



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẢNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0036291

(51)⁷ A61K 39/12; C07K 14/005

(13) B

(21) 1-2019-01815

(22) 14/09/2017

(86) PCT/EP2017/073141 14/09/2017

(87) WO/2018/050747 22/03/2018

(30) 16188866.4 15/09/2016 EP

(45) 25/07/2023 424

(43) 25/09/2019 378A

(73) JANSSEN VACCINES & PREVENTION B.V. (NL)

Archimedesweg 4 2333 CN Leiden, NL

(72) RUTTEN, Lucy (NL); TRUAN, Daphné (CH); STROKAPPE, Nika, Mindy (NL);
LANGEDIJK, Johannes, Petrus, Maria (NL).

(74) Công ty Cổ phần Sở hữu công nghiệp INVESTIP (INVESTIP)

(54) PROTEIN VỎ (ENV) CỦA VIRUT GÂY SUY GIẢM MIỄN DỊCH TÁI TỔ HỢP Ở NGƯỜI, PHƯƠNG PHÁP CẢI THIỆN SỰ TẠO THÀNH TRIME CỦA PROTEIN ENV HIV

(57) Sáng chế đề xuất protein vỏ virus gây suy giảm miễn dịch ở người (Human Immunodeficiency Virus - HIV) có các đột biến mà làm ổn định dạng trime của protein vỏ được đề xuất. Protein vỏ HIV có một số thể axit amin nhất định ở các vị trí cụ thể trong trình tự protein vỏ Protein vỏ HIV được mô tả ở đây có phần trăm tạo thành trime được cải thiện và/hoặc hiệu suất trime được cải thiện so với protein vỏ HIV mà không có một hoặc nhiều trong số các thể axit amin được nêu. Sáng chế cũng đề xuất phân tử axit nucleic và vectơ mã hóa protein vỏ HIV, cũng như chế phẩm chứa protein vỏ HIV, axit nucleic, và vectơ và các phương pháp tạo ra protein Env của HIV tái tổ hợp.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến protein vỏ (env) của virus gây suy giảm miễn dịch tái tổ hợp ở người, phân tử axit nucleic phân lập, vector, phương pháp tạo ra và chế phẩm bao gồm protein này.

Tình trạng kỹ thuật của sáng chế

Virus gây suy giảm miễn dịch ở người (Human Immunodeficiency Virus - HIV) ảnh hưởng đến hàng triệu người trên thế giới, và việc ngăn ngừa HIV thông qua vắc-xin hữu hiệu vẫn được ưu tiên rất cao, ngay cả trong thời đại điều trị kháng retrovirus phổ biến. Sự đa dạng về kháng nguyên giữa các chủng khác nhau và các nhánh của virus HIV gây khó khăn cho việc phát triển vắc-xin với hiệu quả rộng. HIV-1 là chủng virus gây bệnh và phổ biến nhất, với hơn 90% ca HIV/AIDS có nguồn gốc từ sự nhiễm HIV-1 nhóm M. Nhóm M được chia tiếp thành các nhánh hoặc các kiểu phụ khác, trong đó nhánh C là lớn nhất. Vắc-xin hữu hiệu lý tưởng là có khả năng gây ra cả đáp ứng tế bào hiệu nghiệm và kháng thể trung hòa rộng có khả năng làm trung hòa các chủng HIV-1 từ các nhánh khác nhau.

Gai protein vỏ (Env) trên bề mặt HIV bao gồm các trimer của các dimer của glycoprotein gp120 và gp41 (Hình 1A Protein tiền chất gp160 được phân cắt bởi furin thành gp120, mà là đầu của gai và chứa vị trí liên kết thụ thể CD4 cũng như các vòng siêu biến (V1 đến V5), và gp41, mà là thân neo màng của gai protein vỏ. Giống như các protein gây dung hợp loại I khác, gp41 chứa peptit dung hợp đầu tận cùng N (FP), vùng xuyên màng đầu tận cùng C (TM), và vùng tế bào chất. Dung hợp màng giữa HIV và màng tế bào đích cần hàng loạt thay đổi hình dạng trong protein vỏ. Vắc-xin HIV có thể được phát triển dựa trên protein vỏ.

Tuy nhiên, các yếu tố khác nhau làm cho việc phát triển của vắc-xin HIV dựa trên protein vỏ gặp nhiều thách thức, bao gồm cả khả năng biến đổi gen cao của HIV-1, lớp vỏ carbohydrat dày của protein vỏ, và bản chất tương đối động và dễ hủy của cấu trúc gai protein vỏ. Protein vỏ kiểu đại là không ổn định do chức năng của nó. Do đó,

việc cải biến làm ổn định đôi khi được đưa vào cấu trúc vỏ để tạo ra các ứng viên vắc-xin. Protein vỏ là mục tiêu để trung hòa các kháng thể và được glycosyl hóa cao, giúp làm giảm khả năng miễn dịch bằng cách che chắn các epitop protein. Tất cả các kháng thể trung hòa lớn đã biết (bNAb) chứa các glycan này.

Để phát triển vắc-xin, ưu tiên là sử dụng các protein vỏ có thể tạo ra bNAb. Tuy nhiên, hầu hết các bNAb chỉ nhận diện hình dạng protein vỏ tự nhiên trước khi nó trải qua thay đổi hình dạng bất kỳ. Do đó, việc phát triển protein vỏ ổn định ở hình dạng đóng và nén giống tự nhiên của nó, đồng thời giảm thiểu thể hiện các epitop không tự nhiên và do đó không trung hòa, có thể cải thiện hiệu quả của việc tạo ra các bNAb như vậy. Nỗ lực trước đây để tạo ra vắc-xin HIV đã tập trung vào việc phát triển vắc-xin có chứa vùng ngoại bào trước dung hợp của protein vỏ HIV trimer, gp140. Gp140 không có vùng xuyên màng (TM) và vùng tế bào chất, nhưng không giống như gp120, nó có thể tạo thành các cấu trúc trimer. Hơn nữa, các nỗ lực trước đây chủ yếu tập trung vào nhánh A. Tuy nhiên, bề rộng của đáp ứng kháng thể trung hòa đã gây ra vẫn còn hạn chế. Do đó, cũng sẽ có lợi nếu các trimer vỏ tự nhiên được làm ổn định chống nhiều nhánh HIV có sẵn.

Trong hơn hai thập kỷ, những nỗ lực đã được thực hiện để phát triển protein vỏ ổn định ở hình dạng trimer trước dung hợp của nó chỉ có thành công hạn chế trong việc tạo ra các trimer ổn định, tan được của protein vỏ có khả năng tạo ra đáp ứng kháng thể trung hòa rộng. Ví dụ, cái gọi là đột biến SOSIP (501C, 605C và 559P) đã được đưa vào trình tự protein vỏ để cải thiện sự hình thành đoạn trimer gp140 tan được (Sanders et al., (2002), *J. Virol.* 76(17): 8875-89). Cái gọi là đột biến SOSIP bao gồm gốc xystein ở các vị trí 501 và 605, và gốc prolin ở vị trí 559 theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1, mà là sơ đồ đánh số thông thường dùng trong lĩnh vực. Việc đưa vào hai gốc xystein ở các vị trí 501 và 605, mà gần nhau trong cấu trúc protein ba chiều tạo ra cầu disulfua. Protein vỏ đột biến SOSIP, như BG505_SOSIP và B41_SOSIP (protein vỏ từ chủng HIV BG505 và chủng B41 (tức là 9032-08.A1.4685) với đột biến SOSIP), đã được dùng trong nghiên cứu vắc-xin và thể hiện tạo ra Abs trung hòa tự rụng cấp 2 (Sanders et al., *Science* (2015), 349(6224): 139-140).

Tuy nhiên, mặc dù cái gọi là đột biến SOSIP có khả năng làm ổn định dạng trimer của protein vỏ, các đoạn trimer của các đột biến SOSIP này thường dưới 10%, với lượng lớn monome và chất kết tụ vẫn tạo ra. Ngay cả đột biến SOSIP BG505_SOSIP, là một

trong những protein vỏ đột biến SOSIP hứa hẹn nhất đã biết cho đến nay về khả năng làm ổn định dạng trimer thường chỉ mang lại đến 25% dạng trimer (Julien et al., *Proc. Nat. Acad. Sci.* (2015), 112(38), 11947-52). Hơn nữa, trong đoạn trimer này, các trimer không hoàn toàn ổn định vì chúng lộ ra ở đỉnh. Do đó, ngoài các đột biến SOSIP, một số thay thế bổ sung, như E64K, A316W, và 201C-433C, đã được thiết kế để làm ổn định đỉnh và ngăn không cho nó lộ ra (de Taeye et al., *Cell* (2015), 163(7), 1702-15; Kwon et al., (2015) *Nat. Struct. Mol. Biol.* 22(7) 522-31).

Theo đó, cần phải có các trimer được làm ổn định của protein vỏ HIV mà có tỷ lệ phần trăm của hình dạng trimer cải thiện, hiệu suất trimer cải thiện, và/hoặc tính ổn định trimer được cải thiện. Tốt hơn là, các trimer được làm ổn định của protein vỏ HIV cũng có thể cho thấy sự liên kết tốt với các kháng thể trung hòa rộng (bNAb), và liên kết tương đối hạn chế với các Ab không trung hòa rộng (không bNAb). Mục đích của sáng chế là đề xuất protein Env HIV mà có tỷ lệ phần trăm trimer cải thiện, và tốt hơn là hiệu suất trimer cũng cải thiện.

Bản chất kỹ thuật của sáng chế

Sáng chế đề cập đến protein vỏ HIV tái tổ hợp từ các nhánh khác nhau mà có tỷ lệ phần trăm cải thiện của sự tạo thành trimer và/hoặc hiệu suất trimer cải thiện so với trimer vỏ HIV được mô tả trước đây. Sự gấp Env được tối ưu, các dấu hiệu đặc hiệu chúng được sửa, và các vùng của hình dạng đóng trước dung hợp quan trọng cho quá trình dung hợp được làm ổn định bằng đột biến được mô tả ở đây. Điều này tạo ra tiếp cận phổ quát để tối ưu hóa sự gấp và tính ổn định của trimer vỏ HIV-1 đóng trước dung hợp. Các trimer Env HIV ổn định và được gấp gọn tạo ra là hữu dụng cho mục đích chủng ngừa, ví dụ cải thiện cơ hội tạo ra các kháng thể trung hòa rộng và làm giảm cảm ứng của các kháng thể không trung hòa và trung hòa yếu khi sử dụng các trimer Env HIV tái tổ hợp.

Sáng chế còn đề cập đến phân tử axit nucleic phân lập và vector mã hóa protein vỏ HIV tái tổ hợp, tế bào chứa chúng, và chế phẩm của protein vỏ HIV tái tổ hợp, phân tử axit nucleic, vector, và/hoặc tế bào.

Theo một khía cạnh chung, sáng chế đề cập đến protein vỏ virus gây suy giảm miễn dịch ở người (HIV) tái tổ hợp có gốc axit amin cụ thể ở vị trí được nhận diện trong trình tự protein vỏ mà làm ổn định sự hình thành trimer.

Theo các phương án nhất định, protein vỏ (Env) HIV tái tổ hợp theo sáng chế bao gồm trình tự axit amin của protein Env HIV có gốc axit amin được nêu ở ít nhất hai trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp ở vị trí 655;
- (iii) Asn hoặc Gln ở vị trí 535;
- (iv) Val, Ile hoặc Ala ở vị trí 589;
- (v) Phe hoặc Trp ở vị trí 573;
- (vi) Ile ở vị trí 204; và
- (vii) Phe, Met, hoặc Ile ở vị trí 647,

trong đó việc đánh số các vị trí là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1. Theo các phương án được ưu tiên nhất định, gốc axit amin được nêu ở vị trí 651 là Phe; gốc axit amin được nêu ở vị trí 655 là Ile; gốc axit amin được nêu ở vị trí 535 là Asn; và/hoặc gốc axit amin được nêu ở vị trí 573 là Phe.

Theo các phương án nhất định, protein Env của HIV tái tổ hợp theo sáng chế bao gồm trình tự axit amin của protein Env HIV và sự thế axit amin bằng gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp ở vị trí 655;
- (iii) Asn hoặc Gln ở vị trí 535;
- (iv) Val, Ile hoặc Ala ở vị trí 589;
- (v) Phe hoặc Trp ở vị trí 573;
- (vi) Ile ở vị trí 204; và
- (vii) Phe, Met, hoặc Ile ở vị trí 647,

trong đó, protein Env HIV được chọn từ nhóm bao gồm:

- (1) Trình tự axit amin phổ biến Env HIV, ví dụ từ nhánh C (ví dụ bao gồm trình tự axit amin nêu ở SEQ ID NO: 2 hoặc 3) hoặc từ nhánh B (ví dụ

bao gồm trình tự axit amin nêu ở SEQ ID NO: 4 hoặc 5);

(2) protein Env HIV tổng hợp, ví dụ có chứa (a): trình tự axit amin nêu ở SEQ ID NO: 6, hoặc (b): SEQ ID NO: 6 với đột biến Glu thành Arg ở vị trí 166, hoặc (c): (a) hoặc (b) với đột biến của axit amin ở các vị trí 501 và 605 vào gốc Cys và đột biến của axit amin ở vị trí 559 vào gốc Pro, hoặc (d): (a), (b) hoặc (c) có đột biến điểm phân cắt furin nữa, ví dụ thay thế axit amin ở các vị trí 508-511 bằng RRRRRR (SEQ ID NO: 10), hoặc (e) SEQ ID NO: 7, hoặc (f) trình tự Env khảm như Env chứa trình tự axit amin nêu ở SEQ ID NO: 8 hoặc 9; và

(3) protein Env HIV bố mẹ, mà tốt hơn là là protein Env HIV kiểu đại, tốt hơn là của nhánh C, bao gồm ít nhất một đột biến sửa ở gốc axit amin mà thấy ở vị trí tương ứng ở tần suất ít hơn 7,5%, tốt hơn là ít hơn 2%, của trình tự Env HIV trong tập hợp của ít nhất 100, tốt hơn là ít nhất 1000, tốt hơn là ít nhất 10000, trình tự Env HIV kiểu đại, trong đó đột biến sửa là sự thế bằng gốc axit amin mà thấy ở vị trí tương ứng ở tần suất bằng ít nhất 10% của trình tự Env HIV trong tập hợp này và tốt hơn là đột biến sửa là sự thế bằng gốc axit amin mà thấy ở vị trí tương ứng thường xuyên nhất trong tập hợp này;

và việc đánh số các vị trí theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1. Theo các phương án được ưu tiên nhất định, gốc axit amin được nêu ở vị trí 651 là Phe; gốc axit amin được nêu ở vị trí 655 là Ile; gốc axit amin được nêu ở vị trí 535 là Asn; và/hoặc gốc axit amin được nêu ở vị trí 573 là Phe.

Theo các phương án nhất định, protein Env của HIV tái tổ hợp theo sáng chế bao gồm trình tự axit amin của protein Env HIV và sự thế axit amin bằng gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp ở vị trí 655;
- (iii) Asn hoặc Gln ở vị trí 535;
- (iv) Val, Ile hoặc Ala ở vị trí 589;
- (v) Phe hoặc Trp ở vị trí 573;

(vi) Ile ở vị trí 204; và

(vii) Phe, Met, hoặc Ile ở vị trí 647,

trong đó protein Env HIV là protein Env HIV đột biến SOSIP bao gồm ít nhất một đột biến tạo ra gốc axit amin được nêu(s) ở các vị trí được nêu được chọn từ nhóm bao gồm:

(a) Cys ở các vị trí 501 và 605;

(b) Pro ở vị trí 559; và

(c) Cys ở các vị trí 501 và 605 và Pro ở vị trí 559; và

việc đánh số các vị trí là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1. theo các phương án được ưu tiên nhất định, gốc axit amin được nêu ở vị trí 651 là Phe; gốc axit amin được nêu ở vị trí 655 là Ile; gốc axit amin được nêu ở vị trí 535 là Asn; và gốc axit amin được nêu ở vị trí 573 là Phe.

Theo phương án khác, protein Env của HIV tái tổ hợp theo sáng chế còn bao gồm gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm:

(viii) Gln, Glu, Ile, Met, Val, Trp hoặc Phe, tốt hơn là Gln hoặc Glu, ở vị trí 588;

(ix) Lys ở vị trí 64, hoặc Arg ở vị trí 66, hoặc Lys ở vị trí 64 và Arg ở vị trí 66;

(x) Trp ở vị trí 316;

(xi) Cys ở cả vị trí 201 và 433;

(xii) Pro ở vị trí 556, hoặc Pro ở vị trí 558, hoặc Pro ở các vị trí 556 và 558; (xiii) thay thế vòng ở vị trí axit amin 548-568 (vòng HR1) bằng vòng có 7-10 axit amin, tốt hơn là vòng có 8 axit amin, ví dụ có trình tự được chọn từ một trình tự bất kỳ trong số (SEQ ID NO: 12-17);

(xiv) Gly ở vị trí 568, hoặc Gly ở vị trí 569, hoặc Gly ở vị trí 636, hoặc Gly ở cả vị trí 568 và 636, hoặc Gly ở cả vị trí 569 và 636; và/hoặc

(xv) Tyr ở vị trí 302, hoặc Arg ở vị trí 519, hoặc Arg ở vị trí 520, hoặc Tyr ở vị trí 302 và Arg ở vị trí 519, hoặc Tyr ở vị trí 302 và Arg ở vị trí 520, hoặc Tyr ở vị trí 302 và Arg ở cả vị trí 519 và 520.

Theo các phương án nhất định, protein Env của HIV tái tổ hợp theo sáng chế còn bao gồm đột biến ở trình tự phân cắt furin của protein Env HIV, như sự thay thế ở các vị trí 508-511 bằng RRRRRR (SEQ ID NO: 10).

Theo một phương án, protein Env HIV tái tổ hợp là protein gp140.

Theo phương án khác, protein Env HIV tái tổ hợp là protein gp160.

Theo các phương án nhất định, protein Env HIV tái tổ hợp bị cắt cụt trong vùng tế bào chất, ví dụ sau 7 axit amin của vùng tế bào chất.

Theo khía cạnh chung khác, sáng chế đề cập đến phức trime bao gồm oligome không cộng hóa trị của ba trong số bất kỳ của các protein Env của HIV tái tổ hợp được mô tả ở đây.

Khía cạnh chung khác của sáng chế đề xuất phương pháp cải thiện sự gấp và tính ổn định (được đo là phần trăm trime và/hoặc hiệu suất trime tăng) của protein Env HIV bố mẹ, phương pháp bao gồm sửa trình tự axit amin của protein Env HIV bố mẹ bằng cách đưa vào ít nhất một đột biến sửa, tốt hơn là ít nhất 3 đột biến sửa trong protein Env HIV bố mẹ, trong đó đột biến sửa là sự thay thế axit amin ở gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất ít hơn 7,5%, tốt hơn là ít hơn 2%, của trình tự Env HIV trong tập hợp của ít nhất 100, tốt hơn là ít nhất 500, tốt hơn là ít nhất 1000, tốt hơn là ít nhất 10000, trình tự Env HIV kiểu dại, trong đó sự thay thế là bằng gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất bằng ít nhất 10% của trình tự Env HIV trong tập hợp này và tốt hơn là sự thay thế là bằng gốc axit amin mà có mặt ở vị trí tương ứng thường xuyên nhất trong tập hợp này. Sáng chế còn đề xuất protein Env của HIV được sửa mà có thể thu được bằng phương pháp này theo sáng chế để cải thiện sự gấp và tính ổn định (được đo là tỷ lệ phần trăm trime và/hoặc hiệu suất trime) của protein Env HIV. Sáng chế còn đề xuất được phẩm bao gồm protein Env của HIV được sửa này. Sáng chế còn đề xuất phương pháp tạo ra protein Env HIV, bao gồm phương pháp sửa protein Env của HIV được mô tả ở đây, và biểu hiện axit nucleic mã hóa protein Env của HIV đã sửa được làm ổn định trong tế bào chủ tái tổ hợp.

Theo khía cạnh chung khác, sáng chế đề cập đến protein Env của HIV tái tổ hợp bao gồm trình tự axit amin mà là ít nhất 95%, 96%, 97%, 98%, 99% hoặc 100% tương đồng với SEQ ID NO: 2, trong đó tốt hơn là các vị trí 204, 535, 573, 589, 647, 651, và 655, và tốt hơn là các vị trí khác 64, 66, 201, 316, 433, 501, 508-511, 556, 558, 559, 588, 548-568 và 605 không được xem xét khi xác định % tương đồng, và trong đó việc đánh số là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1. Theo các phương án nhất định của chúng, protein Env của HIV tái tổ hợp bao gồm trình tự axit amin mà là ít nhất 98%, 99% hoặc 100% tương đồng với SEQ ID NO: 3, trong đó tốt hơn là các vị trí 204, 535, 573, 589, 647, 651, và 655, và tốt hơn là các vị trí khác 64, 66, 201, 316, 433, 508-511, 556, 558, 588, và 548-568 không được xem xét khi xác định % tương đồng, và trong đó việc đánh số là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1.

Theo khía cạnh chung khác, sáng chế đề cập đến protein Env của HIV tái tổ hợp bao gồm trình tự axit amin mà là ít nhất 95%, 96%, 97%, 98%, 99% hoặc 100% tương đồng với SEQ ID NO: 4, trong đó tốt hơn là các vị trí 204, 535, 573, 589, 647, 651, và 655, và tốt hơn là các vị trí khác 64, 66, 201, 316, 433, 501, 508-511, 556, 558, 559, 588, 548-568 và 605 không được xem xét khi xác định % tương đồng, và trong đó việc đánh số là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1. Theo các phương án nhất định của chúng, protein Env của HIV tái tổ hợp bao gồm trình tự axit amin mà là ít nhất 98%, 99% hoặc 100% tương đồng với SEQ ID NO: 5, trong đó tốt hơn là các vị trí 204, 535, 573, 589, 647, 651, và 655, và tốt hơn là các vị trí khác 64, 66, 201, 316, 433, 508-511, 556, 558, 588, và 548-568 không được xem xét khi xác định % tương đồng, và trong đó việc đánh số là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1.

Theo các khía cạnh và phương án này, một hoặc nhiều trong số axit amin ở các vị trí được nêu mà không được xem xét để xác định % tương đồng, tốt hơn là được chọn từ axit amin được nêu là được ưu tiên ở đây, ví dụ Ile ở vị trí 204; Phe, Ala, Leu, hoặc Trp ở vị trí 651; v.v. (xem các Bảng 1 và 2 dưới đây).

