c, native trimers were present for both Env tested. The approximately three times lower amount of p24 retained on the plaque for B 202G than for the wt B Env could reflect a lower proportion of native trimers for B 202G, or a worse recognition of this envelope with respect to wt B Env, or both. The detection of native trimers for B 202G indicates that this envelope should be able to lead to membrane fusion even if, in the case of a higher proportion of misfolded Env formation for B 202G, this could partly account for the decrease in functionality of this Env.

We then investigated whether the stability of the interaction between gp120 and gp41 was perturbed in B 202G. Decreased stability of this interaction would indeed lead to dissociation of the gp120 from the gp41 in the absence of CD4 making the viral particle unable to recognize and bind the target cell. To evaluate the stability of the gp120 at the surface of the viral particle in the absence of interaction with CD4 we incubated purified virions

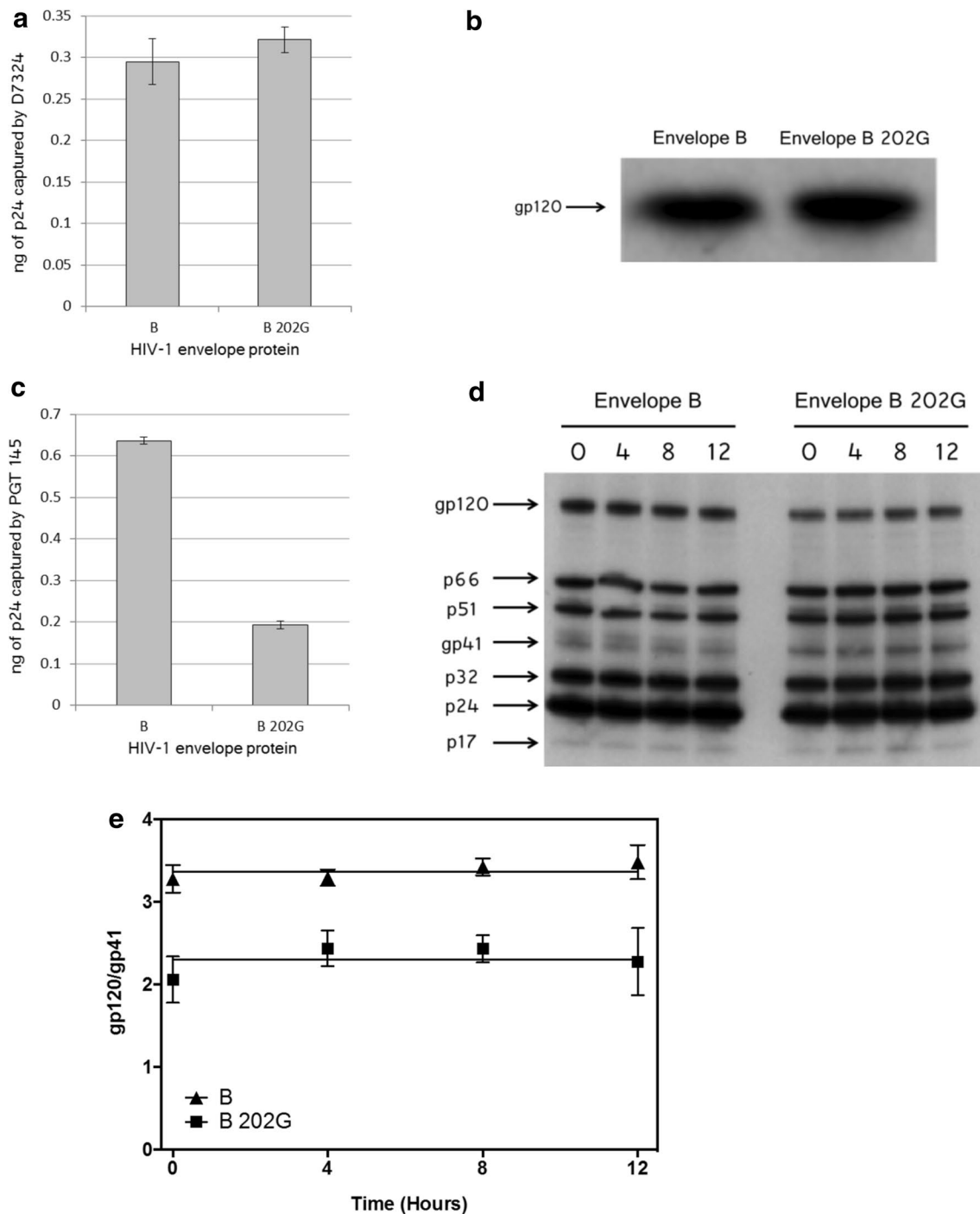
carrying either wt B Env or B 202G in PBS at 37 °C for different times (from 0 to 12 h) before ultracentrifugation and measuring the ratio gp120/gp41 by western blot on the pellet fraction (Fig. 8d). The rationale was that a decreased stability of the interaction between gp120 and gp41 would lead to a progressive shedding of the gp120 from the surface of the viral particle that would result in a curve with a negative slope when the gp120/gp41 ratio is plotted as a function of increasing times of incubation, as done in Fig. 8e. The flat trend of both curves, instead, indicates that both gp120 variants were stably present on the viral particles indicating that the polymorphism at position 202 does not impact the stability of the gp120 at the surface of the viral particle.

The two envelopes differed instead for the height of the curves, in Fig. 8e, which is indicative of an apparent higher gp120/gp41 ratio for wt B Env. The theoretical ratio gp120/gp41 in the viral particle is 1 and the observed ratios most likely reflect how well each of the two proteins is recognised by the serum used for the western blot. Indeed, the serum from patient better recognised gp120, which is known to be highly immunogenic, than gp41 and for both Env. The difference in the heights of the curves in Fig. 8e can be explained by the possibility that the replacement of the amino acid 202 alone affect the antigenic profile of the envelope as discussed below.

#### **The affinity of monomeric B 202G gp120 for CD4 and CCR5 is similar to that of wt B gp120**

The ability of B 202G to bind CD4 and CCR5 was then tested by a biochemical approach. Binding of purified soluble B 202G gp120 to CD4 exposed at the surface of HEK 293T cells was measured by competition with the monoclonal antibody Q4120. This antibody engages the same site at the surface of CD4 that is recognized by HIV gp120 [22] and thus sterically inhibits gp120 binding to the receptor. The equilibrium dissociation constant  $K_i$  value for B 202G binding to CD4 was deduced by using increasing quantities of soluble B 202G gp120 in competition with a constant quantity of Q4120 (0.5 nM). The experiment was performed in parallel with wt B gp120. Both gp120 displayed similar affinities for CD4 ( $K_i$  of 5.0 and 3.9 nM for B 202G and wt B gp120, respectively, Fig. 9a). This indicates that differences in CD4 binding efficiencies cannot account for the observed dramatic differences observed in viral entry with these two envelopes.

The affinity for CCR5 has then been measured by saturation binding experiments of  $^{35}\text{S}$ -labeled gp120 to purified CCR5-expressing HEK cell membranes in the presence of an excess concentration of soluble CD4 (200 nM), as previously described [35] (see “Methods” section). Specific binding of both gp120 to CCR5 could



**Fig. 8** Properties of the defective B 202G Env at the surface of the viral particles. **a** Viral titer (ng of p24) estimated after retention of viral supernatant on ELISA plates coated with the antibody D7324. **b** Western blot on purified viral particles using antibody D7324. **c** Viral titer (ng of p24) estimated after retention of viral supernatant on ELISA plates coated with the antibody PGT 145. **d, e** Measure of the stability of the gp120 protein at the surface of the viral particle. **d** Western blot performed on purified virions carrying either wild type B or B 202G Env proteins as function of the incubation time (indicated, in hours, above each lane) at 37 °C. The identity of the different proteins has been inferred by their molecular mass and is indicated with arrows. The gel shows a representative result of three independent experiments. **e** The bands corresponding to gp120 and to gp41 in three western blots (as the one shown in **d**) have been quantified as indicated in "Methods" section and the corresponding values are expressed as gp120/gp41 ratio as function of the incubation time. Triangles wt B Env; squares B 202G Env

**Table 1 Sensitivity to antibodies and inhibitors of viral entry**

	N332-V3			V2				CD4 binding site		gp120-gp41 interface		Entry inhibitors		
	IC50 ug/mL			IC50 ug/mL				IC50 ug/mL		IC50 ug/mL		IC50 ug/mL		IC50 nM
	PGT121	PGT128	10-1074	PG9	PG16	PG9-PG16-RSH	PGT145	3BNC117	VRC03	35022	8ANC195	sCD4	T20	Tak779
wt B	>10	>10	>10	4.09	2.02	0.96	0.03	3.20	>10	9.29	9.74	0.45	0.12	31.19
wt G	1.06	0.15	0.15	5.16	>10	8.48	0.10	0.12	2.08	>10	2.71	1.11	0.02	14.84
B V3G	>10	>10	>10	1.2	0.08	0.11	0.03	0.56	0.71	>10	>10	0.1	0.05	< 0.3
B 202G-V3G	>10	>10	>10	0.11	0.07	0.09	0.01	0.39	1.21	>10	>10	0.3	0.12	3.84
B 202G-V3C-305/328G	>10	>10	>10	0.18	0.04	0.1	0.02	0.34	1.04	>10	>10	0.77	0.1	2.08

wt B	>10	>10	>10	4.09	2.02	0.96	0.03	3.20	>10	9.29	9.74	0.45	0.12	31.19
wt C	9.17	>10	8.47	0.17	0.1	0.06	9.93	>10	>10	8.53	9.57	4.42	0.14	6.12
B V3C	>10	>10	>10	0.77	0.18	0.11	0.05	0.81	1.31	>10	>10	0.42	0.05	< 0.3
B 202G-V3C	>10	>10	>10	0.31	0.04	0.11	0.04	0.46	2.26	>10	>10	0.31	0.08	1.83
B 202G-V3C-305/328G	>10	>10	>10	0.18	0.04	0.1	0.02	0.34	1.04	>10	>10	0.77	0.1	2.08

Cases are coloured according to the level of  $IC_{50}$ : <0.1, red; 0.1–0.99, orange; 1.0–9.99, yellow; >10.0, green. The table is divided in two groups: above, chimeras involving V3G; below, chimeras involving V3C. The values for wt B Env and B 202G-V3C-305/328G are reported as reference in both parts of the table

be described by hyperbolic curves (Fig. 9b) and from which we derived  $K_d$  values of  $8.1 \pm 1.7$  and  $8.4 \pm 1.5$  nM for B 202G and wt B Env, respectively. These results rule out also CCR5 binding efficiencies from the list of possible defects in functionality of B 202G Env.

#### B 202G efficiently binds target cells but cannot lead to membrane fusion

To evaluate B 202G binding to CD4 in the context of the trimer expressed at the surface of the viral particle, virions encoding the firefly luciferase were added to HEK 293T CD4<sup>+</sup> CCR5<sup>+</sup> cells in culture and incubated at 37 °C for different times, as indicated in Fig. 9c, d. Cells were then detached mechanically by gentle pipetting and centrifuged at low speed, conditions under which free virions are found in the supernatant fraction while internalised or cell-bound virions are associated to the cell pellet (also see “Methods” section). The supernatant was discarded and the cell pellet was gently resuspended and divided into two aliquots. One was used to evaluate viral entry by measuring the expression of the luciferase (Fig. 9c), while the other was used to measure the amount of viral particles associated to the cells (those internalised plus those simply bound to them) through estimating the quantity of p24 associated to the cells (Fig. 9d).

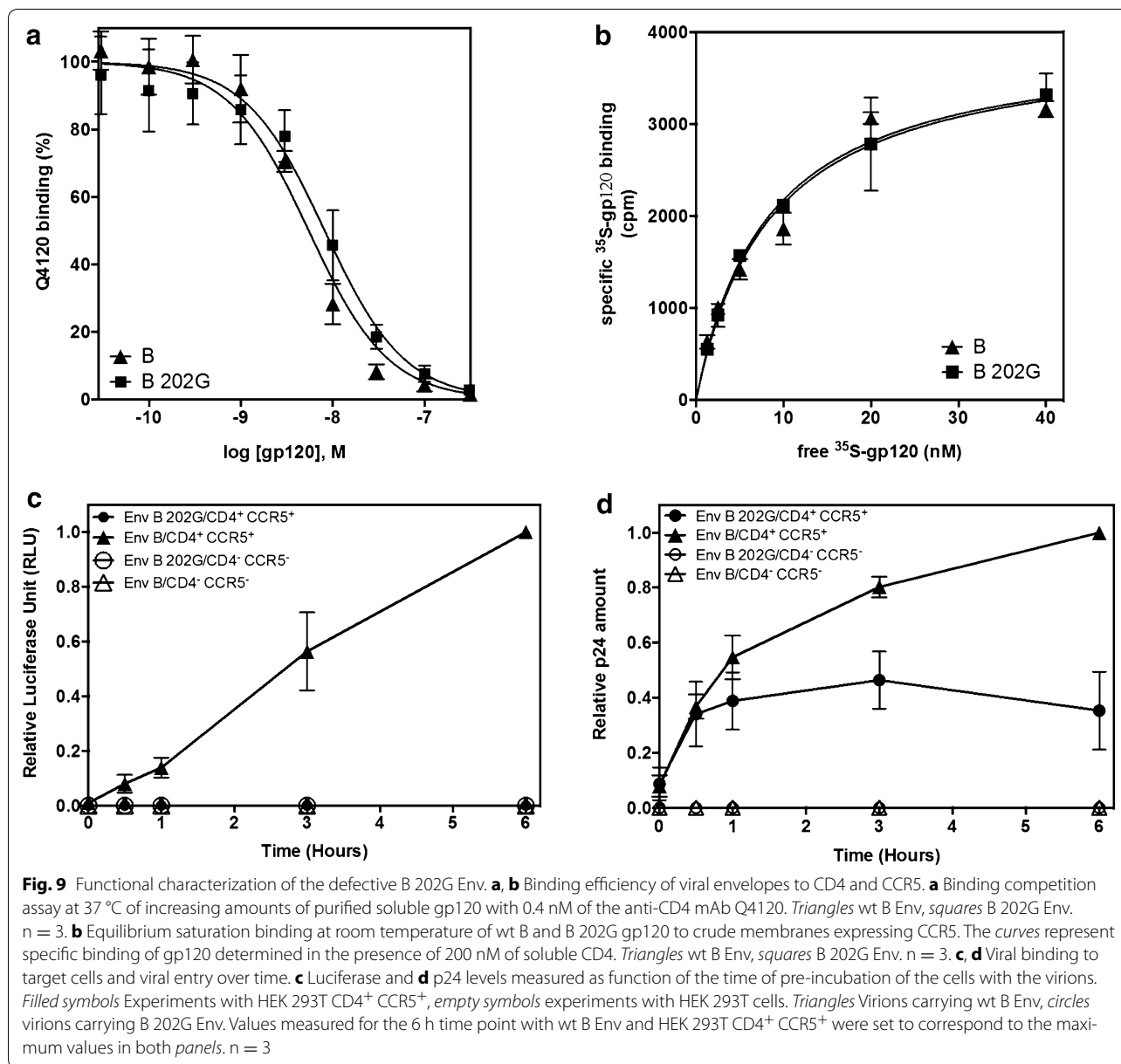
While a progressive increase of viral entry in the target cell was observed with wt B Env for increasing times of incubation of the cells with the viruses, no entry was observed for B 202G Env irrespective of the incubation time, confirming the total lack of functionality of this Env (Fig. 8a). However, despite its lack of functionality

B 202G retained the ability to bind the target cells, as indicated by the amount of p24 associated to the cells (Fig. 8b). No signal of p24 was observed for both Env on the same cells deprived of CD4 and CCR5, indicating the specific nature of the binding signal observed.