Theo khía cạnh chung khác, sáng chế đề cập đến protein Env của HIV tái tổ hợp bao gồm trình tự axit amin mà là ít nhất 95%, 96%, 97%, 98%, 99% hoặc 100% tương đồng với bất kỳ trong số các SEQ ID NO: 2, 3, 4, 5, 20, 22, 24, 26, 27, 28, 29, 30, 31, hoặc 32, trong đó SEQ ID NO: 20, 22, 24, 26, 27, 28, 29, 30, 31, hoặc 32 là được đặc

biệt ưu tiên. Theo một khía cạnh, tốt hơn là các vị trí 204, 535, 573, 589, 647, 651, 655, và 658 và tốt hơn là các vị trí khác 64, 66, 201, 316, 433, 508-511, 556, 558, 588, và 548-568 không được xem xét khi xác định % tương đồng, và trong đó việc đánh số là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1. Cũng theo khía cạnh này, một hoặc nhiều trong số axit amin ở các vị trí được nêu mà không được xem xét để xác định % tương đồng, tốt hơn là được chọn từ axit amin được nêu là được ưu tiên ở đây tại (i)-(vii) của Bảng 1, (viii)-(xv) của Bảng 2, và/hoặc (xvi) của Bảng 1, ví dụ Ile ở vị trí 204; Phe, Leu, Met, hoặc Trp ở vị trí 651; v.v.

Theo khía cạnh chung khác, sáng chế đề cập đến hạt, tốt hơn là liposome hoặc hạt nano, ví dụ hạt nano tự lắp ghép, thể hiện trên protein Env của HIV tái tổ hợp bề mặt của nó theo sáng chế.

Theo khía cạnh chung khác, sáng chế đề cập đến phân tử axit nucleic phân lập mã hóa protein Env của HIV tái tổ hợp theo sáng chế và vector bao gồm phân tử axit nucleic phân lập liên kết hoạt động với trình tự khởi đầu. Theo một phương án, vector là vector virus. Theo phương án khác, vector là vector biểu hiện. Theo một phương án được ưu tiên, vector virus là vector adenovirus.

Khía cạnh chung khác đề cập đến tế bào chủ bao gồm phân tử axit nucleic phân lập hoặc vector mã hóa protein Env HIV tái tổ hợp theo sáng chế. Tế bào chủ này có thể được sử dụng để sản xuất protein tái tổ hợp, biểu hiện protein tái tổ hợp, hoặc sản xuất hạt virus.

Khía cạnh chung khác đề cập đến phương pháp tạo ra protein Env của HIV tái tổ hợp, bao gồm nuôi tế bào chủ bao gồm phân tử axit nucleic phân lập hoặc vector mã hóa protein Env của HIV tái tổ hợp theo sáng chế dưới các điều kiện thích hợp để tạo ra protein Env của HIV tái tổ hợp.

Khía cạnh chung khác nữa đề cập đến chế phẩm bao gồm protein Env của HIV tái tổ hợp, phức trimer, phân tử axit nucleic phân lập, vector, hoặc tế bào chủ như được mô tả ở đây, và chất mang được dụng.

Mô tả vắn tắt các hình vẽ

Phần bản chất kỹ thuật của sáng chế nêu trên, cũng như là phần mô tả chi tiết sáng chế dưới đây, sẽ được hiểu rõ hơn khi tham khảo hình vẽ kèm theo. Cần hiểu rằng sáng chế không bị giới hạn ở các Phương án chính xác được thể hiện trong hình vẽ.

Trên các hình vẽ:

Các Hình 1A và 1B thể hiện hình vẽ sơ đồ của cấu trúc protein vỏ (Env) của HIV;

Hình 1A thể hiện protein Env của HIV chiều dài đầy đủ; và

Hình 1B thể hiện a protein Env của HIV tan được chứa cái gọi là đột biến SOSIP và cắt đứt đầu C bắt đầu ở gốc 664 theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1 (trình tự SOSIP.644);

Các Hình 2A và 2B thể hiện phần trăm của sự tạo thành trime (Hình 2A) và hiệu suất trime (Hình 2B) cho protein Env của HIV tái tổ hợp theo các phương án của sáng chế như được đo bởi thử nghiệm AlphaLISA như được mô tả trong Ví dụ 3; protein Env của HIV tái tổ hợp được thử chứa sự thể đơn, đôi, hoặc ba axit amin được đưa vào trình tự nhánh C phổ biến vỏ của HIV ConC_SOSIP (SEQ ID NO: 3); tỷ lệ phần trăm trime và hiệu suất trime được xác định dựa trên sự liên kết của kháng thể đơn dòng đặc hiệu trime (mAb) PGT145 với mỗi trong số protein Env của HIV tái tổ hợp; hiệu suất trime và tỷ lệ phần trăm tạo thành trime đối với mỗi trong các protein Env của HIV tái tổ hợp theo sáng chế được so với của protein vỏ có trình tự ConC_SOSIP mạch chính mà không có đột biến làm ổn định trime bổ sung bất kỳ được mô tả ở đây;

Hình 3 thể hiện sắc phổ từ phân tích sắc ký loại kích cỡ với tán xạ ánh sáng đa góc (SEC-MALS) của protein Env của HIV tái tổ hợp theo các phương án của sáng chế; protein Env của HIV tái tổ hợp được thử chứa sự thể axit amin đơn được đưa vào trình tự nhánh C phổ biến Env của HIV mạch chính ConC_SOSIP (SEQ ID NO: 3), và được tinh chế bằng sắc ký ái lực lectin như được mô tả trong Ví dụ 2; phân tích SEC-MALS được tiến hành như được mô tả trong Ví dụ 3; đỉnh tương ứng với dạng trime được nêu ở mỗi trong số các sắc phổ;

Hình 4 thể hiện tính ổn định nhiệt của protein Env của HIV tái tổ hợp theo các phương án của sáng chế được báo cáo là tỷ lệ phần trăm của trime còn lại sau khi xử lý nhiệt; protein Env của HIV tái tổ hợp được thử chứa sự thể axit amin đơn, đôi, hoặc ba được đưa vào trình tự nhánh C phổ biến Env của HIV mạch chính ConC_SOSIP (SEQ

ID NO: 3); protein Env của HIV tái tổ hợp được đưa qua xử lý nhiệt, và tỷ lệ phần trăm của trime còn lại sau xử lý nhiệt được xác định bằng thử nghiệm AlphaLISA như được mô tả trong Ví dụ 4; tính ổn định nhiệt của protein vỏ có trình tự ConC_SOSIP mạch chính không có bất kỳ đột biến làm ổn định trime bổ sung cũng được thể hiện;

Các Hình 5A-5B thể hiện phần trăm của sự tạo thành trime (Hình 5A) và hiệu suất trime (Hình 5B) đối với protein Env của HIV tái tổ hợp có sự thế axit amin đơn được đưa vào trình tự nhánh B phổ biến Env của HIV mạch chính ConB_SOSIP (SEQ ID NO: 5) theo các phương án của sáng chế so với của protein vỏ có trình tự ConB_SOSIP mạch chính mà không có bất kỳ đột biến làm ổn định trime bổ sung theo sáng chế như được mô tả trong Ví dụ 5; hiệu suất trime và tỷ lệ phần trăm tạo thành trime được đo bởi thử nghiệm AlphaLISA;

Các Hình 6A-6B thể hiện phần trăm của sự tạo thành trime và hiệu suất trime đối với protein Env của HIV tái tổ hợp có sự thế axit amin được đưa vào trình tự protein vỏ của HIV tổng hợp mạch chính DS_sC4_SOSIP_E166R theo các phương án của sáng chế như được mô tả trong Ví dụ 6; tỷ lệ phần trăm tạo thành trime và hiệu suất trime được đo bởi thử nghiệm AlphaLISA.

Hình 7 thể hiện sắc phổ từ phân tích sắc ký loại kích cỡ với tán xạ ánh sáng đa góc (SEC-MALS) của protein Env của HIV tái tổ hợp theo các phương án của sáng chế; protein Env của HIV tái tổ hợp được thử chứa đột biến K655I đơn với mỗi cạnh biến thể đột biến bổ sung được đưa vào trình tự nhánh C phổ biến Env của HIV mạch chính ConC_SOSIP (SEQ ID NO: 3), và được tinh chế bằng sắc ký ái lực lectin như được mô tả trong Ví dụ 2; phân tích SEC-MALS được thực hiện như được mô tả trong Ví dụ 3; đỉnh thể hiện monome gp140 trong ConC_SOSIP được nêu bởi hộp đánh bóng, ở bên phải của đỉnh trime. Bảng dưới thể hiện phóng to phần dưới của biểu đồ, sao cho nó có thể được thấy là mỗi đột biến bổ sung gây giảm thêm chiều cao của đỉnh monome gp140.

Các Hình 8A-8B thể hiện phần trăm của sự tạo thành trime (Hình 8A) và hiệu suất trime (Hình 8B) đối với BG505_SOSIP (có nguồn gốc từ chủng nhánh A kiểu dại) có thế axit amin đơn và kết hợp của các thế được đưa vào theo các phương án của sáng chế so với của protein vỏ có trình tự BG505_SOSIP mạch chính mà không có bất kỳ đột

biến làm ổn định trime bổ sung theo sáng chế như được mô tả trong Ví dụ 9. Hiệu suất trime và tỷ lệ phần trăm tạo thành trime được đo bởi thử nghiệm AlphaLISA.

Hình 9 thể hiện sắc phổ từ phân tích sắc ký loại kích cỡ với tán xạ ánh sáng đa góc (SEC-MALS) của protein Env của HIV tái tổ hợp theo các phương án của sáng chế; phân tích SEC-MALS được thực hiện trên dịch nổi nuôi cấy của tế bào chuyển nhiễm Env. Đỉnh tương ứng với dạng trime tách rửa trong khoảng từ 7 đến 7,5 phút. Dòng xám đậm là BG505_SOSIP (có nguồn gốc từ chủng nhánh A kiểu đại) và dòng xám nhạt là BG505_SOSIP với sự thế L556P, K655I, M535N, N651F, D589V, K588E.

Các Hình 10A-10B thể hiện hiệu suất trime đối với các biến thể C97ZA_SOSIP, được mô tả trong Ví dụ 10. Hiệu suất trime của C97ZA với ba đột biến làm ổn định (L556P, T651F và M535N) (Hình 10A và B). Trên Hình 10B, trình tự Env được tối ưu hóa thêm bằng các đột biến bổ sung (21 đột biến thêm) mà được thêm để sửa trình tự Env C97ZA theo khung khái niệm được mô tả trong Hình 12 và bằng cách đưa vào đột biến làm ổn định bổ sung (K655I, D589V, A204I và K588E). Hiệu suất trime và tỷ lệ phần trăm tạo thành trime được đo bởi thử nghiệm AlphaLISA. Tín hiệu được chuẩn hóa thành tín hiệu ConC_SOSIP mà được thiết lập ở 1. PNGS là vị trí N-glycosyl hóa tiềm năng.

Hình 11. Hiệu suất trime của chủng HIV-1 Env DU422 với bốn sự thế làm ổn định (xem Ví dụ 11 để biết thêm chi tiết). Tất cả số được chuẩn hóa thành ConC_SOSIP (không thể hiện) mà được thiết lập ở 1.

Hình 12. Khái niệm phổ quát đối với trình tự HIV-1 Env sửa được minh họa cho chủng C97ZA. Gốc với tần suất xuất hiện cao nhất (ở đây gọi là ‘gốc phổ biến’) trong tổng cơ sở dữ liệu HIV-1 (thanh trên) và gốc chủng C97ZA (thanh dưới) được sắp xếp từ phần trăm xuất hiện thấp đến cao của vị trí gốc C97ZA. Vị trí gốc C97ZA để thế cho gốc phổ biến được lựa chọn dựa trên tiêu chuẩn sau: Vị trí với gốc C97ZA mà xuất hiện ít hơn 2% trong trình tự cơ sở dữ liệu Env (thanh đen). Vị trí với gốc C97ZA mà xuất hiện trong khoảng từ 2% đến 7,5% trong trình tự cơ sở dữ liệu Env và bị chôn vùi hoặc bị chôn vùi một phần (thanh xám đen). Các vị trí mà phơi lộ và kỵ nước trong C97ZA và gốc phổ biến ưa nước (hai thanh xám sáng nhất) và vị trí mà là gốc phổ biến (S234N) của vị trí N-glycosyl hóa tiềm năng (PNGS).

Hình 13. Trime ENV_SOSIP của HIV đóng trước dung hợp qua sự sửa trình tự và làm ổn định đột biến. Tín hiệu AlphaLISA trong dịch nổi nuôi cấy tế bào đối với tất cả các đột biến SOSIP được chuẩn hóa thành ConC_SOSIP cho các kháng thể trung hòa rộng.

Hình 14. Đặc điểm SEC phân tích của biến thể Env_SOSIP đối chứng (SOSIP mạch chính), biến thể Env được sửa theo khái niệm được mô tả trong Ví dụ 12 và Hình 12, và biến thể Env với đột biến làm ổn định bổ sung theo bảng 3 sử dụng dịch nổi nuôi cấy tế bào sau khi chuyển nhiễm. Tín hiệu giả được của dịch nổi nuôi cấy tế bào được trừ từ tất cả các đặc điểm. Đỉnh trime được thể hiện với *.

Hình 15. Hiệu suất trime của các biến thể HIV-1 Env ConC không có các cải biến SOSIP làm ổn định.

Hình 16. Hiệu suất trime (A) và tỷ lệ phần trăm trime (B) của ConC_SOSIP với các đột biến ở các vị trí 589, 647, 651 và 655 thành methionin. Tất cả số được chuẩn hóa thành ConC_SOSIP (không thể hiện) mà được thiết lập ở 1. Thanh lỗi được thể hiện ở phía dưới bên phải của các thanh.

Các Hình 17A-17D thể hiện phần trăm của sự tạo thành trime (Hình 17A, B cho các thử nghiệm khác nhau) và hiệu suất trime (Hình 17C, D cho các thử nghiệm khác nhau) cho protein Env của HIV tái tổ hợp với đột biến được nêu như được mô tả trong Ví dụ 15, được đo bởi thử nghiệm AlphaLISA.

Hình 18 thể hiện sắc ký SEC-MALS của protein Env của HIV tái tổ hợp với đột biến được nêu, như được mô tả trong Ví dụ 15.

Mô tả chi tiết sáng chế

Các tài liệu công bố, bài báo và bằng sáng chế khác nhau được viện dẫn hoặc được mô tả trong tình trạng kỹ thuật và trong toàn bộ bản mô tả này; mỗi tài liệu tham khảo này được kết hợp ở đây nhằm mục đích tham khảo đến toàn bộ nội dung của nó. Thảo luận về tài liệu, hành động, nguyên liệu, dụng cụ, vật phẩm hoặc dạng tương tự được bao gồm trong bản mô tả này là nhằm mục đích cung cấp ngữ cảnh cho sáng chế. Thảo luận này không phải là sự thừa nhận rằng bất kỳ hoặc tất cả các vấn đề này tạo thành một phần của tình trạng kỹ thuật đối với sáng chế bất kỳ được bộc lộ hoặc được yêu cầu bảo hộ.

Trừ khi chỉ có dẫn khác, tất cả các thuật ngữ kỹ thuật và khoa học dùng trong bản mô tả này có cùng nghĩa như thường được hiểu bởi chuyên gia trong lĩnh vực mà sáng chế này thuộc về. Mặt khác, các thuật ngữ nhất định được dùng trong bản mô tả này có nghĩa như được nêu trong bản mô tả. Tất cả các bằng sáng chế, đơn sáng chế đã công bố và tài liệu công bố được viện dẫn trong bản mô tả này được kết hợp để tham khảo như thể được nêu đầy đủ trong bản mô tả này. Cần lưu ý rằng khi được sử dụng trong bản mô tả này và các yêu cầu bảo hộ kèm theo, các danh từ số ít bao gồm cả nghĩa số nhiều, trừ khi ngữ cảnh rõ ràng chỉ ra nghĩa khác.

Trừ khi có quy định khác, các giá trị số bất kỳ, như nồng độ hoặc khoảng nồng độ được mô tả ở đây, được hiểu là được biến đổi ở tất cả các trường hợp bằng thuật ngữ "khoảng." Do đó, giá trị số thường bao gồm $\pm 10\%$ giá trị cụ thể. Như được sử dụng ở đây, việc sử dụng khoảng số rõ ràng bao gồm tất cả các khoảng ít hơn có thể, tất cả các giá trị số riêng lẻ trong khoảng đó, bao gồm các số nguyên với các khoảng và đoạn của giá trị trừ khi bối cảnh quy định rõ ràng khác đi.

Axit amin được đề cập trong toàn bộ bản mô tả. Có hai mươi axit amin có trong tự nhiên, cũng như nhiều axit amin không có trong tự nhiên. Mỗi axit amin đã biết, bao gồm cả axit amin tự nhiên và không tự nhiên, có tên đầy đủ, mã viết tắt một chữ, và mã viết tắt ba chữ, tất cả trong đó là đã biết với chuyên gia trong lĩnh vực. Ví dụ, mã viết tắt ba và một chữ cái được dùng cho hai mươi axit amin có trong tự nhiên là như sau: alanin (Ala; A), arginin (Arg; R), axit aspartic (Asp; D), asparagin (Asn; N), cystein (Cys; C), glyxin (Gly; G), axit glutamic (Glu; E), glutamin (Gln; Q), histidin (His; H), isoleuxin (Ile; I), leuxin (Leu; L), lysin (Lys; K), methionin (Met; M), phenylalanin (Phe; F), prolin (Pro; P), serin (Ser; S), threonin (Thr; T), tryptophan (Trp; W), tyrosin (Tyr; Y) và valin (Val; V). Axit amin có thể được gọi bằng tên đầy đủ, mã viết tắt một chữ cái, hoặc mã viết tắt ba chữ cái của chúng.

Trừ khi bối cảnh quy định rõ ràng khác đi, việc đánh số các vị trí trong trình tự axit amin của protein vỏ HIV như được sử dụng ở đây là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1 như ví dụ được nêu trong Korber et al. (Human Retroviruses and AIDS 1998: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. Korber et al., Eds. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N. Mex.), mà được kết hợp vào đây bằng cách viện dẫn toàn bộ. Việc đánh số theo HXB2 là thông thường trong lĩnh vực của

protein Env của HIV. gp160 của thể phân lập HXB2 của HIV-1 có trình tự axit amin được thể hiện trong SEQ ID NO: 1. Việc sắp thẳng hàng trình tự Env của HIV có lợi với trình tự này có thể được dùng để tìm ra sự đánh số axit amin tương ứng trong trình tự có lợi.

Câu “bao gồm trình tự axit amin của protein Env của HIV có gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm” và “bao gồm một hoặc nhiều trong số các gốc sau đây (gốc axit amin)” được sử dụng tráo đổi ở đây.

Thuật ngữ “phần trăm (%) tương đồng trình tự” hoặc “% tương đồng” mô tả số trùng khớp (“các hit”) của axit amin tương đồng của hai hoặc nhiều trình tự axit amin được sắp thẳng hàng so với số gốc axit amin tạo nên chiều dài tổng thể của trình tự axit amin. Nói cách khác, việc sử dụng sự sắp thẳng hàng, cho hai hoặc nhiều trình tự tỷ lệ phần trăm của gốc axit amin mà giống nhau (ví dụ 95%, 97% hoặc 98% tương đồng) có thể được xác định, khi trình tự được so sánh và sắp thẳng hàng để tương ứng tối đa như được đo bằng cách sử dụng thuật toán so sánh trình tự như đã biết trong lĩnh vực, hoặc khi được sắp hàng thủ công và kiểm tra trực quan. Trình tự mà được so sánh để xác định tương đồng trình tự do đó có thể khác bởi (các) thế, (các) thêm hoặc (các) loại bỏ axit amin. Các chương trình thích hợp để sắp thẳng hàng trình tự protein là đã biết đối với chuyên gia trong lĩnh vực. Phần trăm tương đồng trình tự của trình tự protein có thể, ví dụ, được xác định bằng các chương trình như CLUSTALW, Clustal Omega, FASTA hoặc BLAST, ví dụ bằng cách sử dụng thuật toán NCBI BLAST (Altschul SF, et al (1997), *Nucleic Acids Res.* 25:3389-3402).

‘Tập hợp của trình tự Env của HIV’ như được sử dụng ở đây là tập hợp của số đại diện (ví dụ ít nhất 100, hoặc 500, hoặc 1000, hoặc hơn) của trình tự ngẫu nhiên của protein Env của HIV kiểu đại, mà có thể từ cùng nhánh (ví dụ nhánh C) hoặc từ các nhánh khác nhau (ví dụ nhánh A, B, C, v.v.). Các tập hợp thích hợp của các trình tự này có trong cơ sở dữ liệu, hoặc tập hợp con có thể được chiết xuất từ đó, ví dụ Cơ sở dữ liệu trình tự HIV (Los Alamos National Laboratory). Tập hợp này bao gồm tốt hơn là ít nhất 100 trình tự protein Env của HIV, 1000 trình tự protein Env của HIV, ít nhất 10000 trình tự protein Env của HIV, ít nhất 50000 trình tự protein Env của HIV, và có thể chứa nhiều hơn 90000 trình tự protein Env của HIV.

‘Vị trí tương ứng’ trong protein Env của HIV đề cập đến vị trí của gốc axit amin khi ít nhất hai trình tự Env của HIV được sắp thẳng hàng. trừ khi có quy định khác, đánh

số vị trí axit amin nhằm các mục đích này là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1, như thông lệ trong lĩnh vực.

‘Đột biến làm ổn định’ như được sử dụng ở đây là đột biến như được mô tả ở đây theo bất kỳ trong số các mục (i)-(vii), hoặc (xvi), của Bảng 1, hoặc (viii)-(xv) của Bảng 2, mà làm tăng tỷ lệ phần trăm của trime và/hoặc hiệu suất trime (mà có thể ví dụ được đo theo thử nghiệm AlphaLISA hoặc SEC-MALS được mô tả ở đây) của protein Env của HIV so với phân tử bố mẹ khi đột biến được đưa vào bằng cách thế axit amin tương ứng trong phân tử bố mẹ này. Axit amin tạo thành từ đột biến làm ổn định này thường hiếm, nếu có, thấy trong protein Env của thể phân lập HIV kiểu đại.