This result confirms, in the context of the trimeric Env presented at the surface of the viral particle, the observations made for the binding of the soluble gp120 to CD4, further supporting the conclusion that B 202G possess an ability to bind the receptor comparable to that of the original wt B protein. Binding of B 202G rapidly reaches a plateau, likely as a consequence of non-productive binding of this Env. For wt B Env, the time course curve expectedly paralleled the luciferase curve shown in panel A, reflecting progressive viral entry for increasing times of incubation of the cells with the viruses.

#### Revertant Env has an altered antigenic profile and a different sensitivity to the entry inhibitor Tak779

In order to probe the structural arrangement of the wt Env B and of the B 202-V3C-305/328G revertant we tested their immuno- and pharmaco-properties (Table 1). Differences in the recognition by different antibodies or in the efficiency of specific inhibitors of viral entry would be indicative of differences in the conformation of the various envelopes tested. Neutralisation sensitivity to antibodies directed against V3, V2, the gp120/gp41 interface and the CD4 binding sites (CD4bs) as well as sensitivity to the entry inhibitors soluble CD4 (sCD4), T20 and Tak779) was tested. Concerning the antibodies directed against V3 a surprising observation is that the



antigenic properties of V3 from isolates G or C are not retained when these V3 are inserted in the Env B backbone, rather displaying the properties of wt B Env.

This is true not only for the revertant, but also for all the other chimeras tested where V3 was exogenous (B V3G, B V3C, B 202G-V3G, B 202G-V3C) indicating that the conformation of the V3 domain is modulated by the sequence of the rest of the protein. Concerning the antibodies directed against V2 and against the CD4bs, the insertion of the exogenous V3 triggers a change in the parameters of recognition by these two classes of antibodies, in all cases leading to an increased sensitivity

with respect to the starting Env (wt B Env). This is indicative of the influence of V3 on the conformation of V2 and of the accessibility of the CD4 binding site. Noteworthy, recognition of V2 by PG9 is different between B 202G-V3G and BV3G, indicating a modulation of the conformation of V2 by residue 202, either direct or through V3. In the case of PGT145, instead no increased sensitivity was observed and all Env were neutralised with similar efficiency justifying its use for the ELISA tests aiming at the estimation of the amount of trimeric Env mentioned above. Expectedly, no significant alteration of the recognition pattern was observed among the various Env for



the antibodies targeting the gp120/gp41 interface, confirming that this interface was not altered in the variants of Env considered in this work. This result is in accord with the absence of major differences in the sensitivity to the entry inhibitor T20. For the other two entry inhibitors tested, the similarity of sensitivity to sCD4 suggests that all Env undergo similar transitional dynamics between different opening states of the Env trimers at the surface of the viral particle [36] and is indicative of similar affinities of the various Env for CD4. Finally, sensitivity to Tak779, an allosteric antagonist of CCR5, highlights a two-log decrease from wt B Env and B V3G, documenting the occurrence of a conformational change induced by the exogenous V3. Similarly the ten-fold difference observed with Tak779 between B V3G and B 202G-V3G shows that the introduction of the 202G polymorphism also induces a marked conformational change of V3. Overall the results obtained with Tak779 in combination with the biochemical data indicate that the mode of binding to CCR5 differs between wt B Env and the revertant B 202G-V3C-305/328G and that the return to functionality is paralleled by conformational changes in V3.

## Discussion

Coevolution is essential for allowing sequence diversification of proteins, a process particularly important for pathogens that are exposed to the immune response of the host. We report that, despite its overall conservation across HIV-1 clades, replacing the constant region C2 of HIV-1 Env of a primary isolate of subtype A or B by that of an isolate from subtype G or from subtype C triggered a dramatic decrease in functionality with respect to the wild type proteins (Fig. 1c). We focused in particular on the chimeric subtype B isolate where the C2 sequence from isolate G was inserted, since the effect was the most manifest in that case. For this chimera we identified a single polymorphism to be responsible for the total loss of functionality observed when the whole C2 region was exchanged. In particular, replacement of the Thr at position 202 (conserved in subtype B at 96.4 %) by a Lys (conserved at 96.8 % in subtype G) resulted to be sufficient to drop Env functionality to background levels. Functionality of Env B mutated at position 202 was then restored by the insertion of the V3 region from the G isolate, indicating that the coevolving partners of residue 202 were located in V3. In particular, we could identify two residues (positions 305 and 328) that, when of G origin, were sufficient to fully restore functionality (Fig. 7b). Amino acids 305 and 328 are located in the top portion of the ascending part (aa 305) and at the base (aa 328) of the V3 loop, respectively. It has been postulated by bioinformatics approaches the existence of a coevolution link within V3 between residue 328 and positions 306

and 307 [37]. Even if in our case positions 306 and 307 are not involved, it is possible that covariation between position 328 and amino acid 305, adjacent to 306 and 307, is required to provide an appropriate folding of V3. This would “link” positions 328 and 305, requiring their simultaneous replacement in the revertant.

The existence of epistatic interactions between V3 and other portions of the protein had been suggested previously, specifically in the context of the determination of coreceptor use. Indeed, even if the choice of the coreceptor is mostly dictated by interactions with V3, mutations present in the gp41, but also in C1, C2 and C4 alter this choice [38–42]. With respect to those observations that report subtle modifications of the activity of the Env, we report here for the first time that the epistatic interaction between V3 and C2 modulates the transition from completely non-functional to fully functional forms, without impacting coreceptor usage.

The restoration of a marked defect due to a polymorphism in region C2 by covariation of residues in the V3 loop connects functionally two structural elements of the protein that accumulate mutations at different paces. The lower sequence diversity present in constant regions, despite their tendency to induce higher mutation rates than variable regions [43], is intuitively attributable to more strict structural constraints that result in stronger purifying selection in these regions. By documenting the existence of coevolution between constant and variable regions, the present work shows that functional constraints acting on constant regions can influence sequence diversity in the variable ones. Conversely, the structural flexibility of the V3 region increases the genetic robustness of the protein by allowing to compensate for the negative impact of polymorphisms introduced in constant regions and, consequently, to explore a larger sequence space. Disordered regions have been previously reported to increase the sequence plasticity of their respective proteins in a series of organisms [44–46]. Although variable regions of Env cannot be considered as disordered domains and a defined architectural arrangement has been assessed for V1V2 and V3 [47, 48], their genetic diversity underscores a higher flexibility than that of constrained constant portions of the protein. The results we have obtained on V3, suggest that variable regions of Env, could play a function similar to that of disordered parts of proteins, and their presence could increase genetic robustness, besides the well-recognised role they have in the escape from the immune response.