‘Đột biến sửa’ như được sử dụng ở đây là sự thế của gốc axit amin trong protein Env của HIV bố mẹ, mà gốc axit amin có mặt trong ít hơn 7,5%, tốt hơn là ít hơn 2%, ở vị trí tương ứng trong tập hợp của trình tự protein Env của HIV, trong đó sự thế là bằng axit amin mà có mặt ở vị trí tương ứng trong tập hợp này thường xuyên hơn, ví dụ trong ít nhất 10% protein Env của HIV trong tập hợp này, và tốt hơn là bằng axit amin mà có mặt ở vị trí tương ứng trong tập hợp này trong ít nhất 20% protein Env của HIV hoặc axit amin xảy ra thường xuyên nhất ở vị trí tương ứng trong tập hợp này. Axit amin tạo thành từ đột biến sửa này do đó thường thấy ở tỷ lệ phần trăm tương đối cao của protein Env của thể phân lập HIV kiểu đại, và có thể trong một vài trường hợp là giống như ở vị trí tương ứng trong trình tự Env của HIV phổ biến.

Trình tự Env HIV ‘được sửa và được làm ổn định’ như được sử dụng ở đây thường chứa ít nhất một đột biến sửa và ít nhất một đột biến làm ổn định, tốt hơn là nhiều đột biến sửa và nhiều đột biến làm ổn định so với trình tự Env HIV bố mẹ.

Thuật ngữ ‘tự nhiên’ hoặc ‘kiểu đại’ được sử dụng tráo đổi ở đây khi đề cập đến chủng HIV (hoặc protein Env từ đó), và đề cập đến chủng HIV (hoặc protein Env từ đó) như có trong tự nhiên, ví dụ như ở bệnh nhân mắc HIV.

Sáng chế nói chung đề cập đến protein vỏ (Env) tái tổ hợp của HIV bao gồm các thể axit amin nhất định ở vị trí được nêu trong trình tự protein vỏ mà làm ổn định dạng trime của protein vỏ. Việc đưa một hoặc nhiều trong số các thể axit amin được xác định theo sáng chế vào trình tự của protein vỏ của HIV có thể tạo ra tỷ lệ phần trăm tạo thành trime tăng và/hoặc hiệu suất trime tăng. Điều này có thể ví dụ được đo bằng cách sử dụng kháng thể đặc hiệu trime, nhiệt độ nóng chảy, sắc ký loại kích cỡ, và liên kết với kháng thể mà liên kết với protein Env gấp chính xác (trime ổn định) hoặc tùy ý được

gấp không chính xác (không ổn định không trime), và phần trăm trime và/hoặc hiệu suất trime tăng được xem là chỉ báo của protein Env ổn định, tự nhiên, gấp chính xác.

Virus gây suy giảm miễn dịch ở người (HIV) là thành viên của chi *Lentivirinae*, một phần của họ *Retroviridae*. Hai loài HIV lây nhiễm ở người: HIV-1 và HIV-2. HIV-1 là chủng phổ biến nhất của virus HIV, và được biết là gây bệnh nhiều hơn HIV-2. Như được sử dụng ở đây, thuật ngữ “virus gây suy giảm miễn dịch ở người” và “HIV” đề cập đến, nhưng không chỉ giới hạn ở, HIV-1 và HIV-2. Theo phương án được ưu tiên, HIV đề cập đến HIV-1.

HIV được chia thành nhiều nhánh có mức độ sai khác di truyền cao. Như được dùng trong bản mô tả này, thuật ngữ “nhánh HIV” hoặc “kiểu phụ HIV” dùng để chỉ virus gây suy giảm miễn dịch ở người liên quan được phân loại theo mức độ tương tự di truyền của chúng. Nhóm lớn nhất của thể phân lập HIV-1 được gọi là Nhóm M (chủng chính) và bao gồm ít nhất mười nhánh, A đến J.

Theo một khía cạnh chung, sáng chế đề cập đến protein vỏ (Env) HIV tái tổ hợp. Thuật ngữ “tái tổ hợp” khi được dùng viện dẫn đến protein đề cập đến protein mà được tạo ra bằng kỹ thuật tái tổ hợp hoặc bằng tổng hợp hóa học *trong ống nghiệm*. Theo các phương án của sáng chế, protein “tái tổ hợp” có trình tự axit amin nhân tạo trong đó nó chứa ít nhất một chi tiết trình tự (ví dụ, thế, loại, bổ sung, thay thế trình tự axit amin, v.v..) mà không thấy ở trình tự tương đương có trong tự nhiên. Tốt hơn là, protein “tái tổ hợp” là protein vỏ HIV không có trong tự nhiên được tối ưu hóa để gây ra đáp ứng miễn dịch hoặc tạo ra sự miễn dịch chống lại một hoặc nhiều chủng HIV có trong tự nhiên.

Thuật ngữ “protein vỏ HIV,” “Env HIV,” và “protein Env HIV” đề cập đến protein, hoặc đoạn hoặc dẫn xuất của chúng, mà về bản chất được biểu hiện trên vỏ của các hạt virus HIV và cho phép HIV nhắm đích và gắn vào màng huyết tương của tế bào nhiễm HIV. Thuật ngữ “vỏ” và “Env” được sử dụng tráo đổi trong toàn bộ bản mô tả. Gen *env* HIV mã hóa protein gp160 tiền chất, mà được phân cắt bằng cách phân giải protein thành hai glycoprotein vỏ trưởng thành gp120 và gp41. Phản ứng phân cắt được trung gian bằng proteaza tế bào chủ, furin (hoặc bằng protease tương tự furin), ở mô típ trình tự được bảo toàn cao trong tiền chất glycoprotein vỏ retrovirus. Cụ thể hơn, gp160 trime hóa thành (gp160)₃ và sau đó trải qua phân cắt thành hai glycoprotein trưởng thành được kết hợp không cộng hóa trị gp120 và gp41. Sự xâm nhập virus sau đó được điều

khiến bằng trime của dime khác loại gp120/gp41. Gp120 là đoạn liên kết thụ thể, và liên kết với thụ thể CD4 (và đồng thụ thể) trên tế bào đích mà có thụ thể này, như, ví dụ, tế bào T-trợ giúp. Gp41, được liên kết không cộng hóa trị với gp120, là mảnh dung hợp và cung cấp bước thứ hai nhờ đó HIV đi vào tế bào. Gp41 ban đầu được che giấu trong vỏ virut, nhưng khi gp120 liên kết với thụ thể CD4 và đồng thụ thể, gp120 thay đổi hình dạng của nó làm cho gp41 trở nên được lộ ra, khi đó nó có thể giúp đỡ trong việc dung hợp với tế bào chủ. Gp140 là vùng ngoại bào của gp160.

Theo các phương án của sáng chế, “protein vỏ (Env) HIV” có thể là protein gp160 hoặc gp140, hoặc dạng kết hợp, dung hợp, dạng cắt cụt hoặc dẫn xuất của chúng. Ví dụ, “protein vỏ HIV” có thể bao gồm protein gp120 được kết hợp không cộng hóa trị với protein gp41. “Protein vỏ HIV” cũng có thể là protein vỏ HIV bị cắt cụt bao gồm, nhưng không giới hạn, protein vỏ có chứa phần cắt cụt đầu tận cùng C trong vùng ngoại bào (tức là miền mà kéo dài vào không gian ngoại bào), phần cắt cụt trong gp41, như phần cắt cụt trong vùng ngoại bào của gp41, trong vùng xuyên màng của gp41, hoặc phần cắt cụt trong vùng tế bào chất của gp41. Protein vỏ HIV còn có thể là gp140, tương ứng với vùng ngoại bào gp160, hoặc phiên bản mở rộng hoặc cắt cụt của gp140. Biểu hiện của protein gp140 đã được mô tả trong một vài công bố (ví dụ Zhang et al., 2001; Sanders et al., 2002; Harris et al., 2011), và protein có thể còn được đặt từ các nhà cung cấp dịch vụ, ở các biến thể khác nhau ví dụ dựa trên chủng HIV khác nhau. Protein gp140 theo sáng chế có thể có đột biến vị trí phân cắt sao cho vùng gp120 và vùng ngoại bào gp41 không bị phân cắt và liên kết cộng hóa trị, hoặc tùy ý vùng gp120 và vùng ngoại bào gp41 có thể bị phân cắt và liên kết không cộng hóa trị, ví dụ bằng cầu disulphua (như ví dụ ở các biến thể SOSIP). “Protein vỏ HIV” có thể còn là dẫn xuất của protein vỏ HIV có trong tự nhiên có đột biến trình tự, ví dụ như, ở vị trí phân cắt furin, và/hoặc còn được gọi là đột biến SOSIP. Protein vỏ HIV theo sáng chế có thể cũng có vị trí phân cắt sao cho vùng ngoại bào gp120 và gp41 có thể được liên kết không cộng hóa trị.

Theo phương án được ưu tiên của sáng chế, protein Env HIV là protein gp140 hoặc protein gp160, và tốt hơn nữa là protein gp140. Theo phương án được ưu tiên khác protein Env bị cắt cụt, ví dụ bằng cách loại bỏ gốc sau gốc thứ 7 của vùng tế bào chất so với protein Env tự nhiên.

Theo các phương án của sáng chế, “protein vỏ HIV” có thể là trime hoặc monome, và tốt hơn là trime. Trime có thể là homotrime (ví dụ, trime bao gồm ba đơn vị polypeptit tương đồng), hoặc heterotrime (ví dụ, các trime bao gồm ba ddowdown vj polypeptit mà tất cả đều không giống nhau). Tốt hơn là, trime là homotrime. Trong trường hợp gp140 hoặc gp160 bị phân cắt, nó là trime của đơn vị polypeptit mà là dime gp120-gp41, và trong trường hợp tất cả ba trong số các dime này giống nhau, nó được xem là homotrime.

“Protein vỏ HIV” có thể là protein tan được, hoặc protein liên kết màng tế bào. Protein vỏ liên kết màng tế bào thường bao gồm vùng xuyên màng, như ở protein vỏ HIV chiều dài đầy đủ bao gồm vùng xuyên màng (TM) như được thể hiện trong Hình 1A. Protein liên kết màng có thể có vùng tế bào chất, nhưng không cần vùng tế bào chất được liên kết màng. Protein vỏ tan được bao gồm ít nhất sự loại bỏ một phần hoặc toàn bộ vùng xuyên màng. Ví dụ, đầu tận cùng C của protein vỏ HIV chiều dài đầy đủ có thể được cắt cụt để loại bỏ vùng xuyên màng, nhờ đó tạo ra protein tan được, như được thể hiện trong Hình 1B. Tuy nhiên, protein vỏ HIV còn có thể tan được với các vị trí cắt cụt ngắn và vị trí cắt cụt tùy ý với loại được thể hiện trong Hình 1B. Cắt cụt có thể được thực hiện ở nhiều vị trí khác nhau, và các ví dụ không giới hạn là sau axit amin 664, 655, 683, v.v. mà tất cả tạo ra protein tan được. Protein Env liên kết màng theo sáng chế có thể bao gồm đầu tận cùng C toàn bộ hoặc một phần (ví dụ bằng cách loại bỏ một phần vùng tế bào chất đầu C, ví dụ theo các phương án nhất định sau gốc thứ 7 của vùng tế bào chất) so với protein Env tự nhiên.

Peptit tín hiệu thường có mặt ở đầu N của protein Env HIV khi được biểu hiện, nhưng bị phân cắt đi bằng peptidaza tín hiệu và do đó không có mặt trong protein trưởng thành. Peptit tín hiệu có thể được trao đổi với các trình tự tín hiệu khác, và một số ví dụ không giới hạn của peptit tín hiệu được đề xuất ở đây trong SEQ ID NO: 11, 18, 33, và 34.

Theo các phương án của sáng chế, protein vỏ HIV, ví dụ, gp160, hoặc gp140, có thể là có nguồn gốc từ trình tự protein vỏ HIV từ bất kỳ HIV nhánh (hoặc ‘kiểu phụ’), ví dụ, nhánh A, nhánh B, nhánh C, nhánh D, nhánh E, nhánh F, nhánh G, nhánh H, v.v., hoặc kết hợp của chúng (như ở ‘dạng tái tổ hợp tuần hoàn’ hoặc CRF có nguồn gốc từ sự tái tổ hợp giữa các virut của kiểu phụ khác nhau, ví dụ BC, AE, AG, BE, BF, ADG, v.v.). protein vỏ HIV sequence có thể là trình tự có trong tự nhiên, trình tự khảm, trình

tự phổ biến, trình tự tổng hợp, hoặc bất kỳ dẫn xuất hoặc đoạn của chúng. “Trình tự khảm” chứa nhiều epitop có nguồn gốc từ ít nhất ba trình tự vỏ HIV của một hoặc nhiều nhánh HIV, và có thể được thiết kế bởi thuật toán mà tối ưu hóa sự bao phủ của epitop tế bào T. Các ví dụ của trình tự của protein vỏ HIV khảm bao gồm các trình tự được mô tả trong, ví dụ, Barouch et al, *Nat Med* 2010, 16: 319-323; và WO 2010/059732, như ví dụ các trình tự được thể hiện trong SEQ ID NO: 8 và 9. Như được sử dụng ở đây “trình tự phổ biến” nghĩa là trình tự nhân tạo của axit amin dựa trên sự sắp thẳng hàng trình tự axit amin của protein tương đồng, ví dụ như được xác định bởi sự sắp thẳng hàng (ví dụ sử dụng Clustal Omega) của trình tự axit amin của protein tương đồng. Nó là thứ tự được tính toán của gốc axit amin thường xuyên nhất, thấy ở mỗi vị trí trong sự sắp thẳng hàng trình tự, dựa trên các trình tự của Env từ ít nhất 1000 thể phân lập HIV tự nhiên. “Trình tự tổng hợp” là protein vỏ HIV không có trong tự nhiên mà được tối ưu hóa để gây đáp ứng miễn dịch hoặc tạo ra sự miễn dịch chống lại nhiều hơn một chủng HIV có trong tự nhiên. Protein vỏ HIV khảm là các ví dụ không giới hạn của protein vỏ HIV tổng hợp. Theo phương án được ưu tiên của sáng chế, protein Env của HIV là protein Env phổ biến, hoặc protein Env tổng hợp, có ít nhất một trong số các axit amin được nêu ở các vị trí được nêu (i)-(vii) theo sáng chế. Được đặc biệt ưu tiên là protein Env phổ biến có ít nhất một, tốt hơn là ít nhất hai trong số gốc axit amin được nêu ở các vị trí được nêu (i)-(vii) theo sáng chế, tốt hơn là có thêm SOSIP và/hoặc đột biến ở vị trí phân cắt furin như được mô tả dưới đây.

Theo các phương án nhất định theo sáng chế, protein vỏ HIV, bất kể trình tự có trong tự nhiên, trình tự khảm, trình tự phổ biến, trình tự tổng hợp v.v., bao gồm các đột biến trình tự bổ sung ví dụ, ở vị trí phân cắt furin, và/hoặc cái gọi là đột biến SOSIP.

Theo một số phương án theo sáng chế, protein vỏ HIV là "protein Env HIV đột biến SOSIP." Cái gọi là đột biến SOSIP là đột biến làm ổn định trimer mà bao gồm ‘các đột biến SOS’ (gốc Cys ở các vị trí 501 và 605, mà tạo ra việc đưa vào cầu disulfua có thể giữa các gốc cystein mới tạo ra) và ‘đột biến IP’ (gốc Pro ở vị trí 559). Theo các phương án của sáng chế, protein Env thể đột biến SOSIP bao gồm ít nhất một đột biến được chọn từ nhóm bao gồm Cys ở các vị trí 501 và 605; Pro ở vị trí 559; và tốt hơn là Cys ở các vị trí 501 và 605 và Pro ở vị trí 559. Protein Env HIV đột biến SOSIP có thể còn bao gồm các đột biến trình tự khác, ví dụ, ở vị trí phân cắt furin. Ngoài ra, theo các phương án nhất định có thể bổ sung thêm các đột biến sao cho protein Env bao gồm Pro

ở vị trí 556 hoặc vị trí 558 hoặc ở các vị trí 556 và 558, thấy ở đây là có khả năng không chỉ đóng vai trò làm các thay thế cho Pro ở vị trí 559 trong biến thể SOSIP, mà còn làm các đột biến bổ sung mà có thể còn cải thiện sự tạo thành trimer của biến thể SOSIP mà đã có Pro ở vị trí 559.

Theo các phương án được ưu tiên nhất định của sáng chế, protein Env HIV đột biến SOSIP bao gồm Cys ở các vị trí 501 và 605, và Pro ở vị trí 559.

Theo các phương án nhất định, protein vỏ HIV theo sáng chế còn bao gồm đột biến ở vị trí phân cắt furin. Đột biến ở trình tự phân cắt furin có thể là sự thế axit amin, loại bỏ, chèn, hoặc thay thế một trình tự bằng trình tự khác, hoặc thay thế bằng trình tự axit amin khác. Tốt hơn là theo sáng chế, việc tạo đột biến vị trí phân cắt furin có thể được dùng để tối ưu hóa vị trí phân cắt, sao cho sự phân cắt furin được cải thiện so với kiểu đại, ví dụ bằng sự thay thế trình tự ở gốc 508-511 bằng RRRRRR (SEQ ID NO: 10) [tức là thay thế trình tự axit amin chính (ví dụ EK) ở các vị trí 509-510 bằng bốn gốc arginin (tức là hai thay thế và hai thêm), trong khi đó ở các vị trí 508 và 511, đã có gốc arginin có mặt trong hầu hết protein Env của HIV, sao cho chúng thường không cần thay thế, nhưng vì kết quả cuối trong tài liệu tham khảo thường đề cập đến là trình tự axit amin RRRRRR, tác giả giữ danh pháp này ở đây]. Các đột biến khác mà cải thiện phân cắt furin là đã biết và có thể cũng được sử dụng. Tùy ý, có thể thay thế vị trí phân cắt furin bằng liên kết, sao cho phân cắt furin không còn cần nhưng protein sẽ thông qua cấu hình giống tự nhiên (ví dụ được mô tả trong (Sharma et al, 2015) và (Georgiev et al, 2015)).

Theo phương án cụ thể của sáng chế, protein vỏ HIV theo sáng chế còn bao gồm cả cái gọi là đột biến SOSIP (tốt hơn là Cys ở các vị trí 501 và 605, và Pro ở vị trí 559) và trình tự đột biến ở vị trí phân cắt furin, tốt hơn là thay thế trình tự ở gốc 508-511 bằng RRRRRR (SEQ ID NO: 10). Theo các phương án được ưu tiên nhất định, Env HIV bao gồm cả đột biến SOSIP được nêu và đột biến ở vị trí phân cắt furin, và ngoài ra còn bao gồm gốc Pro ở vị trí 556 hoặc 558, tốt nhất là ở cả vị trí 556 và 558.

Theo phương án được ưu tiên của sáng chế, trình tự axit amin của protein vỏ HIV là trình tự phổ biến, như phổ biến nhánh C vỏ HIV hoặc phổ biến nhánh B vỏ HIV. Theo phương án được ưu tiên đặc biệt, trình tự axit amin của protein vỏ HIV là phổ biến nhánh C vỏ HIV.

Protein vỏ HIV ví dụ mà có thể được dùng trong sáng chế bao gồm đồng nhất nhánh C vỏ HIV (SEQ ID NO: 2) và phổ biến nhánh B vỏ HIV (SEQ ID NO: 4). Các trình tự phổ biến nhánh C và nhánh B vỏ HIV có thể bao gồm các đột biến bổ sung mà, ví dụ, tăng cường tính ổn định và/hoặc sự tạo thành trimer, như ví dụ cái gọi là đột biến SOSIP và/hoặc trình tự đột biến ở vị trí phân cắt furin như được mô tả ở trên, như ví dụ trong trình tự ConC_SOSIP được thể hiện trong SEQ ID NO: 3 và trình tự chuỗi nặng nêu trong SEQ ID NO: 5.

Các ví dụ không giới hạn khác của trình tự protein vỏ HIV được ưu tiên mà có thể được dùng trong sáng chế (là phân tử ‘cơ sở’ hoặc ‘bố mẹ’, trong đó sau đó một hoặc nhiều trong số các đột biến theo sáng chế được đưa vào) bao gồm protein Env của HIV tổng hợp, ví dụ bao gồm trình tự axit amin nêu ở SEQ ID NO: 6 hoặc SEQ ID NO: 6 với đột biến Glu thành Arg ở vị trí 166, hoặc là đột biến tùy ý có đột biến SOSIP khác và/hoặc đột biến ở vị trí phân cắt furin như được mô tả ở trên. Ví dụ không giới hạn là SEQ ID NO: 7. Các ví dụ không giới hạn khác là protein vỏ HIV khảm, như chúng có trình tự axit amin nêu ở SEQ ID NO: 8 hoặc 9.