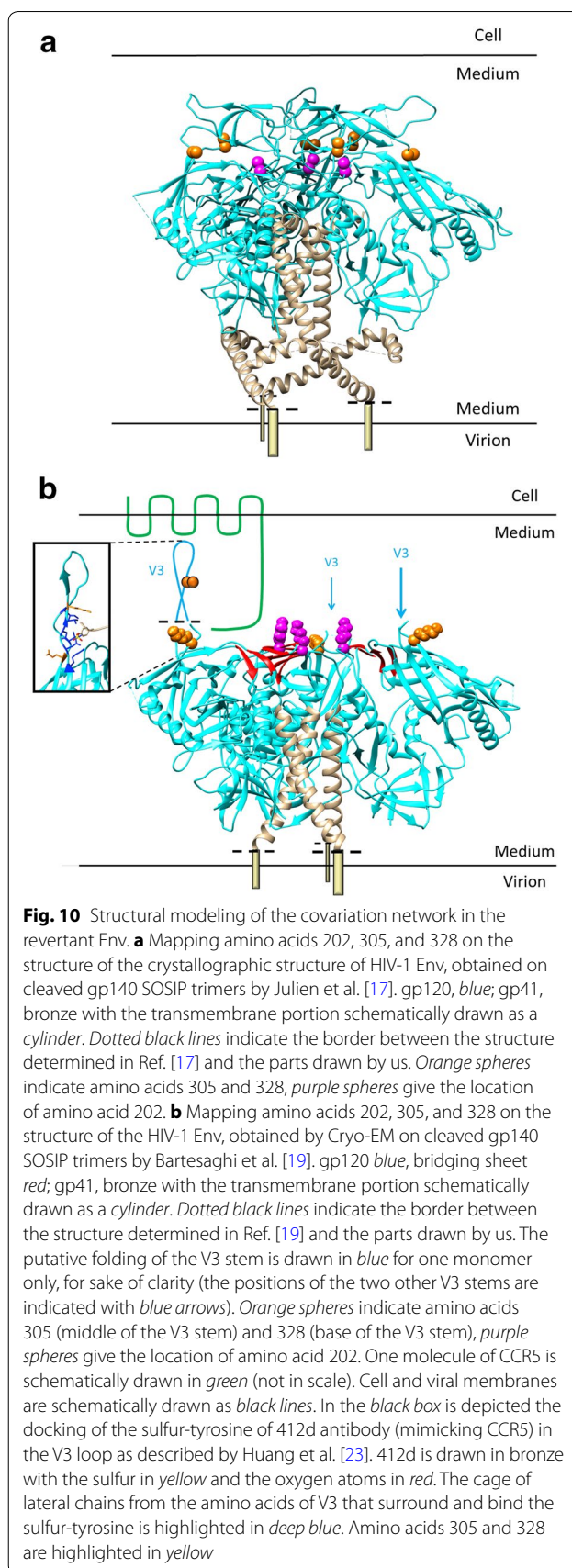
The non-functional Env (B 202G) was neither affected in the stability of the interaction between gp120 and gp41 as judged by western blot on the purified viral particles (Fig. 8d, e), nor in the ability to bind CD4, or CCR5, as shown by biochemical and cell culture data (Fig. 9a, b). Since PGT145 recognizes less efficiently B 202G than

wt B Env (Fig. 8c) it is possible that the mutant yields a lower proportion of native trimers at the surface of the viral particles. This can therefore account for part of the decreased efficiency of entry by B 202G, since a correlation between number of trimeric spikes present and efficiency of viral entry has been described [49, 50]. The total absence of functionality observed, though, is indicative of the existence of an additional, strong block in the entry process.

A second clear defect triggered by the 202 polymorphism was observed downstream co-receptor binding. Biochemical tests indicate, in fact, that B 202G binds CCR5 with an affinity identical to wt B Env despite a total inability to carry out membrane fusion. Since the gp41 portion of B 202G is the same as in wt B Env, failure to carry out membrane fusion by gp41 per se can be ruled out. The defect rather seems to come from the conformational changes that lead from the CCR5-bound structure to the membrane-fusion competent one. The participation of residue 202 in the formation of the bridging sheet, which is involved in the interaction with CCR5, supports this possibility. This interpretation is also compatible with our biochemical data and with the time-course experiments of viral entry into target cells and capture of the viruses at their surface (Fig. 9) that show that, albeit B 202G does not allow viral entry, it retains the ability to bind target cells.

In order to evaluate the potential impact of the polymorphisms at positions 202, 305 and 328 on the conformation of the native trimers, the positions of these residues was mapped on the structure of the SOSIP Env trimer from Julien et al. [17]. In this structure, these residues do not appear to interact directly (Fig. 10a), supporting the hypothesis that the amino acid 202 coevolves with the conformation of the V3 loop as a whole, rather than specifically with residues 305 and 328, potentially explaining the absence of restoration observed for the chimera B 202/328G. Noteworthy, if coevolution occurs with the global folding of a large domain rather than with single specific amino acids, the panel of solutions to retain coevolution networks is expected to be broader and more flexible. Another important element emerging from this analysis is that residue 202 is located just beneath V1V2, which have been described to be involved in the stabilisation of the native conformation of trimers through their interaction with V3 [51]. A mutation occurring in 202 could therefore impact on the correct folding and on the stability of the trimer.

For structural considerations about the post CCR5 binding defect, the appropriate structure to refer to would be a trimeric HIV-1 Env in complex with CD4 and of CCR5. Such structure not being available, we mapped amino acids 202, 305 and 328 on the CD4-bound trimer structure [19]. In the CD4-bound state the V3 loops project away



from the protein core [2, 20, 21, 52] and, consequently, a spatial proximity between residue 202 and its coevolving partners, 305 and 328 does not exist even in this structure. However, the independence of these three amino acids is only apparent here, since these positions could be bridged by a bound CCR5 molecule, as schematically indicated in Fig. 10b. Indeed, CCR5 interacts with the top and the base of the V3 loop as well as the bridging sheet and the three amino acids we describe are located in these regions. The observation that the non-functional B 202G mutant retains an unaltered ability to bind CCR5 (Fig. 9b), can be explained on the basis of the structure of the soluble CD4-gp120 monomer bound to the tyrosine-sulfated antibody 412d that mimics CCR5 binding [23]. This structure indeed shows the existence of a binding pocket for the sulphur-tyrosine of CCR5 in V3. Amino acids 305 and 328 do not participate to the formation of this binding pocket but precisely encompass it (Fig. 10b) explaining why polymorphisms in these positions do not alter the efficiency of CCR5 binding. The most plausible explanation for the loss of functionality of B 202G Env is that the arrangement of the complex Env-CD4-CCR5 formed specifically with the mutant interferes with the appropriate subsequent conformational changes that lead to membrane fusion [5]. How compensation of polymorphism 202 by the polymorphisms 305–328 occurs constitutes an interesting issue for understanding the conformational rearrangements occurring after CCR5 binding.

Finally, comparative immunological and pharmacological probing of wt B and revertant Env confirm the dramatic impact on the conformation of the whole gp120 induced by the simple polymorphisms 202. This is reflected, on one hand, by the tenfold change in the sensitivity to Tak779 observed between B V3G and B 202-V3G. On the other hand, also the antigenic properties of the protein appear strongly modulated. A major observation in this sense is constituted by the change of the recognition of antibodies directed against the V3 domain exerted by the backbone in which V3 is inserted. Similarly, immune recognition of V2 is modulated by the nature of V3. Overall, these observations underline the level of complexity of the requirements for an efficient recognition by the immune system, at least for these regions of Env. Finally, these analyses document the differences in the antigenic properties of the protein observed when transitioning from the wt B Env to the non-functional B 202G variant and, finally to the functional revertant triple mutant (B 202G-V3C-305/328G). The new functional variant, possessing a different antigenic profile broadens the repertoire of infectious variants the immune system has to deal with. Altogether, these observations shed light on a new level of complexity of the interplay between the virus and its host.

## Conclusion

Demonstrating the ability of a variable region of HIV-1 Env to rescue the deleterious effect of a genetic polymorphism in a more structurally constrained region, this work sheds light on a new aspect of the implication of variable regions of Env in the increase of genetic flexibility of this protein. This flexibility allows to generate alternative functional envelopes with new antigenic profiles, widening the repertoire of variants the immune system has to deal with.

## Methods

### Cells

HEK 293T, HEK 293T CD4<sup>+</sup>, HEK 293T CCR5<sup>+</sup> and HEK 293T CD4<sup>+</sup> CCR5<sup>+</sup> were cultured in Dulbecco's Modified Eagle Medium supplemented with 10 % foetal calf serum, penicillin (100 UI/mL) and streptomycin (100 µg/mL) (Invitrogen, Carlsbad, NM, USA) at 37 °C in 5 % CO<sub>2</sub>.

### Plasmids, viral sequences and alignments

The envelope coding sequences have been inserted into the commercial plasmid pcDNA 3.1D/V5-His-TOPO (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA) and were co-transfected with pNL4-3. Luc.E<sup>-</sup> plasmid [53] in order to produce viral particles. The envelope-coding sequences originated from primary isolates belonging to group M. In detail, the following isolates were used: one isolate belonging to subtype A1 (GenBank accession number AF407156, referred herein as Env A); one isolate belonging to subtype B (GenBank accession number AY835448, referred herein as Env B); one isolate belonging to subtype C (GenBank accession number DQ435683, referred herein as Env C); one isolate belonging to subtype G (G-548, from Ref. [29], an isolate cloned in the laboratory from a pool of cDNA kindly provided by Dr. Martine Peeters, issued from the study described in Njai et al. [54], referred herein as Env G). These four envelope genes were used to generate all the variants of envelope described in the result section. Sequences were aligned using Unipro Ugene (<http://ugene.net>) for generating sequence logos and for entropy analyses, and using CLC Sequence Viewer (<http://www.clcbio.com/products/clc-sequence->).

### Construction of chimerical envelopes

Chimerical envelopes between different primary isolates were constructed through overlapping PCR as previously described [16, 27, 28]. Briefly, each fragment, of a given phylogenetic origin, used to constitute the final chimeric genes was amplified independently. Primers used for these PCRs carry short complementary sequences in their extremities that can allow the different fragments to

hybridize. Individual fragments produced containing the overlapping sequences are then mixed in a subsequent PCR where primers are added only after few cycles. Full-length chimera is reconstituted and amplified by adding external primers.

A similar protocol has been used to generate point mutants except that the desired mutations were directly inserted in the desired sequence by amplification with a forward and a reverse complementary primer. The forward primer was used for PCR amplification with a reverse complementary primer annealing at the 3' end of the Env-coding sequence, while the reverse complementary primer was used for amplification with a forward primer annealing at the 5' end of the Env-coding sequence. These PCR products were then assembled to generate the full-length gene as described above for the chimerical envelopes. All constructs were verified by sequencing.

#### Production of HIV-1 particles and viral entry assay

The day before transfection HEK 293T cells were plated in 6 well plates, at a density of  $8.4 \times 10^5$  cells per well. HIV-1 viral particles carrying the envelopes to test and containing the luciferase gene have been produced by cotransfection of HEK 293T with pNL4-3.Luc.E- and pcDNA3.1D.Env plasmids using the polyethylenimine (PEI) method according to manufacturer's instructions (PEI MW 25,000, linear; Polysciences, Warrington, PA, USA). Seventy-two hours after transfection, the recovered medium was filtered and the viral particles were quantified by ELISA directed against p24 capsid protein according to the manufacturer's protocol (Innotest HIV Antigen mAb, Innogenetics, Gent, Belgium). The incorporation of equivalent amounts of the different mutants of envelope proteins in the viral particles was tested by western blot.

Transduction of HEK 293T CD4<sup>+</sup> CCR5<sup>+</sup> with viral particles was performed by adding 25 ng of p24 protein (corresponding to  $1.5 \times 10^8$  HIV virions) to  $1.25 \times 10^5$  cells in 24-well plates. Transduced cells were incubated at 37 °C and harvested 48 h in order to proceed to a luciferase assay by following the instructions (Luciferase Assay System, Promega, Madison, Wisconsin, USA). Luminous signal produced by luciferase have been quantified with a Glomax luminometer (Promega, Fitchburg, WI, USA). Negative controls were constituted by virions produced by cotransfection with pNL4-3.Luc.E- and pcDNA 3.1D/V5-His Topo plasmids not carrying any envelope sequence, while positive controls were constituted by virions carrying the highly functional T-ADA strains envelope.

#### Statistical tests

When a mean has been compared to a reference value, as for the comparison of the functionality of a chimera

with respect to the functionality of a wt reference protein (which was set at 100 %), a Student's one sample *t* test is used. When two means (for example of two chimeras) must be compared, a Fisher-Snedecor's F-test is first used to test the hypothesis of the equality of variances. If F-test rejects the hypothesis of the equality of variances, then a Welch's *t*-test is used to test the independency of the means, otherwise a Student's two samples *t*-test is used to test the independency of the means.

#### Shannon entropy calculations

Clustering of amino acid sequence variation in the C2 region of HIV-1 Env was evaluated by calculating Shannon entropy using the software provided by Los Alamos National Laboratory (LANL) ([http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy\\_one.html](http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one.html)). HIV-1 sequences used for this study have been downloaded from the LANL database. Sequences have been sorted by subtype (A1, B, C and G). For each subtype all available sequences were downloaded after having selected the option provided by the LANL website for random choices of only one sequences per patient. The regions coding for C2 have then been translated into protein that have been subsequently aligned separately and the alignments verified manually. In all cases, the alignments made by the software have been judged satisfying and no manual corrections have been applied. Alignments have finally been uploaded onto the LANL online software for the Shannon entropy test.

#### ELISA

Cells have been transfected as described in upper section in order to produce HIV-1 viral particles. Two days post-transfection, supernatants have been harvested, filtered, and virions have been quantified by ELISA directed against p24 according to the manufacturer's instructions (Innotest HIV Antigen mAb, Innogenetics, Gent, Belgium). Separately, 96 well-flat bottom microtiter plates (Greiner Bio-one GmbH, Frickenhausen, Germany) have been incubated overnight at 4 °C with 50 µL of capture antibody at 5 µg/mL in carbonate-bicarbonate solution (Sigma-Aldrich, St Louis, Missouri, USA). The capture antibodies were either sheep polyclonal antibodies D7324 (Aalto Bio Reagents, Dublin, Ireland), which are raised against the same continuous epitope APTKAKR-RVVQREKR and purified for their affinity toward this epitope, or human monoclonal antibodies PGT 145 (kind gift of Martine Braibant) [34]. After incubation, plates have been washed 5 times with PBS + Tween (0.5 %) and blocked with 200 µL of PBS supplemented with non-fat milk (7 %) for 4 h at room temperature. After a second washing step identical to the first one, virions produced by transfection have been added to each well (100 ng of

p24 for each well) and plates have been incubated for 2 h at 37 °C. Plates have been washed again and, to harvest the content of each well, 100 µL of RIPA have been added to each well. Plates have been incubated for 5 min at 4 °C allow a complete lysis. Content of each well have then been taken, diluted at 1/5 in PBS and an ELISA test directed against p24 has been performed according to the manufacturer's instruction (Innotest HIV Antigen mAb, Innogenetics, Gent, Belgium).

#### Western blot with primary antibody D7324

Supernatants containing viral proteins were collected for western blot analysis. Samples were then normalised in order to contain 30 ng of p24 viral proteins and were separated by electrophoresis in a 4–12 % NuPAGE Bis–Tris gel (NuPAGE Novex, Thermo Fisher Scientific, MA, USA). Proteins were transferred to an Immobilon-P polyvinylidene difluoride (PVDF) membrane (Merck-Millipore, Billerica, Massachusetts, USA) and incubated with antibody D7324. Secondary antibody recognition was performed by incubation with monoclonal anti-sheep donkey antibodies conjugated to the horseradish peroxidase (Sigma-Aldrich, St Louis, Missouri, USA). Immunoblots were revealed by a luminol based enhanced chemiluminescence substrat (Pierce ECL, Life Technologies, Thermo Fischer Scientific, Waltham, Massachusetts, USA).

#### gp120 and gp41 interaction assay

The evaluation of the stability of the interaction between gp120 and gp41 was assessed by western blot. Viral particles were produced by transfection as described above. Twenty-four hours after transfection, the supernatant of the culture was filtered and viral particles were purified by ultracentrifugation on 20 % sucrose cushion, as previously described [29]. Virions were then incubated at 37 °C in PBS for 0, 4, 8 and 12 h. For each time point, samples were collected by centrifugation at 100,000×g for 2 h. Supernatants containing free gp120 were discarded while the viral particles present in the pellet fraction were lysed in RIPA buffer (1× PBS, 1 % NP-40, 0.5 % sodium deoxycholate, 0.05 % SDS). Samples were then incubated for 15 min at 4 °C and centrifuged at 12,000×g for 20 min. Supernatants containing viral proteins were collected for western blot analysis by using the same protocol described in upper sections. The only difference is that primary antibodies that have been used consist in a pool of sera of 2 patients directed against subtypes B and CRF-02 viruses respectively (kind gift of Jean-Christophe Plantier, Rouen, France). Secondary antibody recognition was performed by incubation with monoclonal goat antibodies conjugated to the horseradish peroxidase (Sigma-Aldrich, St Louis, Missouri, USA). The amount of proteins was quantified using Image Lab software

(Bio-Rad, Hercules, California, USA) based on the intensity of the signal obtained for gp120 and gp41 relative to the p24 protein.