Theo các phương án nhất định, phân tử bố mẹ là protein Env HIV kiểu đại, trong đó một hoặc tốt hơn là nhiều axit amin được sửa theo phương pháp được mô tả ở đây. Phân tử bố mẹ này bao gồm ít nhất một đột biến sửa ở gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất ít hơn 7,5%, tốt hơn là ít hơn 2%, của trình tự Env HIV trong tập hợp của ít nhất 100, tốt hơn là ít nhất 500, tốt hơn là ít nhất 1000, tốt hơn là ít nhất 10000, tốt hơn là ít nhất 20000, trình tự Env HIV kiểu đại, trong đó đột biến sửa là sự thế bằng gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất bằng ít nhất 10% của trình tự Env HIV trong tập hợp này. Tốt hơn là sự thế này là bằng gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất bằng ít nhất 15%, ít nhất 20%, ít nhất 25%, của trình tự Env HIV trong tập hợp này. Tốt hơn là, sự thế này là bằng gốc axit amin mà có mặt ở vị trí tương ứng thường xuyên nhất trong tập hợp này. Theo các phương án được ưu tiên nhất định, các phân tử bố mẹ này bao gồm ít nhất 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, hoặc ít nhất 20 trong số đột biến sửa đó. Tốt hơn là ít nhất 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% hoặc tất cả trong số gốc axit amin mà có mặt ở các vị trí tương ứng ở tần suất nhỏ hơn 2% của trình tự Env HIV trong tập hợp này được sửa ở phân tử bố mẹ so với protein Env kiểu đại, theo các phương án nhất định, ít nhất 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% hoặc tất cả trong số gốc

axit amin mà có mặt ở các vị trí tương ứng ở tần suất nhỏ hơn 7,5% của trình tự Env HIV trong tập hợp này được sửa ở phân tử bố mẹ so với protein Env kiểu đại. Theo các phương án nhất định, protein Env HIV kiểu đại là từ chủng nhánh A, B, hoặc C, tốt hơn là từ chủng nhánh C. Kết quả của đột biến sửa này là, phân tử bố mẹ sẽ thể hiện giống nhiều hơn với trình tự phổ biến Env HIV so với chủng kiểu đại gốc, do đó gốc axit amin được sửa đôi khi được gọi ở đây là 'axit amin phổ biến' hoặc 'gốc phổ biến'. Các kết quả của hoạt tính sửa này là các đặc điểm tăng cường lớn hơn của phân tử bố mẹ tạo thành về sự gấp nếp, sự trime hóa, sự biểu hiện, và/hoặc tính ổn định, và phân tử tạo thành ở đây gọi là 'protein Env đã sửa'. Sự bổ sung đột biến làm ổn định theo sáng chế (ví dụ một hoặc nhiều trong số (i)-(vii) (Bảng 1), và/hoặc (xvi) (Bảng 1), và/hoặc tùy ý (viii)-(xv) (Bảng 2), vào phân tử bố mẹ này để cải thiện hơn nữa về một hoặc nhiều trong số tỷ lệ phần trăm trime, hiệu suất trime, tính ổn định, liên kết kháng thể trung hòa rộng, sự gấp nếp, và phân tử tạo thành mà có nguồn gốc từ protein Env của HIV kiểu đại ở đây gọi là 'protein Env được sửa và được làm ổn định'. Sẽ rõ ràng với chuyên gia trong lĩnh vực là việc đưa vào đột biến làm ổn định chuyển đổi thực sự trình tự tạo thành một chút từ trình tự phổ biến, sao cho kết quả thu được của đặc tính tăng cường lớn của phân tử Env HIV được sửa và được làm ổn định là dựa trên hai khái niệm hoàn toàn khác nhau.

Các đột biến tạo ra axit amin được nêu ở các vị trí (i)-(vii) theo sáng chế có thể còn được dùng trong protein Env của HIV trong đó không có các đột biến SOSIP có mặt (ví dụ trong trình tự phổ biến Env hoặc trong protein Env từ thể phân lập HIV kiểu đại) và có khả năng còn cải thiện sự trime hóa của chúng, vì các đột biến theo sáng chế độc lập với các đột biến SOSIP, và ngoài ra thể hiện sự làm việc trong một vài mạch chính protein Env HIV khác nhau. Thật vậy, ở đây thể hiện rằng các đột biến theo sáng chế có thể làm việc trong sự vắng mặt của các đột biến SOS cũng như trong sự vắng mặt của đột biến IP để cải thiện các đặc điểm trime hóa Env HIV.

Protein vỏ HIV tái tổ hợp theo các phương án của sáng chế bao gồm protein vỏ HIV có (các) gốc axit amin nhất định ở các vị trí cụ thể trong trình tự axit amin của protein vỏ HIV. Cụ thể là, bảy vị trí trong protein vỏ được xác định, cũng như gốc axit amin cụ thể được mong muốn ở mỗi trong số các vị trí được xác định. Các vị trí được xác định trong trình tự protein vỏ bao gồm (i) vị trí 651, (ii) vị trí 655, (iii) vị trí 535, (iv) vị trí 589, (v) vị trí 573, (vi) vị trí 204, và (vii) vị trí 647, trong đó việc đánh số các

vị trí là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1. Protein Env HIV theo sáng chế có (các) gốc axit amin đặc hiệu trong ít nhất một trong các vị trí được nêu (i)-(vii), tốt hơn là ở ít nhất hai trong số vị trí được nêu (i)-(vii), tốt hơn nữa là ở ít nhất ba trong số vị trí được nêu (i)-(vii). Gốc axit amin cụ thể mà được mong muốn là ở mỗi trong số các vị trí được xác định theo các phương án của sáng chế được thể hiện trong Bảng 1. Các vị trí được ưu tiên của các tùy chọn này là (i), (ii), (iii), (iv), (vi), và/hoặc (vii). Các vị trí được đặc biệt ưu tiên của các tùy chọn này là (i), (ii), (iii), (iv), và/hoặc (vii). Các tùy chọn được ưu tiên khác là (xvi), được đề cập chi tiết hơn sau đây trong bản mô tả.

Bảng 1: Axit amin mong muốn ở vị trí được nêu trong protein Env của HIV tái tổ hợp theo các phương án của sáng chế

Số	Vị trí ¹	Gốc axit amin mong muốn
(i)	651	Phe, Leu, Met, hoặc Trp (tốt hơn là Phe)
(ii)	655	Phe, Ile, Met, hoặc Trp (tốt hơn là Ile)
(iii)	535	Asn hoặc Gln (tốt hơn là Asn)
(iv)	589	Val, Ile, hoặc Ala (tốt hơn là Val hoặc Ile, tốt nhất là Val)
(v)	573	Phe hoặc Trp (tốt hơn là Phe)
(vi)	204	Ile
(vii)	647	Phe, Met, hoặc Ile (tốt hơn là Phe)
(xvi)	658	Val, Ile, Phe, Met, Ala, hoặc Leu (tốt hơn là Val hoặc Ile, tốt nhất là Val)

¹ Theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1

Trình tự axit amin của protein vỏ HIV mà một hoặc nhiều sự thể axit amin mong muốn (hoặc axit amin được nêu) ở một hoặc nhiều vị trí được nêu được đưa vào đó được gọi là “trình tự vỏ HIV mạch chính” hoặc “trình tự vỏ HIV bố mẹ.” Ví dụ, nếu vị trí 651 trong trình tự ConC_SOSIP nêu ở SEQ ID NO: 3 được đột biến thành Phe, sau đó trình tự ConC_SOSIP được xem là trình tự “mạch chính” hoặc “bố mẹ”. Bất kỳ protein vỏ của HIV có thể được dùng làm trình tự “mạch chính” hoặc “bố mẹ” mà đột biến làm ổn định mới theo phương án theo sáng chế có thể được đưa vào đó, hoặc là một mình hoặc kết hợp với các đột biến khác, như cái gọi là đột biến SOSIP và/hoặc các đột biến

ở vị trí phân cắt furin. Các ví dụ không giới hạn của protein Env HIV mà có thể được dùng là mạch chính bao gồm protein Env của HIV từ thể phân lập HIV tự nhiên, protein Env của HIV tổng hợp, hoặc protein Env HIV phổ biến, và trong các ví dụ không giới hạn nhất định bao gồm loại bao gồm SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8 hoặc SEQ ID NO: 9.

Theo các phương án của sáng chế, protein vỏ HIV có thể có gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm các vị trí 651, 655, 535, 589, 573, 204, và 647, như gốc axit amin được nêu ở một, hai, ba, bốn, năm, sáu, hoặc bảy vị trí. Tốt hơn là, protein vỏ HIV được thể ở một, hai hoặc ba trong số vị trí được nêu, và tốt hơn nữa là protein vỏ HIV được thể ở ít nhất hai trong số vị trí được nêu. Còn tốt hơn nữa là, protein Env HIV được thể ở ba trong số vị trí được nêu, bốn trong số vị trí được nêu, năm trong số vị trí được nêu, sáu trong số vị trí được nêu, hoặc tất cả bảy trong số vị trí được nêu. Tốt hơn là, protein vỏ HIV chứa gốc axit amin được nêu ở ít nhất hai trong số vị trí được nêu. Tốt hơn nữa là, protein vỏ HIV chứa gốc axit amin được nêu ở ba trong số vị trí được nêu. Theo phương án được ưu tiên khác, protein vỏ HIV chứa gốc axit amin được nêu ở bốn, năm, sáu, hoặc tất cả bảy trong số vị trí được nêu.

Phương án của protein Env của HIV có axit amin được nêu ở nhiều vị trí là (các vị trí được đánh số theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1 sau đó là mã axit amin một chữ cái cho gốc có mặt trên vị trí đó, các vị trí trong một phương án protein Env của HIV được phân chia bằng dấu phẩy [ví dụ phương án của protein Env có Phe ở vị trí 651 và Ile ở vị trí 655 được mô tả là 651F, 655I], trong khi đó các phương án khác (tức là protein Env của HIV khác) được phân chia bởi dấu chấm phẩy) bao gồm sau đây.

Đối với các protein Env với axit amin được nêu ở hai vị trí: 651F, 655I; 651F, 655F; 651F, 655M; 651F, 655W; 651F, 535N; 651F, 535Q; 651F, 589V; 651F, 589I; 651F, 589A; 651F, 573F; 651F, 573W; 651F, 204I; 651F, 647F; 651F, 647I; 651F, 647M; 651L, 655I; 651L, 655F; 651L, 655M; 651L, 655W; 651L, 535N; 651L, 535Q; 651L, 589V; 651L, 589I; 651L, 589A; 651L, 573F; 651L, 573W; 651L, 204I; 651L, 647F; 651L, 647I; 651L, 647M; 651M, 655I; 651M, 655F; 651M, 655A; 651M, 655L; 651M, 655W; 651M, 535N; 651M, 535Q; 651M, 589V; 651M, 589I; 651M, 589A; 651M, 573F; 651M, 573W; 651M, 204I; 651M, 647F; 651M, 647I; 651M, 647M;

651W, 655I; 651W, 655F; 651W, 655M; 651W, 655W; 651W, 535N; 651W, 535Q; 651W, 589V; 651W, 589I; 651W, 589A; 651W, 573F; 651W, 573W; 651W, 204I; 651W, 647F; 651W, 647I; 651W, 647M; 655I, 535N; 655I, 535Q; 655I, 589V; 655I, 589I; 655I, 589A; 655I, 573F; 655I, 573W; 655I, 204I; 655I, 647F; 655I, 647I; 655I, 647M; 655F, 535N; 655F, 535Q; 655F, 589V; 655F, 589I; 655F, 589A; 655F, 573F; 655F, 573W; 655F, 204I; 655F, 647F; 655F, 647I; 655F, 647M; 655M, 535N; 655M, 535Q; 655M, 589V; 655M, 589I; 655M, 589A; 655M, 573F; 655M, 573W; 655M, 204I; 655M, 647F; 655M, 647I; 655M, 647M; 655W, 535N; 655W, 535Q; 655W, 589V; 655W, 589I; 655W, 589A; 655W, 573F; 655W, 573W; 655W, 204I; 655W, 647F; 655W, 647I; 655W, 647M; 535N, 589V; 535N, 589I; 535N, 589A; 535N, 573F; 535N, 573W; 535N, 204I; 535N, 647F; 535N, 647I; 535N, 647M; 535Q, 589V; 535Q, 589I; 535Q, 589A; 535Q, 573F; 535Q, 573W; 535Q, 204I; 535Q, 647F; 535Q, 647I; 535Q, 647M; 589V, 573F; 589V, 573W; 589V, 204I; 589V, 647F; 589V, 647I; 589V, 647M; 589I, 573F; 589I, 573W; 589I, 204I; 589I, 647F; 589I, 647I; 589I, 647M; 589A, 573F; 589A, 573W; 589A, 204I; 589A, 647F; 589A, 647I; 589A, 647M; 573F, 204I; 573F, 647F; 573F, 647I; 573F, 647M; 573W, 204I; 573W, 647F; 573W, 647I; 573W, 647M; 204I, 647F; 204I, 647I; 201I, 647M. Mỗi trong số các phương án đó có thể có mặt theo sáng chế trong bất kỳ trình tự Env của HIV, như thể phân lập kiểu dại, hoặc protein Env HIV đột biến SOSIP, hoặc protein Env HIV phổ biến, hoặc protein Env HIV tổng hợp. Mỗi trong số các phương án đó có thể được kết hợp với một trong số các axit amin được ưu tiên theo sáng chế ở vị trí thứ ba của một trong số các vị trí được nêu khác từ (i)-(vii) theo sáng chế. Các phương án này, có gốc axit amin được ưu tiên ở ba vị trí của các vị trí được nêu (i)-(vii) có thể được kết hợp với một trong số các axit amin được ưu tiên ở vị trí thứ tư của một trong số các vị trí được nêu khác từ (i)-(vii) theo sáng chế. Các phương án này, có gốc axit amin được ưu tiên ở bốn vị trí của các vị trí được nêu (i)-(vii) có thể được kết hợp với một trong số các axit amin được ưu tiên ở vị trí thứ năm của một trong số các vị trí được nêu khác từ (i)-(vii) theo sáng chế. Các phương án này, có gốc axit amin được ưu tiên ở năm vị trí của các vị trí được nêu (i)-(vii) có thể được kết hợp với một trong số các axit amin được ưu tiên ở vị trí thứ sáu của một trong số các vị trí được nêu khác từ (i)-(vii) theo sáng chế. Các phương án này, có gốc axit amin được ưu tiên ở sáu vị trí của các vị trí được nêu (i)-(vii) có thể được kết hợp với một trong số các axit amin được ưu tiên ở vị trí thứ bảy của một trong số các vị trí được nêu

khác từ (i)-(vii) theo sáng chế, sao cho protein Env có axit amin được ưu tiên theo sáng chế ở tất cả bảy vị trí (i)-(vii) theo sáng chế. Bất kỳ trong số các phương án khác này có axit amin được ưu tiên theo sáng chế ở ba, bốn, năm, sáu hoặc bảy trong số các vị trí (v)-(vii) theo sáng chế, có thể có mặt ở bất kỳ protein Env của HIV, như từ thể phân lập kiểu đại, biến thể SOSIP, protein Env HIV phổ biến, protein Env HIV tổng hợp, và tương tự.

Các protein Env được ưu tiên theo sáng chế với axit amin được nêu ở hai vị trí là: 651F, 655I; 651F, 535N; 651F, 589V; 651F, 589I; 651F, 573F; 651F, 204I; 651F, 647F; 655I, 535N; 655I, 589V; 655I, 589I; 655I, 573F; 655I, 204I; 655I, 647F; 535N, 589V; 535N, 589I; 535N, 573F; 535N, 204I; 535N, 647F ; 589V, 573F; 589V, 204I; 589V, 647F; 589I, 573F; 589I, 204I; 589I, 647F; 573F, 204I; 573F, 647F; 204I, 647F. Các protein Env được đặc biệt ưu tiên có axit amin được ưu tiên ở ít nhất hai vị trí theo sáng chế bao gồm: 651F, 655I; 655I, 535N; 655I, 589V; 535N, 589V; 535N, 647F.

Một số protein Env của HIV được ưu tiên có gốc axit amin được ưu tiên ở ba vị trí là: 651F, 655I, 535N; 651F, 589V, 535N; 651F, 589I, 535N; 651F, 573F, 535N; 651F, 204I, 535N; 651F, 647F, 535N; 655I, 589V, 535N; 655I, 589I, 535N; 655I, 573F, 535N; 655I, 204I, 535N; 655I, 647F, 535N; 589V, 573F, 535N; 589V, 204I, 535N; 589V, 647F, 535N; 589I, 573F, 535N; 589I, 204I, 535N; 589I, 647F, 535N; 573F, 204I, 535N; 573F, 647F, 535N; 204I, 647F, 535N; 651F, 655I, 589V; 651F, 573F, 589V; 651F, 204I, 589V; 651F, 647F, 589V; 655I, 573F, 589V; 655I, 204I, 589V; 655I, 647F, 589V; 573F, 204I, 589V; 573F, 647F, 589V; 204I, 647F, 589V; 651F, 655I, 589I; 651F, 573F, 589I; 651F, 204I, 589I; 651F, 647F, 589I; 655I, 573F, 589I; 655I, 204I, 589I; 655I, 647F, 589I; 573F, 204I, 589I; 573F, 647F, 589I; 204I, 647F, 589I; 651F, 655I, 573F; 651F, 204I, 573F; 651F, 647F, 573F; 655I, 204I, 573F; 655I, 647F, 573F; 204I, 647F, 573F; 651F, 655I, 204I; 651F, 647F, 204I; 655I, 647F, 204I; 651F, 655I, 647F; 655I, 651F, 647F; 655I, 651F, 535N; 655I, 589V, 573F; 655I, 589V, 204I. Các protein Env được đặc biệt ưu tiên có axit amin được ưu tiên ở ít

nhất ba vị trí theo sáng chế bao gồm: 651F, 655I, 535N; 655I, 589V, 535N; 655I, 573F, 589V; 655I, 204I, 589V; 651F, 655I, 647F.

Một số protein Env của HIV được ưu tiên có gốc axit amin được ưu tiên ở bốn vị trí là: 651F, 655I, 535N, 589V; 651F, 655I, 535N, 573F; 651F, 655I, 589V, 573F; 651F, 535N, 589V, 573F; 655I, 535N, 589V, 573F; 651F, 655I, 535N, 204I; 651F, 655I, 589V, 204I; 651F, 535N, 589V, 204I; 655I, 535N, 589V, 204I; 651F, 655I, 573F, 204I; 651F, 535N, 573F, 204I; 655I, 535N, 573F, 204I; 651F, 589V, 573F, 204I; 655I, 589V, 573F, 204I; 535N, 589V, 573F, 204I; 651F, 655I, 535N, 647F; 651F, 655I, 589V, 647F; 651F, 535N, 589V, 647F; 655I, 535N, 589V, 647F; 651F, 655I, 573F, 647F; 651F, 535N, 573F, 647F; 655I, 535N, 573F, 647F; 651F, 589V, 573F, 647F; 655I, 589V, 573F, 647F; 535N, 589V, 573F, 647F; 651F, 655I, 204I, 647F; 651F, 535N, 204I, 647F; 655I, 535N, 204I, 647F; 651F, 589V, 204I, 647F; 655I, 589V, 204I, 647F; 535N, 589V, 204I, 647F; 651F, 573F, 204I, 647F; 655I, 573F, 204I, 647F; 535N, 573F, 204I, 647F; và 589V, 573F, 204I, 647F.

Một số ví dụ của các protein Env của HIV được ưu tiên có gốc axit amin được ưu tiên ở ít nhất bốn vị trí là: 651F, 655I, 647F, 535N; 651F, 655I, 573F, 589V. Ví dụ được ưu tiên của protein Env của HIV bao gồm gốc axit amin được nêu ở ít nhất bốn vị trí bao gồm 535N, 589V, 651F, 655I. Các ví dụ không giới hạn của các protein Env của HIV này được đề xuất trong SEQ ID NO: 20, 22, 24, 26, 27, 28, 29, 30, 31, và 32. Tốt hơn là protein Env của HIV protein Env của HIV nhánh C hoặc protein Env của HIV nhánh A, tốt nhất là protein Env của HIV nhánh C. Theo các phương án nhất định, protein Env của HIV này còn bao gồm 588E, tức là ít bao gồm ít nhất 535N, 588E, 589V, 651F, 655I. Các ví dụ không giới hạn của such protein Env HIV được đề xuất trong SEQ ID NO: 20, 24, 26, 27, 28, 29, 30, 31, và 32. Theo các phương án nhất định, Env của HIV còn bao gồm 556P, tức là nó bao gồm ít nhất 535N, 556P, 589V, 651F, 655I hoặc ít nhất 535N, 556P, 588E, 589V, 651F, 655I. Các ví dụ không giới hạn của protein Env của HIV này được đề xuất trong SEQ ID NO: 22, 24, 26, 27, 29, 30, 31 và 32.

Theo một phương án, protein Env của HIV tái tổ hợp theo sáng chế bao gồm trình tự axit amin của protein Env HIV có gốc axit amin được nêu ở ít nhất hai trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp ở vị trí 655;

- (iii) Asn hoặc Gln ở vị trí 535;
- (iv) Val, Ile, hoặc Ala ở vị trí 589;
- (v) Phe hoặc Trp ở vị trí 573;
- (vi) Ile ở vị trí 204; và
- (vii) Phe, Met, hoặc Ile ở vị trí 647.

Ví dụ, protein Env HIV tái tổ hợp có thể có một trong số Phe, Leu, Met, hoặc Trp ở vị trí 651, và Asn hoặc Gln ở vị trí 535, tùy ý, gốc axit amin được nêu bổ sung ở vị trí ví được nêu bổ sung. Tốt hơn là, ít nhất một trong các axit amin in (i)-(vii) được đưa vào protein Env của HIV tái tổ hợp bằng sự thế axit amin. Ví dụ, protein Env HIV tái tổ hợp có thể được tạo ra từ protein Env của HIV mà không chứa hoặc chỉ chứa một trong số gốc axit amin tại (i)-(vii) trên sao cho tất cả hoặc một hoặc nhiều trong số ít nhất hai gốc axit amin được nêu được đưa vào protein Env HIV tái tổ hợp bằng sự thế axit amin.

Theo các phương án nhất định, protein Env của HIV tái tổ hợp theo sáng chế còn bao gồm (viii) Gln, Glu, Ile, Met, Val, Trp, hoặc Phe ở vị trí 588, trong đó Gln hoặc Glu được ưu tiên.

Trình tự axit amin của protein Env của HIV trong đó sự thế được mô tả ở trên được đưa vào có thể là protein Env của HIV bất kỳ đã biết trong lĩnh vực theo sáng chế, như, ví dụ trình tự có trong tự nhiên từ HIV nhánh A, nhánh B, nhánh C, v.v.; trình tự khám; trình tự phổ biến, ví dụ, trình tự phổ biến nhánh B hoặc nhánh C; trình tự tổng hợp; hoặc bất kỳ dẫn xuất hoặc đoạn của chúng. Theo các phương án nhất định theo sáng chế, trình tự axit amin của protein Env của HIV bao gồm đột biến bổ sung, như, ví dụ, cái gọi là đột biến SOSIP, và/hoặc đột biến ở vị trí phân cắt furin.

Theo một phương án cụ thể, protein mạch chính Env của HIV là protein Env của HIV đột biến SOSIP bao gồm ít nhất một đột biến được chọn từ nhóm bao gồm Cys ở các vị trí 501 và 605; Pro ở vị trí 559. Theo phương án được ưu tiên, protein Env của HIV đột biến SOSIP bao gồm Cys ở các vị trí 501 và 605, và Pro ở vị trí 559. Theo phương án này, protein Env của HIV tái tổ hợp bao gồm trình tự axit amin của protein Env HIV đột biến SOSIP và sự thế axit amin bằng gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp ở vị trí 655;
- (iii) Asn hoặc Gln ở vị trí 535;

- (iv) Val, Ile, hoặc Ala ở vị trí 589;
- (v) Phe hoặc Trp ở vị trí 573;
- (vi) Ile ở vị trí 204; và
- (vii) Phe, Met, hoặc Ile ở vị trí 647.