#### CD4 and CCR5 binding assay

Soluble gp120 was expressed and purified as previously described [35]. Increasing concentrations of soluble gp120 were bound to  $5 \times 10^4$  CD4-expressing HEK 293T cells in the presence of a 0.5 nM concentration of Q4120 [55]. Specific Q4120 binding was detected with Alexa Fluor 647-conjugated goat anti-mouse IgG and quantified by FACS analysis. Results were normalized for non-specific binding (0 %) and specific binding in the absence of glycoprotein (100 %) and were fitted to a one-site competitive binding model with the software GraphPad Prism (GraphPad Software, LaJolla, California, USA), as described in Ref. [35]. The equilibrium dissociation constants  $K_i$  for the gp120 s were calculated from the curves according to the Cheng and Prusoff equation  $K_i = [IC_{50}/(1 + L/K_D)]$ , where L and  $K_D$  represent the Q4120 concentration (0.5 nM) and the equilibrium dissociation constant of the Q4120-CD4 complex ( $K_D = 0.4$  nM), respectively. For measuring affinities of gp120-CCR5 interactions, equilibrium saturation binding experiments of  $^{35}S$ -labeled gp120 to crude membranes from CCR5-expressing HEK 293T cells were obtained as described in Ref. [35] in the presence of 200 nM of soluble CD4. Specific binding curves were obtained after subtracting from total binding the non-specific binding of glycoproteins measured in the presence of 10 µM maraviroc. Data were fitted to a one-site binding model with the software GraphPad Prism (GraphPad Software, LaJolla, California, USA).

#### Virion-cell binding assay

To evaluate the ability of the envelopes to bind to the target cells and to perform viral entry, virions carrying the envelope of interest (50 ng of p24) were used to infect  $5 \times 10^5$  HEK 293T CD4<sup>+</sup> CCR5<sup>+</sup>. Non-specific binding of viral particles to the cells was evaluated in two ways: measuring binding of virions carrying no envelope proteins to HEK 293T CD4<sup>+</sup> CCR5<sup>+</sup> and measuring binding of virions carrying the Env proteins to study to HEK 293T deprived of CD4 and CCR5. The signal obtained with virions carrying no envelope proteins and HEK 293T CD4<sup>+</sup> CCR5<sup>+</sup> was used to subtract background signal. Virions and cells were incubated at 37 °C for different times and the cells were then pelleted by centrifugation at 400g for 3 min. Supernatants containing free virions were discarded and the cell pellets gently resuspended in DMEM. Half of each sample was incubated in duplicate at 37 °C, 5 % CO<sub>2</sub> for 48 h and used for luciferase assay. The other half was centrifuged again 3 min

at 400 g, supernatants were discarded and cellular pellets were lysed with RIPA buffer using the same protocol described above. The amount of p24 protein in pellets was estimated by ELISA, as described above.

#### Env-pseudotyped virus production and titration for neutralization assays

Env-pseudotyped viruses were produced as previously described [56] by cotransfecting  $3 \times 10^6$  293T cells with 4  $\mu\text{g}$  of each pCDNA3.1-env clone and 8  $\mu\text{g}$  of pNL4.3.LUC.R\_E [53] using FuGene<sup>®</sup>-6 transfection reagent (Promega, Madison, Wisconsin, USA). Virus stocks were harvested 72 h later, purified by filtration (0.45  $\mu\text{m}$  filter) and stored as aliquots at  $-80^\circ\text{C}$ . Viral infectivity was monitored by infection of  $1 \times 10^4$  TZM-bl cells, with serial fivefold dilutions of viral supernatants in quadruplicate, in the presence of 30  $\mu\text{g}/\text{mL}$  DEAE-dextran. Infection levels were determined after 48 h by measuring the luciferase activity of cell lysates using the Bright-Glo luciferase assay (Promega, Madison, Wisconsin, USA) and a Centro LB 960 luminometer (Berthold Technologies, Bad Wildbad, Germany) [57]. Wells producing relative luminescence units (RLU) >2.5 times the background were scored as positive. The  $\text{TCID}_{50}$  was calculated as described previously [58].

#### Neutralization assay

Sensitivity to HuMoNABs of the pseudotyped viruses was assessed in duplicate in TZM-bl cells, as previously described. After titration, virus stocks were diluted to 400  $\text{TCID}_{50}/\mu\text{L}$ . Aliquots of 50  $\mu\text{L}$  were then incubated for 1 h at  $37^\circ\text{C}$  with 50  $\mu\text{L}$  of threefold serial dilutions of PGT121, PGT128, 10-1074, PGT145, PG9, PG16, PG9-PG16-RSH, VRC03, 3BNC117, 35O22 and 8ANC195 (10–0.0046  $\mu\text{g}/\text{mL}$ ). The virus-antibody mixture was then used to infect 10,000 TZM-bl cells in the presence of 30  $\mu\text{g}/\text{mL}$  DEAE-dextran. Infection levels were determined after 48 h by measuring the luciferase activities of cell lysates, as described above. Results were expressed as mean values.  $\text{IC}_{50}$  values were defined as the antibody concentration required to reduce RLUs by 50 %.

#### Inhibition of entry by enfuvirtide (T20), TAK779 and sCD4-183

TZM-bl cells were employed in quadruplicate to assess the sensitivity of pseudotyped viruses to ENF, TAK779 and sCD4-183 (NIH AIDS Research and Reference Reagent Program). After titration, pseudotyped virus stocks were diluted to obtain 400  $\text{TCID}_{50}/\text{mL}$  in growth medium. Aliquots of 50  $\mu\text{L}$  were then incubated for 1 h at  $37^\circ\text{C}$  with 50  $\mu\text{L}$  of threefold serial dilutions of ENF or sCD4-183 (10–0.0046  $\mu\text{g}/\text{mL}$ ). The virus-inhibitor mixture was then used to infect  $1 \times 10^4$  TZM-bl cells in

the presence of 30  $\mu\text{g}/\text{mL}$  DEAE-dextran. Infection levels were determined after 48 h by measuring the luciferase activities of cell lysates.  $\text{IC}_{50}$  values were defined as the reciprocal of the inhibitor concentration required to reduce RLUs by 50 %. Results were expressed as mean values. For TAK779 inhibition,  $8 \times 10^3$  TZM-bl cells per well were plated the day prior infection. Cells were first treated for 1 h at  $37^\circ\text{C}$  with 150  $\mu\text{L}$  of three-fold serial dilutions of TAK779 (6  $\mu\text{M}$ –0.3 nM) before adding 50  $\mu\text{L}$  of pseudotyped viruses normalized to 400  $\text{TCID}_{50}/\text{mL}$ . 100  $\mu\text{L}$  of DMEM medium supplemented with 30  $\mu\text{g}/\text{mL}$  DEAE-dextran were then added to cells. Luciferase activity was measured 48 h after infection as described above. CCR5 antagonist susceptibility was expressed as maximal percent inhibition (MPI) and  $\text{IC}_{50}$  values. Results were expressed as mean values.

#### Authors' contributions

RG conceived the study and the experimental strategy; carried out all the work described (with the exception of the neutralization tests), participated in writing the manuscript. MH begun the study and conceived its initial part. MP participated to the experimentation. ZZ and PC produced and purified gp120 monomers and contributed to investigation of their interaction with receptor and coreceptor in vitro. CV carried out the neutralization tests, which were conceived and directed by MB. BL directed the study of Env interaction with receptor and coreceptor in vitro. MN conceived, designed and directed the study, wrote the manuscript. All authors read and approved the final manuscript.

#### Author details

<sup>1</sup> Architecture et Réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC, 15 rue René Descartes, 67084 Strasbourg, France. <sup>2</sup> Present Address: U1016, UMR 8104, INSERM-CNRS, Institut Cochin, Université Paris Descartes, Paris, France. <sup>3</sup> INSERM U1108, Viral Pathogenesis Unit, Department of Virology, Institut Pasteur, Paris, France. <sup>4</sup> INSERM U966, Université François Rabelais, Tours, France.

#### Acknowledgements

The following reagents were obtained through the AIDS, NIAID, and NIH: A-Q461ENVe2, Q842ENVd12, and Q168ENVa2 from Dr. Julie Overbaugh [59]; pTHRO4156 clone 18 (SVPB15) from Drs. B.H. Hahn and D.L. Kothe [60]; CAP210.2.00.E8, SVPC17, and CAP45.2.00.G3, SVPC16 from Drs. L. Morris, K. Misana, and D. Montefiori [61]; ZM249M.PL1, SVPC10, and ZM197M.PB7, SVPC6 from Drs. B.H. Hahn, Y. Li, and J.F. Salazar-Gonzalez [61]. We thank Pascal Poinard and the International AIDS Vaccine Initiative, Inc. (IAVI) for providing us with PG9, PG16, PGT121, PGT128, PGT145; Marie Pancera, Peter Kwong and the Vaccine Research Center (VRC), NIAID, NIH for providing us with PG9-PG16-RSH; Michel C. Nussenzweig for providing us with 10-1074, 3BNC117 and 8ANC195. The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: pNL4.3.LUC.R-E- from N. Landau; TZM-bl cells from J. C. Kappes, X. Wu, and Tranzyme Inc.; TAK-779 from M. Baba; sCD4-183 from Progenics; VRC03 from Dr. John Mascola; anti-HIV-1 gp41/gp120 Monoclonal 35O22 from Drs. Jinghe Huang and Mark Connors; T-20, Fusion Inhibitor from Roche.

#### Competing interests

The authors declare that they have no competing interests.

#### Funding

This work was supported by grants from Sidaction, and the French Agency for AIDS and Hepatitis Research (ANRS) to M. Negroni. R. Gasser was supported by a grant from the French "Ministère de l'Enseignement Supérieur et de la Recherche" and the ANRS, M. Hamoudi by a grant from the French "Ministère de l'Enseignement Supérieur et de la Recherche" and Sidaction. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Received: 24 March 2016 Accepted: 18 July 2016

Published online: 30 July 2016

## References

- Dalglish AG, Beverley PCL, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The Cd4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*. 1984;312:763–7.
- Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, Subramaniam S. Molecular architecture of native HIV-1 gp120 trimers. *Nature*. 2008;455:109–13.
- Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun CX, Grayson T, Wang SY, Li H, et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA*. 2008;105:7552–7.
- Mild M, Kvist A, Esbjornsson J, Karlsson I, Fenyo EM, Medstrand P. Differences in molecular evolution between switch (R5 to R5X4/X4-tropic) and non-switch (R5-tropic only) HIV-1 populations during infection. *Infect Genet Evol*. 2010;10:356–64.
- Pancera M, Zhou TQ, Druz A, Georgiev IS, Soto C, Gorman J, Huang JH, Acharya P, Chuang GY, Ofek G, et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*. 2014;514:455.
- de Souza MS, Ratto-Kim S, Chuenarom W, Schuetz A, Chantakulkij S, Nuntapinit B, Valencia-Micolta A, Thelian D, Nitayaphan S, Pitisuttithum P, et al. The Thai phase III trial (RV144) vaccine regimen induces T cell responses that preferentially target epitopes within the V2 region of HIV-1 envelope. *J Immunol*. 2012;188:5166–76.
- Liao HX, Bonsignori M, Alam SM, McLellan JS, Tomaras GD, Moody MA, Kozink DM, Hwang KK, Chen X, Tsao CY, et al. Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. *Immunity*. 2013;38:176–86.
- Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Premisi N, Namwat C, de Souza M, Adams E, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N Engl J Med*. 2009;361:2209–20.
- Rolland M, Edlefsen PT, Larsen BB, Tovanabutra S, Sanders-Buell E, Hertz T, deCamp AC, Carrico C, Menis S, Magaret CA, et al. Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature*. 2012;490:417–20.
- Patel PH, Preston BD. Marked infidelity of human-immunodeficiency-virus type-1 reverse-transcriptase at RNA and DNA-template ends. *Proc Natl Acad Sci USA*. 1994;91:549–53.
- Cuevas JM, Geller R, Garijo R, Lopez-Aldeguer J, Sanjuan R. Extremely high mutation rate of HIV-1 in vivo. *Plos Biol*. 2015;13:e1002251.
- Baird HA, Galetto R, Gao Y, Simon-Loriere E, Abreha M, Archer J, Fan J, Robertson DL, Arts EJ, Negroni M. Sequence determinants of breakpoint location during HIV-1 intersubtype recombination. *Nucleic Acids Res*. 2006;34:5203–16.
- Zhuang JL, Jetzt AE, Sun GL, Yu H, Klarmann G, Ron Y, Preston BD, Dougherty JP. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol*. 2002;76:11273–82.
- Galetto R, Moumen A, Giacomoni V, Veron M, Charneau P, Negroni M. The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. *J Biol Chem*. 2004;279:36625–32.
- Drummond DA, Silberg JJ, Meyer MM, Wilke CO, Arnold FH. On the conservative nature of intragenic recombination. *Proc Natl Acad Sci USA*. 2005;102:5380–5.
- Simon-Loriere E, Rossolillo P, Negroni M. RNA structures, genomic organization and selection of recombinant HIV. *RNA Biol*. 2011;8:280–6.
- Julien JP, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse PJ, Burton DR, Sanders RW, Moore JP, et al. Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science*. 2013;342:1477–83.
- Lyumkis D, Julien JP, de Val N, Cupo A, Potter CS, Klasse PJ, Burton DR, Sanders RW, Moore JP, Carragher B, et al. Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science*. 2013;342:1484–90.
- Bartesaghi A, Merk A, Borgnia MJ, Milne JLS, Subramaniam S. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nat Struct Mol Biol*. 2013;20:1352–7.
- Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*. 1998;393:648–59.
- Pancera M, Majeed S, Ban YEA, Chen L, Huang CC, Kong L, Kwon YD, Stuckey J, Zhou TQ, Robinson JE, et al. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc Natl Acad Sci USA*. 2010;107:1166–71.
- Hu GQ, Liu J, Taylor KA, Roux KH. Structural comparison of HIV-1 envelope spikes with and without the V1/V2 loop. *J Virol*. 2011;85:2741–50.
- Huang CC, Lam SN, Acharya P, Tang M, Xiang SH, Hussan SSU, Stanfield RL, Robinson J, Sodroski J, Wilson IA, et al. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science*. 2007;317:1930–4.
- Lee B, Sharron M, Blanpain C, Doranz BJ, Vakili J, Setoh P, Berg E, Liu G, Guy HR, Durell SR, et al. Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *J Biol Chem*. 1999;274:9617–26.
- DoKwon Y, Pancera M, Acharya P, Georgiev IS, Crooks ET, Gorman J, Joyce MG, Guttman M, Ma XC, Narpala S, et al. Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat Struct Mol Biol*. 2015;22:522.
- Mann JK, Ndung'u T. HIV-1 vaccine immunogen design strategies. *Virology*. 2015;12:3.
- Galli A, Kearney M, Nikolaitchik OA, Yu S, Chin MPS, Maldarelli F, Coffin JM, Pathak VK, Hu WS. Patterns of human immunodeficiency virus type 1 recombination *ex vivo* provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. *J Virol*. 2010;84:7651–61.
- Simon-Loriere E, Galetto R, Hamoudi M, Archer J, Lefevre P, Martin DP, Robertson DL, Negroni M. Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. *Plos Pathog*. 2009;5:e1000418.
- Hamoudi M, Simon-Loriere E, Gasser R, Negroni M. Genetic diversity of the highly variable V1 region interferes with human immunodeficiency virus type 1 envelope functionality. *Retrovirology*. 2013;10:114.
- Scharf L, Scheid JF, Lee JH, West AP, Chen C, Gao H, Gnanapragasam PNP, Mares R, Seaman MS, Ward AB, et al. Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Rep*. 2014;7:785–95.
- Scharf L, Wang HQ, Gao H, Chen SY, McDowall AW, Bjorkman PJ. Broadly neutralizing antibody 8ANC195 recognizes closed and open states of HIV-1 Env. *Cell*. 2015;162:1379–90.
- West AP, Scharf L, Horwitz J, Klein F, Nussenzweig MC, Bjorkman PJ. Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc Natl Acad Sci USA*. 2013;110:10598–603.
- Moore JP, Mckeating JA, Weiss RA, Sattentau QJ. Dissociation of Gp120 from HIV-1 virions induced by soluble CD4. *Science*. 1990;250:1139–42.
- Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, et al. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature*. 2011;477:466–U117.
- Garcia-Perez J, Staropoli I, Azoulay S, Heinrich JT, Cascajero A, Colin P, Lortat-Jacob H, Arenzana-Seisdedos F, Alcami J, Kellenberger E, Lagane B. A single-residue change in the HIV-1 V3 loop associated with maraviroc resistance impairs CCR5 binding affinity while increasing replicative capacity. *Retrovirology*. 2015;12:50.
- Munro JB, Mothes W. Structure and dynamics of the native HIV-1 Env trimer. *J Virol*. 2015;89:5752–5.
- Travers SAA, Tully DC, McCormack GP, Fares MA. A study of the coevolutionary patterns operating within the Env gene of the HIV-1 group M subtypes. *Mol Biol Evol*. 2007;24:2787–801.
- Cashin K, Sterjovski J, Harvey KL, Ramsland PA, Churchill MJ, Gorry PR. Covariance of charged amino acids at positions 322 and 440 of HIV-1 Env contributes to coreceptor specificity of subtype B viruses, and can be used to improve the performance of V3 sequence-based coreceptor usage prediction algorithms. *Plos One*. 2014;9:e109771.
- Dimonte S, Mercurio F, Svicher V, D'Arrigo R, Perno CF, Ceccherini-Silberstein F. Selected amino acid mutations in HIV-1 B subtype gp41 are Associated with Specific gp120(V3) signatures in the regulation of co-receptor usage. *Retrovirology*. 2011;8:33.