Protein Env HIV đột biến SOSIP có thể còn bao gồm đột biến ở vị trí phân cắt furin, như thay thế ở các vị trí 608-511 by SEQ ID NO: 10.

Theo phương án khác, protein Env của HIV tái tổ hợp theo sáng chế bao gồm trình tự axit amin của protein Env HIV và sự thế axit amin bằng gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp ở vị trí 655;
- (iii) Asn hoặc Gln ở vị trí 535;
- (iv) Val, Ile, hoặc Ala ở vị trí 589;
- (v) Phe hoặc Trp ở vị trí 573;
- (vi) Ile ở vị trí 204; và
- (vii) Phe, Met, hoặc Ile ở vị trí 647,

trong đó protein Env HIV được chọn từ nhóm bao gồm:

- (1) trình tự phổ biến Env của HIV, như trình tự phổ biến nhánh C hoặc nhánh B, ví dụ bao gồm trình tự axit amin nêu ở SEQ ID NO: 2, 3, 4, hoặc 5;
- (2) protein Env HIV tổng hợp, ví dụ có chứa trình tự axit amin nêu ở (a) SEQ ID NO: 6; (b) SEQ ID NO: 6 với đột biến Glu thành Arg ở vị trí 166; (c) SEQ ID NO: 7; (d) SEQ ID NO: 8 hoặc 9, (a), (b) hoặc (d) tùy ý còn có đột biến SOSIP và/hoặc đột biến ở vị trí phân cắt furin như được mô tả ở trên.

Tốt hơn là, protein Env của HIV tái tổ hợp bao gồm trình tự axit amin của protein Env của HIV và sự thế axit amin bằng gốc axit amin được nêu ở ít nhất hai trong số vị trí được nêu được chọn từ nhóm bao gồm (i)-(vii) ở trên, như hai vị trí hoặc ba vị trí. Tuy nhiên, protein Env của HIV tái tổ hợp có thể bao gồm sự thế axit amin bằng gốc axit amin được nêu ở một hoặc nhiều trong số vị trí được nêu, như một, hai, ba, bốn, năm, sáu, hoặc bảy trong số các vị trí được nêu.

Theo một phương án cụ thể, protein mạch chính Env của HIV là nhánh C phổ biến Env của HIV bao gồm trình tự axit amin mà là ít nhất 95%, 96%, 97%, 98%, 99% hoặc 100% tương đồng với trình tự axit amin nêu ở SEQ ID NO: 2. Tốt hơn là, trình tự

nhánh C phổ biến HIV nêu ở SEQ ID NO: 2 còn bao gồm cái gọi là đột biến SOSIP, tức là, Cys ở các vị trí 501 và 605, và Pro ở vị trí 559, và tốt hơn nữa là còn bao gồm cái gọi là đột biến SOSIP và đột biến ở vị trí phân cắt furin, như ví dụ thể ở các vị trí 508-511 bằng SEQ ID NO: 10. Theo phương án được ưu tiên đặc biệt, protein mạch chính Env của HIV bao gồm trình tự được thể hiện trong SEQ ID NO: 3, hoặc trình tự ít nhất 95% tương đồng với nó, trong đó tốt hơn là axit amin ở các vị trí 501, 559, 605, và 508-511 như được thay thế bởi SEQ ID NO: 10, không đột biến so với SEQ ID NO: 3.

Theo phương án được ưu tiên đặc biệt khác, protein mạch chính Env của HIV là nhánh B phổ biến Env của HIV bao gồm trình tự axit amin mà là ít nhất 95%, 96%, 97%, 98%, 99% hoặc 100% tương đồng với trình tự axit amin nêu ở SEQ ID NO: 4. Tốt hơn là, trình tự nhánh B phổ biến của HIV nêu ở SEQ ID NO: 4 còn bao gồm cái gọi là đột biến SOSIP, tức là, Cys ở các vị trí 501 và 605, và Pro ở vị trí 559, và tốt hơn nữa là còn bao gồm cái gọi là đột biến SOSIP và đột biến ở vị trí phân cắt furin, như ví dụ thể ở các vị trí 508-511 bằng SEQ ID NO: 10. Theo phương án được ưu tiên đặc biệt, protein mạch chính Env của HIV bao gồm trình tự được thể hiện trong SEQ ID NO: 5, hoặc trình tự ít nhất 95% tương đồng với nó, trong đó tốt hơn là axit amin ở các vị trí 501, 559, 605, và 508-511 như được thay thế bởi SEQ ID NO: 10, không đột biến so với SEQ ID NO: 5.

Theo một phương án khác nữa, protein mạch chính Env HIV là protein Env HIV tổng hợp, ví dụ có chứa trình tự axit amin nêu ở (a) SEQ ID NO: 6; (b) SEQ ID NO: 6 với đột biến Glu thành Arg ở vị trí 166; (c) SEQ ID NO: 7; hoặc (d) SEQ ID NO: 8 hoặc 9, (a), (b) hoặc (d) tùy ý còn có đột biến SOSIP (501C, 605C, 559P) và/hoặc đột biến ở vị trí phân cắt furin (508-511RRRRRR) như được mô tả ở trên.

Theo phương án cụ thể khác nữa, protein mạch chính Env của HIV là protein Env của HIV từ virus HIV nhánh A hoặc nhánh C kiểu đại, tùy ý bao gồm các đột biến để sửa trình tự theo phương pháp được mô tả ở đây.

Các kết hợp ví dụ của hai vị trí trong protein Env của HIV mà có thể được thể đồng thời bao gồm các gốc 535,589; 535,647; và 589,655; như ví dụ trong thể đột biến kép I535N, D589V; I535N, E647F; và D589V, K655I. Các thể đột biến kép khác bao gồm K655I, I535N; N651F, K655I; và K655I, I573F. Kết hợp ví dụ của ba vị trí trong protein Env của HIV mà có thể được thể đồng thời bao gồm 535,589,655, như ví dụ trong

thể đột biến ba I535N, D589V, K655I. Thể đột biến ba khác bao gồm K655I, D589V, I573F; và K655I, N651F, I535N.

Theo các phương án nhất định theo sáng chế, protein Env của HIV tái tổ hợp theo sáng chế có thể còn bao gồm gốc axit amin được nêu (ví dụ qua sự thể) ở một hoặc nhiều vị trí được nêu bổ sung được chọn từ nhóm bao gồm các vị trí (viii) 588, (ix) 64 hoặc 66, (x) 316, (xi) 201/433, (xii) 556 hoặc 558 hoặc 556 và 558, (xiii) 548-568, (xiv) 568, 569 và 636, hoặc (xv) 302, 519 hoặc 520, như được thể hiện trong Bảng 2 dưới đây. Một số trong các thể axit amin này (ví dụ (viii)) được thấy bởi tác giả sáng chế là kết hợp rất tốt với (kết hợp của) các đột biến (i)-(vii) theo sáng chế như được mô tả ở trên. Còn lại trong số các sự thể axit amin đã được báo cáo trước đây trong tài liệu tham khảo. Ví dụ, De Taeye et al. (*Cell* (2015) 163(7), 1702-15) báo cáo protein vỏ của HIV có đột biến đôi E64K và T316W, và protein Env của HIV có đột biến 66R; và Kwon et al. (*Nat. Struct. Mol. Biol.* (2015) 22(7) 522-31) báo cáo protein vỏ HIV có thể disulfua I204C, A433C; và Guenaga et al. (*Immunity* (2017) 46, 792-803) báo cáo protein vỏ của HIV có sự thể ba L568G, T569G hoặc N636G, và N302Y, F519R, L520R. Tuy nhiên, theo hiểu biết của tác giả, các đột biến đã được mô tả trước đây không được mô tả kết hợp với bất kỳ trong số các thể mới được mô tả ở đây, ví dụ, sự thể được liệt kê trong Bảng 1. Các đột biến axit amin này kết hợp với sự thể axit amin theo sáng chế có thể còn làm tăng hiệu suất trime và/hoặc tỷ lệ phần trăm tạo thành trime. Sự thể axit amin này có thể được đưa vào bất kỳ của các protein Env của HIV tái tổ hợp được mô tả ở đây ngoài sự thể bằng gốc axit amin được nêu ở một hoặc nhiều trong số vị trí được nêu như được mô tả trong Bảng 1.

Bảng 2: Các vị trí bổ sung của thể axit amin và gốc của sự thể

Số	Vị trí ¹	Gốc axit amin được nêu
(viii)	588	Gln, Glu, Ile, Met, Val, Trp, hoặc Phe (tốt hơn là Gln hoặc Glu)
(ix)	64 hoặc 66	Lys ở vị trí 64; hoặc Arg ở vị trí 66
(x)	316	Trp
(xi)	201 và 433	Cys ở cả vị trí
(xii)	556 hoặc 558 hoặc 556 và 558	Pro ở một trong hai hoặc cả hai vị trí

(xiii)	548-568 (vòng HR1)	Thay thế bởi vòng ngắn hơn và ít linh hoạt hơn có 7-10 axit amin, tốt hơn là vòng có 8 axit amin, ví dụ có trình tự được chọn từ một trình tự bất kỳ trong số (SEQ ID NO: 12-17)
(xiv)	568, 569, 636	Gly ở bất kỳ trong số các vị trí này, hoặc Gly ở cả vị trí 568 và 636, hoặc Gly ở cả vị trí 569 và 636
(xv)	302, 519, 520	Tyr ở vị trí 302, hoặc Arg ở vị trí 519, hoặc Arg ở vị trí 520; hoặc Tyr ở vị trí 302 và Arg ở vị trí 519; hoặc Tyr ở vị trí 302 và Arg ở vị trí 520; hoặc Tyr ở vị trí 302 và Arg ở cả vị trí 519 và 520

¹ Theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1

Sự thể được xác định ở các vị trí được nêu theo sáng chế [(i)-(vii), xem ví dụ Bảng 1] là không hoặc hiếm có mặt ở trình tự tự nhiên, đã thấy trong kết hợp trong trình tự protein Env của HIV được báo cáo trước đây, và không được đề xuất trước đây để tạo ra sự trime hóa của protein Env của HIV cải thiện, hiệu suất trime cải thiện và/hoặc tính ổn định trime tăng. Các đột biến (ix)-(xi) trong Bảng 2 (mà được báo cáo trước đây bởi người khác) là đều trong vùng gp120, mà kháng thể đặc hiệu trime PGT145 liên kết vào đó. Các đột biến này giữ trime cần ở đỉnh (mà là ở đỉnh của phân tử). Sự thể (xii) và (xiii) đều ở HR1 của gp41. Ngoại trừ vị trí 204, các đột biến của sáng chế trong Bảng 1 đều trong vùng gp41 (ở phần đáy của phân tử), nhưng bên ngoài vùng HR1. Rõ ràng là, các đột biến được mô tả trước đây không tạo ra gợi ý bất kỳ để đưa vào các đột biến theo sáng chế, hướng chỉ là các tác dụng đáng ngạc nhiên của chúng lên sự tạo thành trime với đỉnh gần như được đo bằng liên kết PGT145. Ngoài các đột biến điểm (viii)-(xii) trong Bảng 2, cũng có thể thay thế vòng HR1 của protein Env (gốc axit amin 548-568 trong trình tự kiểu đại, với đánh số theo gp160 của thể phân lập HXB2) bằng vòng ngắn hơn và ít linh hoạt hơn có 7-10 axit amin, tốt hơn là vòng có 8 axit amin, ví dụ có trình tự được chọn từ một trình tự bất kỳ trong số (SEQ ID NO: 12-17), xem ví dụ Kong et al (Nat Commun. 2016 Jun 28;7:12040. doi: 10.1038/ncomms12040) mà mô tả vòng ngắn hơn này thay thế vòng HR1. Biến thể Env này, còn có gốc axit amin được nêu ở ít nhất một và tốt hơn là ở ít nhất hai trong số các vị trí được nêu (i)-(vii) theo sáng chế, cũng là phương án theo sáng chế. Các đột biến được nêu ở (viii)-(xiii) có thể theo các phương án nhất định theo sáng chế được thêm vào protein Env của HIV theo sáng chế, tức là có một hoặc nhiều trong số axit amin được nêu ở các vị trí (i)-(vii). Cũng vậy, các kết hợp trong nhóm (viii)-(xiii) có thể được thực hiện, ví dụ không giới hạn là kết hợp

của các đột biến (ngoài ít nhất một đột biến trong số (i)-(vii)) ở (viii) và (xii) (ví dụ I535N, A556P, K588E). Một số ví dụ không giới hạn của thể đột biến đôi mà được thực hiện trong nền Env HIV với đột biến SOSIP và kết hợp các đột biến ở ít nhất một trong số các vị trí (i)-(vii) và ở ít nhất một trong số các vị trí (viii)-(xiii) bao gồm: 535,588; 588,589; 655,588; 558,535; và 655,556; như ví dụ I535N, K588E; 588Q, D589V; K655I, K588E; A558P, I535N; và K655I, L556P. Một số ví dụ không giới hạn của đột biến ba này bao gồm 558,535,588; 558,535,589; 558,535,655; và 558,535,651, như ví dụ A558P, I535N, K588E; A558P, I535, D589V; A558P, I535N, K655I; và A558P, I535N, N651F.

Các ví dụ không giới hạn khác của kết hợp theo sáng chế bao gồm 655I, 573F, 589V, 588E; 651F, 655I, 573F, 589V, 588E; 651F, 655I, 573F, 589V, 588E, 535N; 651F, 655I, 573F, 589V, 588E, 535N, 204I; 651F, 655I, 556P; 651F, 535N, 556P; 651F, 589V, 556P; 651F, 589I, 556P; 651F, 573F, 556P; 651F, 204I, 556P; 651F, 647F, 556P; 655I, 535N, 556P; 655I, 589V, 556P; 655I, 589I, 556P; 655I, 573F, 556P; 655I, 204I, 556P; 655I, 647F, 556P; 535N, 589V, 556P; 535N, 589I, 556P; 535N, 573F, 556P; 535N, 204I, 556P; 535N, 647F, 556P; 589V, 573F, 556P; 589V, 204I, 556P; 589V, 647F, 556P; 589I, 573F, 556P; 589I, 204I, 556P; 589I, 647F, 556P; 573F, 204I, 556P; 573F, 647F, 556P; 651F, 655I, 558P; 651F, 535N, 558P; 651F, 589V, 558P; 651F, 589I, 558P; 651F, 573F, 558P; 651F, 204I, 558P; 651F, 647F, 558P; 655I, 535N, 558P; 655I, 589V, 558P; 655I, 589I, 558P; 655I, 573F, 558P; 655I, 204I, 558P; 655I, 647F, 558P; 535N, 589V, 558P; 535N, 589I, 558P; 535N, 573F, 558P; 535N, 204I, 558P; 535N, 647F, 558P; 589V, 573F, 558P; 589V, 204I, 558P; 589V, 647F, 558P; 589I, 573F, 558P; 589I, 204I, 558P; 589I, 647F, 558P; 573F, 204I, 558P; 573F, 647F, 558P; 655I, 589V, 535N, 556P; 651F, 655I, 535N, 556P; 556P, 651F; 556P, 651F, 655I, 535N; 655I, 589V, 573F, 651F, 588E, 556P; 556P, 651F, 655I, 535N, 573F; 556P, 651F, 655I, 535N, 573F, 589V; 556P, 651F, 655I, 535N, 573F, 589V, 204I; 556P, 651F, 655I, 535N, 573F, 589V, 204I, 588Q; 556P, 651F, 655I, 535N, 573F, 589V, 204I, 588Q, 647F; 556P, 651F, 535N, 573F; 556P, 651F, 535N, 573F, 589V; 556P, 651F, 535N, 573F, 589V, 204I; 556P, 651F, 535N, 573F, 589V, 204I, 588Q; 556P, 651F, 535N, 573F, 589V, 204I, 588Q, 647F; 556P, 655I, 535N, 573F; 556P, 655I, 535N, 573F, 589V; 556P, 655I, 535N, 573F, 589V, 204I; 556P, 655I, 535N, 573F, 589V, 204I, 588Q, 647F; 556P, 655I, 535N, 573F; 556P, 655I, 535N, 573F, 589V; 556P, 655I, 535N, 573F, 589V, 204I; 556P, 655I, 535N, 573F, 589V, 204I,

588Q; 556P, 655I, 535N, 573F, 589V, 204I, 588Q, 647F. Một lần nữa, bất kỳ trong số các phương án có thể là trong bất kỳ protein Env của HIV, ví dụ thể phân lập kiểu đại, Env phổ biến, protein Env tổng hợp, protein Env thể đột biến SOSIP, thể phân lập kiểu đại chứa đột biến sửa theo khái niệm được mô tả ở đây, v.v. Một số kết hợp được ưu tiên theo sáng chế bao gồm 655I, 589V, 573F, 651F, 588E, 535N, 204I; 556P, 655I, 535N, 573F, 589V, 204I, 588Q; 204I, 535N, 556P, 588E, 589V, 651F, 655I; 535N, 556P, 589V, 651F, 655I; và 535N, 556P, 588E, 589V, 651F, 655I.

Theo các phương án được ưu tiên nhất định, protein Env của HIV bao gồm trình tự mà là ít nhất 95% tương đồng với, tốt hơn là ít nhất 96%, 97%, 98%, 99% tương đồng với, tốt hơn là 100% tương đồng với, bất kỳ trong số các SEQ ID NO: 20, 22, 24, 26, 27, 28, 29, 30, 31 và 32. Để xác định % tương đồng, tốt hơn là các vị trí (i)-(xv) của Bảng 1 và 2, và tốt hơn là cả các vị trí 501, 559 và 605 không được xem xét. Tốt hơn là gốc axit amin ở các vị trí này là các gốc trong trình tự của SEQ ID NO: 20, 22, 24, 26, 27, 28, 29, 30, 31 hoặc 32, tương ứng.

Theo các phương án nhất định, protein Env của HIV theo sáng chế còn bao gồm: (xvi) gốc axit amin được chọn từ Val, Ile, Phe, Met, Ala, hoặc Leu ở vị trí 658. Tốt hơn là axit amin ở vị trí 658 là Val hoặc Ile, tốt nhất là Val. Phát hiện ra rằng tỷ lệ phần trăm trime và hiệu suất trime tăng mạnh của protein Env, hoặc là một mình hoặc kết hợp với các đột biến được chọn từ (i)-(vii) của Bảng 1 và/hoặc (viii)-(xv) của Bảng 2 được mô tả ở đây.

Theo các phương án của sáng chế, protein Env của HIV tái tổ hợp có ít nhất một trong số (a) tỷ lệ phần trăm tạo thành trime cải thiện và (b) hiệu suất trime được cải thiện so với protein Env HIV không có gốc axit amin được nêu ở một hoặc nhiều trong số các vị trí 651, 655, 535, 589, 573, 204, và 647 như được thể hiện trong Bảng 1.

Như được sử dụng ở đây “tỷ lệ phần trăm tạo thành trime cải thiện” nghĩa là tỷ lệ phần trăm trime lớn hơn được tạo thành khi trình tự mạch chính của protein vỏ HIV chứa một hoặc nhiều trong số sự thể axit amin theo sáng chế so với tỷ lệ phần trăm của trime mà được tạo thành khi trình tự mạch chính của trình tự vỏ HIV không chứa sự thể axit amin này. Như được sử dụng ở đây “hiệu suất trime cải thiện” nghĩa là tổng lượng của dạng trime của protein vỏ thu được khi trình tự mạch chính của protein vỏ HIV chứa một hoặc nhiều trong số sự thể axit amin theo sáng chế so với tổng lượng của dạng trime

của protein vỏ mà được tạo thành khi trình tự mạch chính của trình tự vỏ HIV không chứa sự thể axit amin này.

Sự tạo thành trime có thể là được đo bằng thử nghiệm liên kết kháng thể sử dụng các kháng thể mà liên kết đặc hiệu với dạng trime của protein Env của HIV. Các ví dụ của kháng thể đặc hiệu trime mà có thể được dùng để dò dạng trime bao gồm, nhưng không giới hạn ở, kháng thể đơn dòng (mAbs) PGT145, PGDM1400, PG16, và PGT151. Tốt hơn là, kháng thể đặc hiệu trime là mAb PGT145. Thử nghiệm liên kết kháng thể bất kỳ đã biết trong lĩnh vực liên quan đến sáng chế có thể được dùng để đo tỷ lệ phần trăm tạo thành trime của protein Env của HIV tái tổ hợp theo sáng chế, như ELISA, AlphaLISA, v.v.

Theo hương án, sự tạo thành trime được đo bằng AlphaLISA. AlphaLISA là thử nghiệm lân cận dựa trên hạt trong đó phân tử oxy đơn, được tạo ra bởi chiếu xạ năng lượng cao của hạt cho, được chuyển cho hạt nhận mà nằm trong khoảng cách xấp xỉ 200 nm so với hạt cho. Việc chuyển phân tử oxy đơn cho hạt nhận khởi đầu chuỗi thác của các phản ứng hóa học gây ra tín hiệu phát quang mà sau đó có thể được phát hiện (Eglen et al. *Curr. Chem. Genomics*, 2008, 25(1): 2-10). Ví dụ, protein vỏ HIV tái tổ hợp được dán nhãn bằng thể Flag-His có thể được ủ với mAb đặc hiệu trime, hạt cho liên hợp với kháng thể mà liên kết với mAb đặc hiệu trime, hạt cho liên hợp niken, hạt nhận liên hợp với kháng thể kháng His, và hạt nhận liên hợp với kháng thể kháng Flag. Lượng trime tạo thành có thể được xác định bằng cách đo tín hiệu phát quang tạo ra từ cặp hạt cho liên hợp với kháng thể mà liên kết với mAb đặc hiệu trime và hạt nhận liên hợp với kháng thể kháng His. Tổng lượng protein vỏ HIV được biểu hiện có thể được xác định bằng cách đo tín hiệu phát quang sinh ra từ cặp hạt cho liên hợp niken và hạt nhận liên hợp kháng Flag. Ví dụ, lượng trime và tổng protein vỏ được biểu hiện có thể là được đo bằng thử nghiệm AlphaLISA như được mô tả chi tiết trong Ví dụ 3. Tỷ lệ phần trăm tạo thành trime có thể được tính toán bằng cách chia lượng trime tạo thành cho tổng lượng protein vỏ được biểu hiện.

Lượng trime tạo thành và tổng lượng protein vỏ được biểu hiện có thể cũng được xác định bằng cách sử dụng các kỹ thuật sắc ký mà có khả năng tách dạng trime từ các dạng khác của protein vỏ HIV, ví dụ, dạng monome. Các ví dụ của kỹ thuật này mà có thể được sử dụng bao gồm, nhưng không giới hạn ở sắc ký loại kích cỡ tán xạ ánh sáng đa góc (SEC-MALS). Theo các phương án nhất định, tỷ lệ phần trăm tạo thành được

xác định bằng cách sử dụng SEC-MALS. Theo các phương án nhất định, hiệu suất trimer được xác định bằng cách sử dụng SEC-MALS.

Axit nucleic, vector, và tế bào

Theo khía cạnh chung khác, sáng chế đề xuất phân tử axit nucleic mã hóa protein Env của HIV tái tổ hợp theo sáng chế, và vector bao gồm phân tử axit nucleic. Phân tử axit nucleic theo sáng chế có thể ở dạng ARN hoặc ở dạng ADN thu được bằng cách tách dòng hoặc sản xuất bằng cách tổng hợp. ADN có thể là sợi kép hoặc sợi đơn. ADN có thể bao gồm ví dụ cADN, ADN bộ gen, hoặc kết hợp của chúng. Phân tử axit nucleic và vector có thể là được dùng để tạo ra protein tái tổ hợp, biểu hiện protein trong tế bào chủ, hoặc tạo ra hạt virus.

Theo các phương án của sáng chế, axit nucleic mã hoá protein vỏ HIV tái tổ hợp được liên kết hoạt động với gen khởi động, có nghĩa là axit nucleic này chịu sự kiểm soát của gen khởi động. Gen khởi động có thể là gen khởi động cùng loại (tức là, có nguồn gốc từ cùng nguồn di truyền với vector) hoặc gen khởi động khác loại (tức là, có nguồn gốc từ vector khác hoặc nguồn di truyền khác). Các ví dụ của gen khởi động thích hợp bao gồm gen khởi động ngay sớm của cytomegalovirus ở người (hCMV IE, hoặc gọi tắt là “CMV”) và gen khởi động của virus Rous Sarcoma (RSV). Tốt hơn là, gen khởi động nằm ngược chiều với axit nucleic ở trong băng biểu hiện.

Theo Phương án của sáng chế, vector có thể là vector biểu hiện. Vector biểu hiện bao gồm, nhưng không chỉ giới hạn ở, vector để biểu hiện protein tái tổ hợp và vector để chuyển axit nucleic vào đối tượng để biểu hiện ở mô của đối tượng, như vector virus. Ví dụ về vector virus thích hợp được sử dụng theo sáng chế bao gồm, nhưng không chỉ giới hạn ở vector adenovirus, vector virus kết hợp adeno, vector virus thủy đậu, vector Vaccinia Ankara sửa đổi MVA, vector virus ruột, vector virus gây viêm não ngựa Venezuela, vector virus rừng Semliki, vector virus khản thuốc lá, vector lentivirus, v.v.. Vector cũng có thể là vector không phải là virus. Ví dụ về vector không phải là virus bao gồm, nhưng không chỉ giới hạn ở plasmid, nhiễm sắc thể nhân tạo của vi khuẩn, nhiễm sắc thể nhân tạo của nấm men, thể thực khuẩn, v.v..

Theo các phương án nhất định của sáng chế, vector là vector adenovirus, ví dụ vector adenovirus tái tổ hợp. Vector adenovirus tái tổ hợp có thể ví dụ có nguồn gốc từ adenovirus ở người (HAdV, hoặc AdHu), hoặc adenovirus của khỉ như adenovirus của tinh tinh hoặc khỉ đột (ChAd, AdCh, hoặc SAdV) hoặc adenovirus của khỉ nâu (rhAd).

Tốt hơn là, vectơ adenovirut tái tổ hợp ở người, ví dụ kiểu huyết thanh adenovirut ở người tái tổ hợp 26, hoặc bất kỳ trong số các kiểu huyết thanh adenovirut ở người tái tổ hợp 5, 4, 35, 7, 48, v.v. theo phương án khác, vectơ adenovirut là vectơ rhAd, ví dụ rhAd51, rhAd52 hoặc rhAd53.

Điều chế vectơ adenovirut tái tổ hợp đã được biết rõ trong lĩnh vực kỹ thuật này. Ví dụ, điều chế vectơ adenovirut 26 tái tổ hợp được mô tả, trong, ví dụ, WO 2007/104792 và trong Abbink *et al.*, (2007) *Virology* 81(9): 4654-63. Trình tự bộ gen ví dụ của adenovirut 26 được tìm thấy trong Mã số truy cập GenBank EF 153474 và trong SEQ ID NO: 1 của WO 2007/104792. Trình tự bộ gen ví dụ cho rhAd51, rhAd52 và rhAd53 được đề xuất trong US 2015/0291935.

Theo các phương án của sáng chế, bất kỳ trong số các protein Env của HIV tái tổ hợp được mô tả ở đây có thể là được biểu hiện và/hoặc được mã hóa bằng vectơ được mô tả ở đây. Căn cứ vào sự thoái hóa của mã di truyền, chuyên gia trong lĩnh vực kỹ thuật này hiểu rõ rằng một vài trình tự axit nucleic có thể được thiết kế mà mã hoá cùng protein, theo phương pháp hoàn toàn thông thường trong lĩnh vực kỹ thuật này. Axit nucleic mã hóa protein Env HIV tái tổ hợp theo sáng chế có thể tùy ý được tối ưu hóa bộ ba mã hóa để đảm bảo biểu hiện đúng trong tế bào chủ (ví dụ, tế bào vi khuẩn hoặc động vật có vú). Sự tối ưu hóa bộ ba mã hóa là công nghệ được áp dụng rộng rãi trong lĩnh vực kỹ thuật này.

Sáng chế còn đề xuất tế bào, tốt hơn là tế bào được phân lập, có chứa bất kỳ trong số các phân tử axit nucleic được mô tả trong bản mô tả này. Tế bào có thể ví dụ được sử dụng để sản xuất protein tái tổ hợp, hoặc sản xuất hạt virus.

Do đó các phương án của sáng chế cũng đề cập đến phương pháp tạo ra protein Env HIV tái tổ hợp. Phương pháp này bao gồm bước chuyển nhiễm tế bào chủ bằng vectơ biểu hiện có chứa axit nucleic mã hoá protein Env HIV tái tổ hợp theo phương án của sáng chế liên kết hoạt động với gen khởi động, nuôi tế bào đã được chuyển nhiễm trong điều kiện thích hợp để biểu hiện protein Env HIV tái tổ hợp, và tinh chế tùy ý hoặc phân lập protein Env HIV tái tổ hợp được biểu hiện trong tế bào. Protein Env HIV tái tổ hợp có thể được phân lập hoặc được thu gom từ tế bào bằng phương pháp bất kỳ đã biết trong lĩnh vực bao gồm sắc ký ái lực, sắc ký loại kích thước, v.v. Kỹ thuật được sử dụng để biểu hiện protein tái tổ hợp được biết rõ đối với chuyên gia trong lĩnh vực dựa vào sáng chế. Protein Env HIV tái tổ hợp được biểu hiện có thể cũng được nghiên cứu mà

không cần tinh chế hoặc phân lập protein được biểu hiện, ví dụ, bằng cách phân tích dịch nổi của tế bào chuyển nhiễm với vector biểu hiện mã hóa protein Env HIV tái tổ hợp và nuôi dưới các điều kiện thích hợp để biểu hiện protein Env HIV.

Theo phương án được ưu tiên, protein Env HIV tái tổ hợp được biểu hiện được tinh chế dưới các điều kiện mà cho phép tập hợp protein để tạo thành phức trime được làm ổn định. Ví dụ, tế bào động vật có vú được chuyển nhiễm với vector biểu hiện mã hóa protein Env HIV tái tổ hợp liên kết hoạt động với trình tự khởi đầu (ví dụ trình tự khởi đầu CMV) có thể được nuôi cấy ở 33-39°C, ví dụ 37°C, và 2-12% CO₂, ví dụ 8% CO₂. Biểu hiện có thể được thực hiện trong hệ biểu hiện tùy ý như tế bào côn trùng hoặc tế bào nấm men, tất cả đều đã biết trong lĩnh vực. Protein Env HIV được biểu hiện sau đó có thể được phân lập từ nuôi cấy tế bào ví dụ sắc ký ái lực lectin, mà liên kết glycoprotein. Protein Env HIV liên kết với cột có thể được tách rửa bằng mannopyranosit. Protein Env HIV được tách rửa từ cột có thể được đưa vào bước tinh chế thêm, như sắc ký loại kích cỡ, nếu cần, để loại bất kỳ tạp chất dư, ví dụ, tạp chất tế bào, mà cả chất kết tụ Env, monome gp140 và monome gp120. Các phương pháp tinh chế tùy ý, các ví dụ không giới hạn bao gồm sắc ký ái lực kháng thể, chọn lọc âm bằng tinh chế không-bNAb, kháng thể, hoặc các phương pháp sắc ký khác như sắc ký trao đổi ion v.v., cũng như các phương pháp khác đã biết trong lĩnh vực, có thể cũng được sử dụng để phân lập protein Env HIV được biểu hiện.

Phân tử axit nucleic và vector biểu hiện mã hóa protein Env của HIV tái tổ hợp theo sáng chế có thể được tạo ra bằng phương pháp bất kỳ đã biết trong lĩnh vực liên quan đến sáng chế. Ví dụ, axit nucleic mã hóa protein Env HIV tái tổ hợp có thể được điều chế bằng cách đưa ít nhất một trong số các sự thể axit amin ở các vị trí được nêu vào trình tự vỏ HIV mạch chính bằng cách sử dụng công nghệ kỹ thuật di truyền và kỹ thuật sinh học phân tử, ví dụ, gây đột biến hướng điểm, phản ứng chuỗi polymeraza (PCR), v.v., mà là các phương pháp đã biết đối với chuyên gia trong lĩnh vực. Phân tử axit nucleic sau đó có thể được đưa hoặc “được tách dòng” vào vector biểu hiện cũng bằng cách sử dụng các kỹ thuật sinh học phân tử tiêu chuẩn. Protein vỏ HIV tái tổ hợp sau đó có thể được biểu hiện từ vector biểu hiện trong tế bào chủ, và protein được biểu hiện được tinh chế từ nuôi cấy tế bào bằng phương pháp bất kỳ đã biết trong lĩnh vực liên quan đến sáng chế.

Phức trime

Theo khía cạnh chung khác, sáng chế đề cập đến phức trime bao gồm oligome không cộng hóa trị của ba trong số các protein Env của HIV tái tổ hợp theo sáng chế. Phức trime có thể bao gồm bất kỳ của các protein Env của HIV tái tổ hợp được mô tả ở đây. Tốt hơn là phức trime bao gồm ba monome giống nhau (hoặc heterodime giống nhau nếu gp140 bị phân cắt) của protein Env của HIV tái tổ hợp theo sáng chế. Phức trime có thể được tách từ dạng khác của protein vỏ HIV, như dạng monome, hoặc phức trime có thể có mặt cùng với dạng khác của protein vỏ HIV, như dạng monome.

Chế phẩm và Phương pháp

Theo khía cạnh chung khác, sáng chế đề cập đến chế phẩm bao gồm protein Env của HIV tái tổ hợp, phức trime, axit nucleic phân lập, vector, hoặc tế bào chủ, và chất mang được dụng. Chế phẩm có thể bao gồm bất kỳ trong số các protein Env của HIV tái tổ hợp, phức trime, phân tử axit nucleic phân lập, vector, hoặc tế bào chủ được mô tả ở đây.

Chất mang có thể bao gồm một hoặc nhiều tá được dụng như chất liên kết, chất phân rã, tác nhân làm phòng, chất tạo huyền phù, chất nhũ hóa, chất làm ẩm, chất bôi trơn, hương liệu, chất làm ngọt, chất bảo quản, thuốc nhuộm, chất làm tan và chất phủ. Bản chất chính xác của chất mang hoặc vật liệu khác có thể phụ thuộc vào đường dùng, ví dụ, các đường trong cơ, trong da, dưới da, qua đường miệng, trong tĩnh mạch, qua da, trong niêm mạc (ví dụ như, ruột), trong mũi hoặc trong màng bụng. Đối với chế phẩm tiêm được dạng lỏng, ví dụ, huyền phù và dung dịch, chất mang và chất phụ trợ thích hợp bao gồm nước, glycol, dầu, rượu, chất bảo quản, chất tạo màu và chất tương tự. Đối với chế phẩm dùng qua đường miệng dạng rắn, ví dụ, bột, viên nang, viên nén hình elip, gelcap và viên nén, chất mang và chất phụ trợ thích hợp bao gồm tinh bột, đường, chất pha loãng, chất tạo hạt, chất bôi trơn, chất gắn, chất làm phân rã và chất tương tự. Đối với hỗn hợp chất phun/thuốc xông qua đường mũi, dung dịch/huyền phù trong nước có thể có chứa nước, glycol, dầu, chất làm mềm, chất ổn định, chất làm ẩm, chất bảo quản, chất thơm, hương vị, và chất tương tự làm chất mang và chất phụ trợ thích hợp.

Chế phẩm theo sáng chế có thể được tạo chế phẩm trong vật chất bất kỳ thích hợp để sử dụng cho đối tượng để làm thuận lợi cho việc sử dụng và cải thiện hiệu quả, bao gồm, nhưng không giới hạn ở, việc sử dụng qua đường miệng (trong ruột) và tiêm ngoài đường tiêu hóa. Việc tiêm ngoài đường tiêu hóa bao gồm tiêm hoặc truyền trong

tĩnh mạch, tiêm dưới da, tiêm trong da, và tiêm trong cơ. Chế phẩm theo sáng chế cũng có thể được tạo chế phẩm cho các đường dùng khác bao gồm qua niêm mạc, qua mắt, qua trực tràng, cấy ghép tác động lâu dài, sử dụng dưới lưỡi, dưới lưỡi, từ niêm mạc miệng bỏ qua sự tuần hoàn tĩnh mạch, xông, hoặc trong mũi.

Phương án theo sáng chế còn đề cập đến phương pháp tạo ra chế phẩm. Theo các phương án của sáng chế, phương pháp tạo ra chế phẩm bao gồm trộn protein Env của HIV tái tổ hợp, phức trime, axit nucleic phân lập, vectơ, hoặc tế bào chủ theo sáng chế với một hoặc nhiều chất mang được dùng. Chuyên gia trong lĩnh vực kỹ thuật này quen thuộc với các kỹ thuật thông thường được sử dụng để điều chế các chế phẩm này.

Kháng nguyên HIV (ví dụ, protein hoặc đoạn của chúng có nguồn gốc từ HIV *gag*, *pol*, và/hoặc sản phẩm gen *env*) và vectơ, như vectơ virus, biểu hiện kháng nguyên HIV đã được dùng trước đây trong chế phẩm gây miễn dịch và vắc-xin để chủng ngừa đối tượng chống nhiễm HIV, hoặc để tạo ra đáp ứng miễn dịch chống nhiễm HIV ở đối tượng. Như được sử dụng ở đây, "đối tượng" có nghĩa là động vật bất kỳ, tốt hơn là động vật có vú, tốt nhất là người, sẽ được dùng hoặc đã dùng chế phẩm sinh miễn dịch theo các phương án của sáng chế. Thuật ngữ "động vật có vú" như được dùng trong bản mô tả này, bao hàm động vật có vú bất kỳ. Ví dụ về động vật có vú bao gồm, nhưng không chỉ giới hạn ở, chuột nhắt, chuột cống, thỏ, chuột lang, khỉ, người, v.v., tốt hơn là người. Protein Env của HIV tái tổ hợp theo sáng chế có thể còn được dùng làm kháng nguyên để gây đáp ứng miễn dịch chống lại virus gây suy giảm miễn dịch ở người (HIV) ở đối tượng cần chúng. Đáp ứng miễn dịch có thể là chống một hoặc nhiều nhánh HIV, như nhánh A, nhánh B, nhánh C, v.v. Chế phẩm có thể chứa vectơ mà protein Env HIV tái tổ hợp được biểu hiện từ đó, hoặc chế phẩm có thể chứa protein Env HIV tái tổ hợp được phân lập theo phương án theo sáng chế.

Ví dụ, chế phẩm bao gồm protein HIV tái tổ hợp hoặc phức trime của chúng có thể được dùng cho đối tượng cần chúng để gây đáp ứng miễn dịch chống nhiễm HIV ở đối tượng. Chế phẩm bao gồm vectơ, như vectơ adenovirus, mã hóa protein Env của HIV tái tổ hợp theo sáng chế, trong đó protein Env HIV tái tổ hợp được biểu hiện bởi vectơ, có thể còn được dùng cho đối tượng cần chúng để gây đáp ứng miễn dịch chống nhiễm HIV ở đối tượng. Phương pháp được mô tả ở đây còn bao gồm việc dùng chế phẩm theo sáng chế kết hợp với một hoặc nhiều kháng nguyên HIV bổ sung (ví dụ, protein hoặc đoạn của chúng có nguồn gốc từ sản phẩm gen HIV *gag*, *pol*, và/hoặc *env*

mà tốt hơn là được biểu hiện từ một hoặc nhiều vector, như vector adenovirut hoặc vector MVA, bao gồm phương pháp môi và tăng cường đáp ứng miễn dịch.

Theo các phương án nhất định, protein Env HIV có thể được thể hiện trên hạt, như liposome, hạt tương tự virut (VLP), hạt nano, thể virut, hoặc túi có nguồn gốc tế bào, tùy ý kết hợp với tá được nội sinh và/hoặc ngoại sinh. Khi so với protein Env tan được hoặc monome trên chính nó, các hạt này thường thể hiện hiệu quả tăng cường thể heinej kháng nguyên trong cơ thể.

Các ví dụ của VLP mà thể hiện protein Env của HIV có thể được điều chế ví dụ bằng cách đồng biểu hiện protein Env của HIV với protein virut tự lắp ghép như lõi HIV Gag hoặc protein Gag retrovirut khác. VLP giống virut, nhưng không lây nhiễm vì chúng không chứa vật liệu gen virut. Biểu hiện của protein cấu trúc virut, như vỏ hoặc bao, có thể tạo ra sự tự lắp ghép của VLP. VLP là đã biết đối với chuyên gia trong lĩnh vực, và sử dụng chúng trong vắc-xin được mô tả ví dụ trong (Kushnir et al, 2012).

Theo các phương án nhất định được ưu tiên, vật truyền là adenovirut tái tổ hợp. Liposome là túi cầu có ít nhất một lớp kép lipid. Protein trime Env HIV có thể ví dụ được liên hợp không cộng hóa trị với liposom này bằng tương tác tĩnh điện, ví dụ bằng cách thêm thể His vào đầu C của trime Env HIV và nguyên tử chelat hóa hóa trị hai như Ni^{2+} hoặc Co^{2+} được kết hợp vào nhóm đầu của lipid được dẫn xuất trong liposome. Theo các phương án không giới hạn và ví dụ nhất định, liposome bao gồm 1,2-distearoyl-sn-glyxero-3-phosphocholin (DSPC), cholesterol, và muối Niken hoặc Cô ban của 1,2-dioleoyl-sn-glyxero-3-[(N-(5-amino-1-carboxypentyl)iminodioxetic axit)succinyl] (DGS-NTA(Ni^{2+}) hoặc DGS-NTA(Co^{2+})) ở tỷ lệ mol 60:36:4. Theo phương án được ưu tiên, protein trime vỏ HIV được liên hợp cộng hóa trị với bề mặt liposom, ví dụ qua nhóm chức maleimit được tích hợp trong bề mặt liposome. Theo các phương án ví dụ không giới hạn nhất định của chúng, liposome bao gồm DSPC, cholesterol, và 1,2-dipalmitoyl-sn-glyxero-3-phosphoetanolamin-N-[4-(p-maleimidometyl)xyclohexan-carboxamit] lipid ở tỷ lệ mol bằng 54:30:16. Protein Env của HIV có thể được liên hợp vào đó ví dụ qua xystein đầu C bổ sung trong protein Env của HIV. Các biến thể liên hợp cộng hóa trị là ổn định hơn, tạo ra chuẩn độ IgG đặc hiệu kháng nguyên cao và epitop ở ‘đáy’ ít liên quan đến kháng nguyên của trime Env được giấu đi. Phương pháp điều chế trime Env của HIV được liên hợp với liposom, cũng như đặc tính của chúng, là đã biết và đã được mô tả ví dụ trong (Bale et al, 2017), được kết hợp vào đây bằng

cách viện dẫn. Sáng chế còn đề xuất protein Env HIV theo sáng chế được dung hợp vào và/hoặc được thể hiện trên liposome.

Theo các phương án nhất định, protein Env HIV theo sáng chế được dung hợp vào các hạt tự lắp ghép, hoặc được thể hiện trên các hạt nano. Các hạt nano kháng nguyên là sự lắp ghép của các polypeptit mà có nhiều bản sao kháng nguyên, ví dụ protein Env của HIV theo sáng chế, mà tạo ra nhiều vị trí liên kết (ái lực) và có thể tạo ra tính ổn định kháng nguyên và khả năng miễn dịch cải thiện. Việc điều chế và sử dụng các hạt nano protein tự lắp ghép để dùng trong vắc-xin là đã biết đối với chuyên gia trong lĩnh vực, xem ví dụ (Zhao et al, 2014), (López-Sagaseta et al, 2016). Là ví dụ không giới hạn, các hạt nano tự lắp ghép có thể là dựa trên feritin, bacterioferritin, hoặc DPS. Các hạt nano DPS thể hiện protein trên bề mặt của chúng ví dụ được mô tả trong WO2011/082087. Mô tả của kháng nguyên HIV-1 trime trên các hạt này đã được mô tả ví dụ trong (He et al, 2016). Các hạt nano protein tự lắp ghép khác cũng như điều chế chúng, được bộc lộ ví dụ trong WO 2014/124301, và US 2016/0122392, được kết hợp vào đây bằng cách viện dẫn. Sáng chế còn đề xuất protein Env HIV theo sáng chế được dung hợp vào và/hoặc được thể hiện trên các hạt nano tự lắp ghép. Sáng chế còn đề xuất chế phẩm bao gồm VLP, liposome, hoặc các hạt nano tự lắp ghép theo sáng chế.