40. Huang W, Toma J, Fransen S, Stawiski E, Reeves JD, Whitcomb JM, Parkin N, Petropoulos CJ. Coreceptor tropism can be influenced by amino acid substitutions in the gp41 transmembrane subunit of human immunodeficiency virus type 1 envelope protein. *J Virol*. 2008;82:5584–93.
41. Lombardi F, Nakamura KJ, Chen T, Sobrera ER, Tobin NH, Aldrovandi GM. A conserved glycan in the C2 domain of HIV-1 envelope acts as a molecular switch to control X4 utilization by clonal variants with identical V3 loops. *Plos One*. 2015;10:e0128116.
42. Dimonte S, Babakir-Mina M, Mercurio F, Di Pinto D, Ceccherini-Silberstein F, Svicher V, Perno CF. Selected amino acid changes in HIV-1 subtype-C gp41 are associated with specific gp120(V3) signatures in the regulation of co-receptor usage. *Virus Res*. 2012;168:73–83.
43. Geller R, Domingo-Calap P, Cuevas J, Rossolillo P, Negroni M, Sanjuan R. The external domains of the HIV-1 envelope are a mutational cold spot. *Nat Commun*. 2015;6:8571.
44. Gitlin L, Hagai T, LaBarbera A, Solovey M, Andino R. Rapid evolution of virus sequences in intrinsically disordered protein regions. *Plos Pathog*. 2014;10:e1004529.
45. Nilsson J, Grahn M, Wright APH. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol*. 2011;12:R65.
46. Schlessinger A, Schaefer C, Vicedo E, Schmidberger M, Punta M, Rost B. Protein disorder—a breakthrough invention of evolution? *Curr Opin Struct Biol*. 2011;21:412–8.
47. Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, et al. Structure of a V3-containing HIV-1 gp120 core. *Science*. 2005;310:1025–8.
48. Pancera M, Shahzad-ul-Hussan S, Doria-Rose NA, McLellan JS, Bailer RT, Dai KF, Loesgen S, Louder MK, Staupel RP, Yang YP, et al. Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1–V2-directed antibody PG16. *Nat Struct Mol Biol*. 2013;20:804.
49. Brandenburg OF, Magnus C, Rusert P, Regoes RR, Trkola A. Different infectivity of HIV-1 strains is linked to number of envelope trimers required for entry. *Plos Pathog*. 2015;11:e1004595.
50. Klasse PJ. Modeling how many envelope glycoprotein trimers per virion participate in human immunodeficiency virus infectivity and its neutralization by antibody. *Virology*. 2007;369:245–62.
51. Rusert P, Krarup A, Magnus C, Brandenburg OF, Weber J, Ehler AK, Regoes RR, Gunthard HF, Trkola A. Interaction of the gp120 V1V2 loop with a neighboring gp120 unit shields the HIV envelope trimer against cross-neutralizing antibodies. *J Exp Med*. 2011;208:1419–33.
52. White TA, Bartesaghi A, Borgnia MJ, Meyerson JR, de la Cruz MJV, Bess JW, Nandwani R, Hoxie JA, Lifson JD, Milne JLS, Subramaniam S. Molecular architectures of trimeric SIV and HIV-1 envelope glycoproteins on intact viruses: strain-dependent variation in quaternary structure. *Plos Pathog*. 2010;6:e1001249.
53. Connor RI, Chen BK, Choe S, Landau NR. Vpr is required for efficient replication of human-immunodeficiency-virus type-1 in mononuclear phagocytes. *Virology*. 1995;206:935–44.
54. Njai HF, Gali Y, Vanham G, Clybergh C, Jennes W, Vidal N, Butel C, Mpoudi-Ngolle E, Peeters M, Arien KK. The predominance of human immunodeficiency virus type I (HIV-I) circulating recombinant form 02 (CRF02\_AG) in West Central Africa may be related to its replicative fitness. *Retrovirology*. 2006;3:40.
55. Healey D, Dianda L, Moore JP, McDougal JS, Moore MJ, Estess P, Buck D, Kwong PD, Beverley PCL, Sattentau QJ. Novel anti-Cd4 monoclonal-antibodies separate human-immunodeficiency-virus infection and fusion of CD4+ cells from virus binding. *J Exp Med*. 1990;172:1233–42.
56. Braibant M, Xie J, Samri A, Agut H, Autran B, Barin F. Disease progression due to dual infection in an HLA-B57-positive asymptomatic long-term nonprogressor infected with a nef-defective HIV-1 strain. *Virology*. 2010;405:81–92.
57. Montefiori D. Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol Biol*. 2009;485:395–405.
58. Johnson VA, Byington RE. Infectivity assay (virus yield assay). In: Aldrovani A, Walker BD, editors. *Techniques in HIV research*. New York: Stockton Press; 1990. p. 71–6.
59. Long EM, Rainwater SM, Lavreys L, Mandaliya K, Overbaugh J. HIV type 1 variants transmitted to women in Kenya require the CCR5 coreceptor for entry, regardless of the genetic complexity of the infecting virus. *AIDS Res Hum Retroviruses*. 2002;18:567–76.
60. Li M, Gao F, Mascola JR, Stamatatos L, Polonis VR, Koutsoukos M, Voss G, Goepfert P, Gilbert P, Greene KM, et al. Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol*. 2005;79:10108–25.
61. Li M, Salazar-Gonzalez JF, Derdeyn CA, Morris L, Williamson C, Robinson JE, Decker JM, Li Y, Salazar MG, Polonis VR, et al. Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular env clones from acute and early heterosexually acquired infections in Southern Africa. *J Virol*. 2006;80:11776–90.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