Theo các phương án nhất định, tá được được bao gồm trong chế phẩm theo sáng chế hoặc được dùng đồng thời với chế phẩm theo sáng chế. Sử dụng tá được là tùy ý, và có thể còn tăng cường đáp ứng miễn dịch khi chế phẩm được dùng cho mục đích chủng ngừa. Tá được thích hợp để đồng sử dụng hoặc bao gồm trong chế phẩm theo sáng chế tốt hơn là các loại mà an toàn tiềm năng, dung hợp tốt và hiệu quả ở người. Tá được này là đã biết đối với chuyên gia trong lĩnh vực, và các ví dụ không giới hạn bao gồm QS-21, Detox-PC, MPL-SE, MoGM-CSF, TiterMax-G, CRL- 1005, GERBU, TERamide, PSC97B, Adjuver, PG-026, GSK-I, GcMAF, B-aletin, MPC-026, Adjuvax, CpG ODN, Betafectin, muối nhôm như Nhôm Phosphat (ví dụ AdjuPhos) hoặc Nhôm Hydroxit, và MF59.

Các khía cạnh khác theo sáng chế đề cập đến protein vỏ HIV tái tổ hợp bao gồm trình tự axit amin mà là ít nhất 95%, 96%, 97%, 98%, 99% hoặc 100% tương đồng với trình tự axit amin nêu ở SEQ ID NO: 2 hoặc SEQ ID NO: 4, mà thể hiện trình tự nhánh C phổ biến và nhánh B phổ biến của vỏ HIV, tương ứng. Các trình tự phổ biến này đã thấy trong bất kỳ trình tự có trong tự nhiên, và do đó được tin là protein vỏ HIV mới.

Protein vỏ HIV tái tổ hợp bao gồm trình tự axit amin mà là ít nhất 95%, 96%, 97%, 98%, 99% hoặc 100% tương đồng với trình tự axit amin nêu ở SEQ ID NO: 2 hoặc SEQ ID NO: 4 có thể tùy ý còn bao gồm cái gọi là đột biến SOSIP và/hoặc đột biến ở vị trí phân cắt furin, như, ví dụ trong các trình tự được thể hiện trong SEQ ID NO: 3 hoặc SEQ ID NO: 3 còn bao gồm Pro ở vị trí 558 và/hoặc vị trí 556; và SEQ ID NO: 5 hoặc SEQ ID NO: 5 còn bao gồm Pro ở vị trí 558 và/hoặc vị trí 556. Khi xác định % tương đồng cho các trình tự này, axit amin ở vị trí phân cắt furin đột biến và ở các vị trí 501, 605, 559, 556 và 558 tốt hơn là không được xem xét. Đã phát hiện một cách ngạc nhiên là protein này được biểu hiện ở mức cao và có mức ổn định và tạo thành trimer cao. Các protein Env HIV này có thể theo các phương án nhất định được dùng làm protein mạch chính, trong đó các đột biến được mô tả ở trên có thể có thể được thực hiện để thu được phân tử theo sáng chế. Phân tử axit nucleic phân lập mã hóa các trình tự này, vector bao gồm các trình tự này liên kết hoạt động với trình tự khởi đầu, và chế phẩm bao gồm protein, phân tử axit nucleic phân lập, hoặc vector cũng được bao hàm bởi sáng chế.

Phương án thực hiện sáng chế

Phương án 1 là protein Env của HIV tái tổ hợp, bao gồm trình tự axit amin của protein Env HIV có gốc axit amin được nêu ở ít nhất hai trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp, tốt hơn là Phe, ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp, tốt hơn là Ile, ở vị trí 655;
- (iii) Asn hoặc Gln, tốt hơn là Asn, ở vị trí 535;
- (iv) Val, Ile hoặc Ala, tốt hơn là Val hoặc Ile, ở vị trí 589;
- (v) Phe hoặc Trp, tốt hơn là Phe, ở vị trí 573;
- (vi) Ile ở vị trí 204; và
- (vii) Phe, Met, hoặc Ile, tốt hơn là Phe, ở vị trí 647,

trong đó việc đánh số các vị trí là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1.

Phương án 2 là protein Env của HIV tái tổ hợp, bao gồm trình tự axit amin của protein Env HIV và sự thể axit amin bằng gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp, tốt hơn là Phe, ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp, tốt hơn là Ile, ở vị trí 655;

- (iii) Asn hoặc Gln, tốt hơn là Asn, ở vị trí 535;
- (iv) Val, Ile hoặc Ala, tốt hơn là Val hoặc Ile, ở vị trí 589;
- (v) Phe hoặc Trp, tốt hơn là Phe ở vị trí 573;
- (vi) Ile ở vị trí 204; và
- (vii) Phe, Met, hoặc Ile, tốt hơn là Phe, ở vị trí 647,

Theo một số phương án, phân tử liên kết được chọn từ nhóm bao gồm:

- (1) protein Env HIV có trình tự phổ biến, ví dụ từ nhánh C hoặc từ nhánh B, ví dụ bao gồm trình tự axit amin của (SEQ ID NO: 2, 3, 4 hoặc 5); hoặc
 - (2) protein vỏ HIV tổng hợp, ví dụ có chứa trình tự axit amin nêu ở (a) SEQ ID NO: 6; (b) SEQ ID NO: 6 với đột biến Glu thành Arg ở vị trí 166; (c) SEQ ID NO: 7; hoặc (d) SEQ ID NO: 8 hoặc SEQ ID NO: 9), trong đó (a), (b) hoặc (d) tùy ý có thể có thêm đột biến SOSIP (Cys ở các vị trí 501 và 605 và Pro ở vị trí 559) và/hoặc đột biến ở vị trí phân cắt furin (ví dụ SEQ ID NO: 10 thay thế axit amin 508-511); hoặc
 - (3) protein Env HIV kiểu đại, tốt hơn là của nhánh C, bao gồm ít nhất một đột biến sửa ở gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất ít hơn 7,5%, tốt hơn là ít hơn 2%, của trình tự Env HIV trong tập hợp của ít nhất 100, tốt hơn là ít nhất 1000, tốt hơn là ít nhất 10000, trình tự Env HIV kiểu đại, trong đó đột biến sửa là sự thế bằng gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất bằng ít nhất 10% của trình tự Env HIV trong tập hợp này và tốt hơn là đột biến sửa là sự thế bằng gốc axit amin mà có mặt ở vị trí tương ứng thường xuyên nhất trong tập hợp này; và
- trong đó việc đánh số các vị trí là theo đánh số trong gp160 của thể phân lập

HXB2 của HIV-1.

Phương án 3 là protein Env của HIV tái tổ hợp, bao gồm trình tự axit amin của protein Env HIV và sự thế axit amin bằng gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (i) Phe, Leu, Met, hoặc Trp, tốt hơn là Phe, ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp, tốt hơn là Ile, ở vị trí 655;
- (iii) Asn hoặc Gln, tốt hơn là Asn, ở vị trí 535;
- (iv) Val, Ile hoặc Ala, tốt hơn là Val hoặc Ile, ở vị trí 589;
- (v) Phe hoặc Trp, tốt hơn là Phe, ở vị trí 573;

(vi) Ile ở vị trí 204; và

(vii) Phe, Met, hoặc Ile, tốt hơn là Phe, ở vị trí 647,

trong đó protein Env HIV là protein Env HIV đột biến SOSIP bao gồm ít nhất một đột biến được chọn từ nhóm bao gồm:

(a) Cys ở các vị trí 501 và 605;

(b) Pro ở vị trí 559;

(c) Cys ở các vị trí 501 và 605 và Pro ở vị trí 559; và

việc đánh số các vị trí là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1.

Phương án 4 là protein Env HIV tái tổ hợp theo phương án 2, bao gồm gốc axit amin được nêu ở ít nhất hai trong số vị trí được nêu được chọn từ nhóm bao gồm (i) đến (vii).

Phương án 5 là protein Env HIV tái tổ hợp theo phương án 3, bao gồm gốc axit amin được nêu ở ít nhất hai trong số vị trí được nêu được chọn từ nhóm bao gồm (i) đến (vii).

Phương án 6 là protein Env của HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1, 2 và 4, còn bao gồm Cys ở các vị trí 501 và 605 hoặc Pro ở vị trí 559, tốt hơn là Cys ở các vị trí 501 và 605 và Pro ở vị trí 559.

Phương án 7 là protein Env của HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 6, bao gồm gốc axit amin được nêu ở ít nhất ba trong số vị trí được nêu được chọn từ nhóm bao gồm (i) đến (vii).

Phương án 8 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 6, bao gồm gốc axit amin được nêu ở ít nhất bốn trong số vị trí được nêu được chọn từ nhóm bao gồm (i) đến (vii).

Phương án 9 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 6, bao gồm gốc axit amin được nêu ở ít nhất năm trong số vị trí được nêu được chọn từ nhóm bao gồm (i) đến (vii).

Phương án 10 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 6, bao gồm gốc axit amin được nêu ở ít nhất sáu trong số vị trí được nêu được chọn từ nhóm bao gồm (i) đến (vii).

Phương án 11 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 6, bao gồm gốc axit amin được nêu ở bảy trong số các vị trí được nêu được chọn từ nhóm bao gồm (i) đến (vii).

Phương án 12 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1, 4, và 5, trong đó ít nhất hai vị trí và gốc được nêu là kết hợp được chọn từ nhóm bao gồm: 651F, 655I; 651F, 535N; 651F, 589V; 651F, 589I; 651F, 573F; 651F, 204I; 651F, 647F; 655I, 535N; 655I, 589V; 655I, 589I; 655I, 573F; 655I, 204I; 655I, 647F; 535N, 589V; 535N, 589I; 535N, 573F; 535N, 204I; 535N, 647F ; 589V, 573F; 589V, 204I; 589V, 647F; 589I, 573F; 589I, 204I; 589I, 647F; 573F, 204I; 573F, 647F; và 204I, 647F.

Phương án 13 là protein Env HIV tái tổ hợp theo phương án 7, trong đó ít nhất ba vị trí và gốc được nêu là kết hợp được chọn từ nhóm bao gồm: 651F, 655I, 535N; 651F, 589V, 535N; 651F, 589I, 535N; 651F, 573F, 535N; 651F, 204I, 535N; 651F, 647F, 535N; 655I, 589V, 535N; 655I, 589I, 535N; 655I, 573F, 535N; 655I, 204I, 535N; 655I, 647F, 535N; 589V, 573F, 535N; 589V, 204I, 535N; 589V, 647F, 535N; 589I, 573F, 535N; 589I, 204I, 535N; 589I, 647F, 535N; 573F, 204I, 535N; 573F, 647F, 535N; 204I, 647F, 535N; 651F, 655I, 589V; 651F, 573F, 589V; 651F, 204I, 589V; 651F, 647F, 589V; 655I, 573F, 589V; 655I, 204I, 589V; 655I, 647F, 589V; 573F, 204I, 589V; 573F, 647F, 589V; 204I, 647F, 589V; 651F, 655I, 589I; 651F, 573F, 589I; 651F, 204I, 589I; 651F, 647F, 589I; 655I, 573F, 589I; 655I, 204I, 589I; 655I, 647F, 589I; 573F, 204I, 589I; 573F, 647F, 589I; 204I, 647F, 589I; 651F, 655I, 573F; 651F, 204I, 573F; 651F, 647F, 573F; 655I, 204I, 573F; 655I, 647F, 573F; 204I, 647F, 573F; 651F, 655I, 204I; 651F, 647F, 204I; 655I, 647F, 204I; 651F, 655I, 647F; 655I, 651F, 647F; 655I, 651F, 535N; 655I, 589V, 573F; và 655I, 589V, 204I.

Phương án 14 là protein Env HIV tái tổ hợp theo phương án 8, trong đó ít nhất bốn vị trí và gốc được nêu là kết hợp được chọn từ nhóm bao gồm: 651F, 655I, 535N, 589V; 651F, 655I, 535N, 573F; 651F, 655I, 589V, 573F; 651F, 535N, 589V, 573F; 655I, 535N, 589V, 573F; 651F, 655I, 535N, 204I; 651F, 655I, 589V, 204I; 651F, 535N, 589V, 204I; 655I, 535N, 589V, 204I; 651F, 655I, 573F, 204I; 651F, 535N, 573F, 204I; 655I, 535N, 573F, 204I; 651F, 589V, 573F, 204I; 655I, 589V, 573F, 204I; 535N, 589V, 573F, 204I; 651F, 655I, 535N, 647F; 651F, 655I, 589V, 647F; 651F, 535N, 589V, 647F; 655I, 535N, 589V, 647F; 651F, 655I, 573F, 647F; 651F, 535N, 573F, 647F; 655I, 535N,

573F, 647F; 651F, 589V, 573F, 647F; 655I, 589V, 573F, 647F; 535N, 589V, 573F, 647F; 651F, 655I, 204I, 647F; 651F, 535N, 204I, 647F; 655I, 535N, 204I, 647F; 651F, 589V, 204I, 647F; 655I, 589V, 204I, 647F; 535N, 589V, 204I, 647F; 651F, 573F, 204I, 647F; 655I, 573F, 204I, 647F; 535N, 573F, 204I, 647F; và 589V, 573F, 204I, 647F.

Phương án 15 là protein Env HIV tái tổ hợp theo phương án 9, trong đó ít nhất năm vị trí và gốc được nêu là kết hợp được chọn từ nhóm bao gồm: 651F, 655I, 535N, 589V, 573F; 651F, 655I, 535N, 589V, 204I; 651F, 655I, 535N, 573F, 204I; 651F, 655I, 589V, 573F, 204I; 651F, 535N, 589V, 573F, 204I; 655I, 535N, 589V, 573F, 204I; 651F, 655I, 535N, 589V, 647F; 651F, 655I, 535N, 573F, 647F; 651F, 655I, 589V, 573F, 647F; 651F, 535N, 589V, 573F, 647F; 655I, 535N, 589V, 573F, 647F; 651F, 655I, 535N, 204I, 647F; 651F, 655I, 589V, 204I, 647F; 651F, 535N, 589V, 204I, 647F; 655I, 535N, 589V, 204I, 647F; 651F, 655I, 573F, 204I, 647F; 651F, 535N, 573F, 204I, 647F; 655I, 535N, 573F, 204I, 647F; 651F, 589V, 573F, 204I, 647F; 655I, 589V, 573F, 204I, 647F; và 535N, 589V, 573F, 204I, 647F.

Phương án 16 là protein Env HIV tái tổ hợp theo phương án 10, trong đó ít nhất sáu vị trí và gốc được nêu là kết hợp được chọn từ nhóm bao gồm: 651F, 655I, 535N, 589V, 573F, 204I; 651F, 655I, 535N, 589V, 573F, 647F; 651F, 655I, 535N, 589V, 204I, 647F; 651F, 655I, 535N, 573F, 204I, 647F; 651F, 655I, 589V, 573F, 204I, 647F; 651F, 535N, 589V, 573F, 204I, 647F; và 655I, 535N, 589V, 573F, 204I, 647F.

Phương án 17 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 16, còn bao gồm sự thể axit amin bằng gốc axit amin được nêu ở ít nhất một trong số vị trí được nêu được chọn từ nhóm bao gồm:

- (viii) Gln, Glu, Ile, Met, Val, Trp, hoặc Phe, tốt hơn là Gln hoặc Glu, ở vị trí 588;
- (ix) Lys ở vị trí 64 hoặc Arg ở vị trí 66 hoặc cả Lys ở vị trí 64 và Arg ở vị trí 66;
- (x) Trp ở vị trí 316;
- (xi) Cys ở cả vị trí 201 và 433;
- (xii) Pro ở vị trí 556 hoặc 558 hoặc ở cả vị trí 556 và 558; và
- (xiii) thay thế vòng ở vị trí axit amin 548-568 (vòng HR1) bằng vòng có 7-10 axit amin, tốt hơn là vòng có 8 axit amin, ví dụ có trình tự được chọn từ một trình tự bất kỳ trong số (SEQ ID NO: 12-17);
- (xiv) Gly ở vị trí 568, hoặc Gly ở vị trí 569, hoặc Gly ở vị trí 636, hoặc Gly ở cả vị trí 568 và 636, hoặc Gly ở cả vị trí 569 và 636; và/hoặc

(xv) Tyr ở vị trí 302, hoặc Arg ở vị trí 519, hoặc Arg ở vị trí 520, hoặc Tyr ở vị trí 302 và Arg ở vị trí 519, hoặc Tyr ở vị trí 302 và Arg ở vị trí 520, hoặc Tyr ở vị trí 302 và Arg ở cả vị trí 519 và 520, trong đó việc đánh số các vị trí là theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1.

Phương án 18 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 17, còn bao gồm đột biến trong trình tự phân cắt furin của protein Env HIV.

Phương án 19 là protein Env HIV tái tổ hợp theo phương án 18, trong đó đột biến ở vị trí phân cắt furin là thay thế ở các vị trí 508-511 bằng RRRRRR (SEQ ID NO: 10).

Phương án 20 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 19, là gp140 hoặc gp160.

Phương án 21 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 20, trong đó protein Env HIV tái tổ hợp có ít nhất một trong số phần trăm tạo thành trime cải thiện và hiệu suất trime được cải thiện so với protein Env HIV không có một hoặc nhiều trong số gốc axit amin được nêu ở các vị trí được nêu được chọn từ nhóm bao gồm (i) đến (vii).

Phương án 22 là protein Env HIV tái tổ hợp theo phương án 21, trong đó sự tạo thành trime được đo bằng sắc ký loại kích cỡ với tán xạ ánh sáng đa góc (SEC-MALS).

Phương án 23 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 22, còn bao gồm gốc axit amin được chọn từ Val, Ile, Phe, Met, Ala, hoặc Leu, tốt hơn là Val hoặc Ile, tốt nhất là Val, ở vị trí 658.

Phương án 24 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 23, bao gồm kết hợp của axit amin được chọn từ nhóm bao gồm:

(a) 655I, 589V, 573F, 651F, 588E, 535N, 204I;

(b) 556P, 655I, 535N, 573F, 589V, 204I, 588Q;

(c) 204I, 535N, 556P, 588E, 589V, 651F, 655I;

(d) 535N, 556P, 589V, 651F, 655I; và

(e) 535N, 556P, 588E, 589V, 651F, 655I.

Phương án 25 là protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 24, bao gồm trình tự axit amin mà là ít nhất 95%, 96%, 97%, 98%, 99% tương đồng với, hoặc 100% tương đồng với, bất kỳ trong số các SEQ ID NO: 3, 5,

20, 22, 24, 26, 27, 28, 29, 30, 31, hoặc 32, tốt hơn là ít nhất 98% tương đồng với bất kỳ trong số các SEQ ID NO: 20, 22, 24, 26, 27, 28, 29, 30, 31, hoặc 32.

Phương án 26 là phức trime bao gồm oligome không cộng hóa trị của ba trong số các protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 25.

Phương án 27 là hạt, ví dụ liposome hoặc hạt nano, ví dụ hạt nano tự lắp ghép, thể hiện protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án từ 1-25 hoặc phức trime theo phương án 26.

Phương án 28 là phân tử axit nucleic phân lập mã hóa protein Env của HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 25.

Phương án 29 là vector chứa phân tử axit nucleic phân lập theo phương án 28 liên kết hoạt động với trình tự khởi đầu.

Phương án 30 là vector theo phương án 29, mà là vector adenovirut.

Phương án 31 là tế bào chủ bao gồm phân tử axit nucleic phân lập theo phương án 28 hoặc vector theo phương án 29 hoặc 30.

Phương án 32 là phương pháp tạo ra protein Env của HIV tái tổ hợp, bao gồm nuôi tế bào chủ theo phương án 31 dưới các điều kiện thích hợp để tạo ra protein Env HIV tái tổ hợp.

Phương án 33 là phương pháp tạo ra protein Env của HIV tái tổ hợp bao gồm thu được vector biểu hiện bao gồm axit nucleic phân lập theo phương án 28 liên kết hoạt động với trình tự khởi đầu; chuyển nhiễm tế bào với vector biểu hiện; nuôi tế bào chuyển nhiễm dưới các điều kiện thích hợp để biểu hiện protein Env HIV tái tổ hợp; và tinh chế protein Env HIV tái tổ hợp dưới các điều kiện mà cho phép hình thành phức trime được làm ổn định.

Phương án 34 là phương pháp tạo ra protein Env của HIV tái tổ hợp theo bất kỳ trong số các phương án 1 đến 25, bao gồm đưa vào ít nhất một sự thế axit amin tạo ra gốc axit amin được nêu ở vị trí được chọn từ nhóm bao gồm (i)-(vii) vào trình tự protein vỏ HIV mạch chính.

Phương án 35 là phương pháp theo phương án 34, trong đó trình tự nucleotit mã hóa sự thế axit amin được đưa vào axit nucleic mã hóa trình tự protein vỏ HIV mạch chính.

Phương án 36 là phương pháp theo phương án 34 hoặc 35, trong đó trình tự protein vỏ HIV mạch chính được chọn từ nhóm bao gồm: SEQ ID NO: 2; SEQ ID NO: 3; SEQ ID NO: 4; SEQ ID NO: 5; SEQ ID NO: 6; SEQ ID NO: 7; SEQ ID NO: 8; SEQ ID NO: 9; SEQ ID NO: 6 có đột biến của Glu thành Arg ở vị trí 166; SEQ ID NO: 6 có Cys ở các vị trí 501 và 605 và Pro ở vị trí 559 và/hoặc có SEQ ID NO: 10 thay thế axit amin 508-511; SEQ ID NO: 6 có đột biến của Glu thành Arg ở vị trí 166, còn có Cys ở các vị trí 501 và 605 và Pro ở vị trí 559 và/hoặc có SEQ ID NO: 10 thay thế axit amin 508-511; SEQ ID NO: 8 có Cys ở các vị trí 501 và 605 và Pro ở vị trí 559 và/hoặc có SEQ ID NO: 10 thay thế axit amin 508-511; SEQ ID NO: 9 có Cys ở các vị trí 501 và 605 và Pro ở vị trí 559 và/hoặc có SEQ ID NO: 10 thay thế axit amin 508-511; và protein Env HIV kiểu đại có các đột biến mà tạo ra ít nhất (a), (b) hoặc (c), tốt hơn là ít nhất hai trong số (a), (b) và (c), tốt nhất là (a), (b) và (c) của các gốc sau đây: (a) Cys ở các vị trí 501 và 506 và Pro ở vị trí 559,

(b) có SEQ ID NO: 10 thay thế axit amin 508-511, và/hoặc

(c) ít nhất một đột biến sửa ở gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất ít hơn 7,5%, tốt hơn là ít hơn 2%, của trình tự Env HIV trong tập hợp của ít nhất 100, tốt hơn là ít nhất 1000, tốt hơn là ít nhất 10000, trình tự Env HIV kiểu đại, trong đó đột biến sửa là sự thế bằng gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất bằng ít nhất 10% của trình tự Env HIV trong tập hợp này và tốt hơn là đột biến sửa là sự thế bằng gốc axit amin mà có mặt ở vị trí tương ứng thường xuyên nhất trong tập hợp này.

Phương án 37 là chế phẩm bao gồm protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án 1 đến 25, phức trime theo phương án 26, hạt theo phương án 27, phân tử axit nucleic phân lập theo phương án 28, vector theo phương án 29 hoặc 30, hoặc tế bào chủ theo phương án 31, và chất mang được dụng.

Phương án 38 là chế phẩm theo phương án 37, còn bao gồm tá dược.

Phương án 39 là phương pháp tạo ra chế phẩm theo phương án 37, bao gồm trộn protein Env HIV tái tổ hợp, phức trime, particle, axit nucleic phân lập, vector, hoặc tế bào chủ với một hoặc nhiều chất mang được dụng.

Phương án 40 là phương pháp chủng ngừa đối tượng chống nhiễm HIV bao gồm dùng cho đối tượng chế phẩm bao gồm protein vỏ HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án từ 1 đến 25, phức trime theo phương án 26, hạt theo phương án 27, hoặc vector theo phương án 29 hoặc 30.

Phương án 41 là phương pháp tạo ra đáp ứng miễn dịch chống nhiễm HIV ở đối tượng cần chủng, bao gồm việc dùng cho đối tượng chế phẩm chứa protein vỏ HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án từ 1 đến 25, phức trime theo phương án 26, hạt theo phương án 27, hoặc vector theo phương án 29 hoặc 30.

Phương án 42 là protein Env HIV tái tổ hợp có chứa trình tự axit amin nêu ở SEQ ID NO: 2, hoặc trình tự mà là ít nhất 95% tương đồng với nó.

Phương án 43 là protein Env HIV tái tổ hợp có chứa trình tự axit amin nêu ở SEQ ID NO: 3, hoặc trình tự mà là ít nhất 95% tương đồng với nó.

Phương án 44 là protein Env HIV tái tổ hợp có chứa trình tự axit amin nêu ở SEQ ID NO: 4, hoặc trình tự mà là ít nhất 95% tương đồng với nó.

Phương án 45 là protein Env HIV tái tổ hợp có chứa trình tự axit amin nêu ở SEQ ID NO: 5, hoặc trình tự mà là ít nhất 95% tương đồng với nó.

Phương án 46 là phân tử axit nucleic phân lập mã hóa protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án từ 42 đến 45.

Phương án 47 là vector chứa phân tử axit nucleic phân lập theo phương án 46 liên kết hoạt động với trình tự khởi đầu.

Phương án 48 là vector theo phương án 47, mà là vector adenovirut.

Phương án 49 là tế bào chủ bao gồm phân tử axit nucleic phân lập theo phương án 46 hoặc vector theo phương án 47 hoặc 48.

Phương án 50 là chế phẩm bao gồm protein Env HIV tái tổ hợp theo một phương án bất kỳ trong số các phương án từ 42 đến 45, phân tử axit nucleic phân lập theo phương án 46, vector theo phương án 47 hoặc 48, hoặc tế bào chủ theo phương án 49, và chất mang được dùng.

Phương án 51 là phương pháp cải thiện phần trăm trime và/hoặc hiệu suất trime (thể hiện sự gấp và tính ổn định) của protein Env HIV bố mẹ, phương pháp bao gồm việc sửa trình tự axit amin của protein Env HIV bố mẹ bằng cách đưa vào ít nhất một đột biến sửa, tốt hơn là ít nhất 3 đột biến sửa trong protein Env HIV bố mẹ, trong đó đột biến sửa là sự thế axit amin ở gốc axit amin mà thấy ở vị trí tương ứng ở tần suất ít hơn 7,5%, tốt hơn là ít hơn 2%, của trình tự Env HIV trong tập hợp của ít nhất 100, tốt hơn là ít nhất 500, tốt hơn là ít nhất 1000, tốt hơn là ít nhất 10000, trình tự Env HIV kiểu đại, trong đó sự thế là bằng gốc axit amin mà thấy ở vị trí tương ứng ở tần suất bằng ít nhất

10% của trình tự Env HIV trong tập hợp này và tốt hơn là sự thế là bằng gốc axit amin mà thấy ở vị trí tương ứng thường xuyên nhất trong tập hợp này.

Phương án 52 là phương pháp theo phương án 51, trong đó ít nhất 50%, tốt hơn là ít nhất 80%, của gốc axit amin trong protein Env HIV bố mẹ mà thấy ở các vị trí tương ứng ở tần suất ít hơn 7,5% của trình tự Env HIV trong tập hợp này được sửa.

Phương án 53 là phương pháp theo phương án 51, trong đó ít nhất 50%, tốt hơn là ít nhất 80%, của gốc axit amin trong protein Env HIV bố mẹ mà thấy ở các vị trí tương ứng ở tần suất ít hơn 2% của trình tự Env HIV trong tập hợp này được sửa.

Phương án 54 là phương pháp theo một phương án bất kỳ trong số các phương án từ 51 đến 53, trong đó protein Env HIV bố mẹ là từ nhánh C.

Phương án 55 là phương pháp theo một phương án bất kỳ trong số các phương án từ 51 đến 54, trong đó protein Env HIV bố mẹ là protein Env HIV kiểu dại.

Phương án 56 là phương pháp theo một phương án bất kỳ trong số các phương án từ 51 đến 54, trong đó protein Env HIV bố mẹ một hoặc nhiều trong số sau đây:

- (a) Cys ở các vị trí 501 và 506 và Pro ở vị trí 559;
- (b) đột biến trong trình tự phân cắt furin của protein Env HIV, ví dụ có SEQ ID NO: 10 thay thế axit amin 508-511;
- (c) Phe ở vị trí 651;
- (d) Ile ở vị trí 655;
- (e) Asn ở vị trí 535;
- (f) Val ở vị trí 589;
- (g) Phe ở vị trí 573;
- (h) Ile ở vị trí 204;
- (i) Phe ở vị trí 647;
- (j) Val ở vị trí 658;
- (k) Gln hoặc Glu ở vị trí 588; và/hoặc
- (l) Pro ở vị trí 556, 558, hoặc 556 và 558.

Phương án 57 là protein Env của HIV tái tổ hợp có thể thu được bằng phương pháp theo một phương án bất kỳ trong số các phương án 51 đến 56.

Ví dụ thực hiện sáng chế

Ví dụ 1: Tạo ra trình tự phổ biến nhánh C và nhánh B của vỏ HIV

Trình tự phổ biến nhánh C của vỏ HIV

Trình tự phổ biến của protein vỏ (Env) nhánh C của HIV được phát triển là trình tự mạch chính đến nghiên cứu hiệu quả của nhiều đột biến khác nhau lên sự tạo thành trimer của protein Env HIV. Sự sắp hàng trình tự của 3.434 trình tự protein vỏ từ thể phân lập virus HIV đã biết được tải từ cơ sở dữ liệu Los Alamos (<http://www.hiv.lanl.gov/content/index>). Từ 3.434 trình tự, 1.252 trình tự của chỉ nhánh C được chọn lọc để tạo ra trình tự phổ biến của protein Env nhánh C của HIV. Ở các vị trí mà gốc phổ biến có thể không được nhận diện một cách rõ ràng dựa trên sự sắp hàng, trình tự phổ biến được dùng để nhận diện trình tự kiểu đại gần nhất bằng cách tra cứu BLAST. Gốc phổ biến ở các vị trí này sau đó lựa chọn là axit amin trong trình tự kiểu đại gần nhất được nhận diện từ tra cứu BLAST. Trình tự phổ biến nhánh C Env HIV được thể hiện trong SEQ ID NO: 2. Hai trình tự với sự tương đồng cao nhất với SEQ ID NO: 2 sử dụng BLAST là trình tự với số Genbank ADM30337.1 và ADM30340.1, đều có 90% tương đồng trình tự với SEQ ID NO: 2.

Trình tự phổ biến nhánh C Env của HIV được cải biến thêm bằng cách đưa vào cái gọi là đột biến SOSIP, mà bao gồm gốc cysteine ở các vị trí 501 và 605 và gốc proline ở vị trí 559, cũng như tối ưu hóa vị trí phân cắt furin bằng cách thay thế vị trí furin ở gốc 508-511 với 6 gốc arginine. Ngoài ra, Val ở vị trí 295 được làm đột biến thành Asn (V295N), để tạo ra vị trí glycosyl hóa liên kết N có trong phần lớn chủng HIV và có thể cải thiện liên kết với các kháng thể nhất định được dùng trong một số thử nghiệm. Hơn nữa, đầu C bị cắt cụt ở gốc 664, tạo ra trình tự mã hóa protein gp140 HIV tan được. Tất cả các vị trí thế/biến đổi được mô tả ở trên là so với đánh số trong gp160 của thể phân lập HXB2 của HIV-1. Trình tự gp140 HIV tạo thành, gọi là “ConC_SOSIP,” được thể hiện trong (SEQ ID NO: 3). Trình tự ConC_SOSIP được dùng làm trình tự vỏ HIV mạch chính hoặc bố mẹ mà các đột biến bổ sung, ví dụ, thế axit amin đơn và đôi, được đưa vào đó để tạo ra protein Env của HIV tái tổ hợp theo các phương án của sáng chế.

Trình tự phổ biến nhánh B vỏ HIV

Trình tự phổ biến nhánh B Env HIV được tạo ra bằng cách sử dụng quy trình tương tự như được mô tả ở trên để tạo ra trình tự phổ biến nhánh C Env của HIV. Trình tự phổ biến nhánh B được tạo ra bằng cách sử dụng 1,708 trình tự protein vỏ nhánh B từ các thể phân lập virus nhánh B đã biết. Trình tự phổ biến nhánh B Env HIV được thể hiện trong SEQ ID NO: 4.

Trình tự phổ biến nhánh B Env HIV được cải biến thêm bằng cách đưa vào cái gọi là đột biến SOSIP, tối ưu hóa vị trí phân cắt furin bằng cách thay thế vị trí furin với 6 gốc arginin, và cắt cụt đầu C ở gốc 664, như được mô tả ở trên, tạo ra trình tự mã hóa trình tự phổ biến nhánh B gp140 HIV tan được. Trình tự protein Env gp140 HIV tạo thành, gọi là “ConB_SOSIP” được thể hiện trong (SEQ ID NO: 5).

Phát hiện một cách ngạc nhiên là các phân tử trên cơ sở phổ biến có mức biểu hiện cải thiện so với các phân tử dựa trên thể phân lập tự nhiên, và hơn nữa đã có mức trime hóa cải thiện. Do đó, các phân tử có SEQ ID NO: 2-5 đã có các đặc điểm có lợi đáng ngạc nhiên.

Ví dụ 2: Biểu hiện và tinh chế protein Env HIV tái tổ hợp

Protein Env của HIV tái tổ hợp được biểu hiện và được tinh chế làm protein gp140 tan được. Các đột biến đơn (sự thay thế axit amin) và kết hợp của chúng (ví dụ, các đột biến đôi và ba) được đưa vào trình tự phổ biến mạch chính ConC_SOSIP để tạo ra chuỗi các biến thể protein Env của HIV tái tổ hợp.

Sự tạo thành và biểu hiện cấu trúc Env gp140 HIV và các biến thể

ADN mã hóa trình tự phổ biến Env nhánh C của HIV ConC_SOSIP được thể hiện trong SEQ ID NO: 3 được tổng hợp và được tối ưu hóa bộ ba mã hóa ở GenScript (Piscataway, NJ 08854) hoặc Gene Art (Life Technologies, Carlsbad, CA). Trình tự bộ ba mã hóa được tối ưu hóa sau đó được tách dòng vào vector pcDNA2004 để tạo ra cấu trúc Env gp140 nhánh C của HIV, mà được dùng làm trình tự vỏ HIV mạch chính để đưa vào đột biến khác. Các đột biến được đưa vào trình tự mạch chính ConC_SOSIP bằng đột biến hướng điểm và phản ứng chuỗi polymeraza (PCR) được thực hiện trên cấu trúc Env gp140 nhánh C của HIV pcDNA2004. Tế bào HEK-Expi293F hoặc tế bào HEK293F được chuyển nhiễm tạm thời với 90% vector pcDNA2004 mã hóa trình tự ConC_SOSIP hoặc biến thể của chúng và 10% vector pcDNA2004 mã hóa furin proteaza (furin-pcDNA2004) theo hướng dẫn của nhà sản xuất. Tế bào chuyển nhiễm được nuôi cấy trong 5 ngày ở 37°C và 10% CO₂. Dịch nổi nuôi cấy được quay trong 10 phút ở 1250 x g. Dịch nổi đã quay sau đó được lọc vô trùng bằng cách sử dụng bộ lọc chân không 0,22 µm và được lưu trữ ở 4°C đến khi sử dụng tiếp theo.

Để biểu hiện ở định dạng 96 giếng tế bào được nuôi cấy trong 3 ngày ở 37°C và 10% CO₂. 4 µL Optimem (môi trường nuôi cấy) được trộn với 4 µL 100 ng/µL ADN và 8 µL hỗn hợp Expi293F (54 µL/mL Optimem) như được thêm và ủ trong 20 phút. Sau

đó 200 uL/giếng tế bào Expi293F được thêm ở tỷ lệ $2,5 \times 10^6$ tế bào/mL. Dịch nổi nuôi cấy được thu hoạch và quay trong 5 phút ở 300 g để loại bỏ các tế bào và mảnh tế bào. Tiếp theo, dịch nổi được quay được lọc vô trùng bằng cách sử dụng máy lọc chân không 0,22 μ m và lưu trữ tại 4°C cho đến khi sử dụng.

Tinh chế Protein Env gp140 của HIV

Protein Env gp140 của HIV được biểu hiện từ vector pcDNA2004 được tinh chế theo nguyên tắc tinh chế hai bước bằng cách sử dụng cột *Galantus nivalis*-lectin (Vectorlabs, AL-1243) để tinh chế ban đầu, và cột tăng Superdex200 (GE) trong bước tiếp theo để loại tạp chất dư. Đối với bước đầu sử dụng cột *Galantus nivalis*-lectin, dịch nổi nuôi cấy được pha loãng với chất đệm (40 mM Tris, 500 mM NaCl pH 7,5) và được đưa qua cột 4 mL CV Tricorn 10-50 Lectin Agarose ở tốc độ 4 mL mỗi phút. Sau đó, cột được rửa bằng bốn thể tích cột chất đệm (40 mM Tris, 500 mM NaCl pH 7,5) và được tách rửa bằng bốn thể tích cột column 40 mM Tris, 500 mM NaCl, và 1 M mannopyranoside pH 7,5 với dòng hướng lên bằng 1,6 mL/phút, nghĩa là hướng của dòng chảy thay đổi từ dưới lên trên để tăng tốc độ rửa giải của protein vỏ và giảm thể tích rửa giải. Chất rửa giải được cô bằng cách sử dụng thiết bị cô quay (50K, Amicon Ultra, Millipore).

Protein Env gp140 của HIV được tinh chế thêm trên cột Superdex200 bằng cách sử dụng 50 mM Tris, 150 mM NaCl pH 7,4 làm chất đệm chạy. Định thứ hai mà được tách rửa từ cột chứa Protein Env gp140 của HIV. Các phân đoạn chứa đỉnh này được gom, và nhận diện của đỉnh được xác nhận là protein Env gp140 của HIV bằng cách sử dụng phân tích Western blot và SDS-PAGE, và/hoặc SEC-MALS. Nồng độ của protein Env gp140 của HIV được tinh chế được xác định bằng cách đo mật độ quang ở 280 nm, và protein Env gp140 của HIV được tinh chế được lưu trữ ở 4°C đến khi sử dụng tiếp theo.

Phân tích SDS-PAGE và Western Blot

Dịch nổi nuôi cấy tế bào chứa các mẫu protein Env gp140 của HIV được biểu hiện và protein Env gp140 của HIV được tinh chế được phân tích trên 4-12% (khối lượng/thể tích) Bis-Tris NuPAGE gel, 1X MOPS (Life Technologies) dưới các điều kiện khử hoặc không khử, và được tạo blot bằng cách sử dụng công nghệ iBlot (Life Technologies). Tất cả các quy trình này được thực hiện theo hướng dẫn của nhà sản xuất. Để phân tích độ tinh khiết, gel được nhuộm bằng vết protein hồng ngoại Krypton

(Thermo Scientific) hoặc vết protein Rubi SYPRO (Bio-Rad). Để phân tích Western blot, các màng được dò bằng kháng thể kháng-6x-thể Histidin (kháng-His-HRP). Gel và màng blot được quét trên thiết bị Odyssey (Li-Cor), và ảnh được phân tích bằng cách sử dụng phần mềm Odyssey 3.0 (Li-Cor).

Sự tạo ảnh sự tạo thành trime Env của HIV bằng Kính hiển vi điện tử vết âm Kính hiển vi điện tử vết âm (NS-EM) được sử dụng để tạo ảnh trime của protein vỏ có trình tự mạch chính ConC_SOSIP, mà được tinh chế bằng cách sử dụng *Galanthus nivalis* lectin sau đó là sắc ký loại kích cỡ, và được tiến hành như được mô tả trong Julien et al. 2015 (*Proc. Natl. Acad. Sci.* (2015) 112(38) 11947-52). Các mẫu trime được pha loãng đến khoảng từ 0,01-0,5 mg/mL trong muối đệm Tris (TBS), pH 7,4 và được gắn lên lưới sàng 200 Cu phủ cacbon (EMS CF200-Cu) mà được thoát sáng trong không khí 2×10^{-1} mbar, 25 mA, 30 giây, trước khi sử dụng. Sau đó, giọt 3 μ L mẫu trime được pha loãng được áp dụng lên lưới trong 1 phút sau đó là blot với giấy lọc (Whatman số 1 hoặc 4). Lưới được làm khô trong một phút, sau đó được nhuộm với 3 μ L uranyl axetat 2,3% (UAc) trong 60 giây. Dữ liệu được gom bằng cách sử dụng kính hiển vi điện tử FEI Tecnai F20 hoạt động ở 120 keV, với độ phóng đại bằng 25,000x mà tạo ra kích thước điểm ảnh bằng 4,68 Å ở mặt phẳng mẫu. Ảnh được thu bằng Gatan BM ultrascan. Hầu hết tất cả các hạt trong ảnh (của vật liệu giàu trime này) là các trime đóng được tạo thành (dữ liệu không thể hiện).

Ví dụ 3: Sàng lọc các biến thể Env gp140 HIV tái tổ hợp về hiệu suất trime và tỷ lệ phần trăm tạo thành trime

Các biến thể protein Env HIV tái tổ hợp được tạo ra trong Ví dụ 2 được sàng lọc về sự tạo thành trime để nhận diện các đột biến mà cải thiện tỷ lệ phần trăm của trime tạo thành và/hoặc hiệu suất trime cải thiện so với trình tự mạch chính ConC_SOSIP. Sàng lọc kỹ tỷ lệ phần trăm trime và hiệu suất trime được thực hiện bằng cách sử dụng thử nghiệm AlphaLISA để đánh giá liên kết của bảng kháng thể HIV trung hòa rộng (bNAb) và không-bNAb với protein Env của HIV tái tổ hợp. Kết quả của thử nghiệm AlphaLISA được xác nhận bởi sắc ký loại kích cỡ và tán xạ ánh sáng đa góc (SEC-MALS).

Phân tích thử nghiệm AlphaLISA®

Tổng biểu hiện của protein Env gp140 của HIV và tổng lượng của trime âm gấp nếp đúng trong hơn 200 biến thể gp140 HIV với thể axit amin đơn được đưa vào trình

tự ConC_SOSIP được tạo ra như được mô tả trong Ví dụ 2 được đo trong dịch nổi nuôi cấy tế bào bằng thử nghiệm AlphaLISA. Các biến thể gp140 HIV chứa các đột biến đôi và ba cũng được thử. Protein Env HIV có trình tự ConC_SOSIP mà không có bất kỳ đột biến được thử để so sánh.

Kháng thể đơn dòng sau đây (mAbs) không kể những cái khác được dùng để phân tích: mAb PGT145, mAb PGDM1400, mAb PG16, mAb PGT151, mAb 35O22, mAb PGT128, mAb PG9, mAb F105, mAb B6, mAb 447-52d, mAb 14e, và mAb 17b. MAbs 447-52D (AB014), PG9 (AB015), và PG16 (AB016) được mua từ Polymun Scientific Immunbiologische Forschung GmbH (Klosterneuburg, Úc). Kháng thể không trung hòa b6 được lấy từ Dennis R. Burton (The Scripps Research Institute, La Jolla, CA), và kháng thể không trung hòa 14e được lấy từ James E. Robinson (Tulane University, New Orleans, LA). Đối với mAbs PGT145 (PDB: 3U1S), PGDM1400 (PDB: 4RQQ), PGT151 (PDB: 4NUG), 35O22 (PDB: 4TVP), F105 (PDB: 1U6A), PGT128 (PDB: 3TYG), và 17b (PDB: 4RQS) axit nucleic mã hóa trình tự công bố được tách dòng vào vector biểu hiện và được tạo ra để đánh giá protein Env HIV. Ngoại trừ mAbs F105, B6, 447-52d, 14e, và 17b, các kháng thể được dùng để phân tích là các kháng thể trung hòa rộng (bNAb). bNAb có khả năng trung hòa nhiều chủng virus HIV. Trong số bNAb, PGT145, PGDM1400, và PG16 là các chất liên kết đỉnh và đặc hiệu trime. PGT151 cũng đặc hiệu trime, nhưng liên kết ở giao diện của hai gen khởi động gp120 và gp41, và phụ thuộc phân cắt. Liên kết của không-bNAb là chỉ báo gấp nếp không chính xác hoặc hình dạng trime mở.

Sự gấp nếp protein được thử nghiệm bằng cách đo sự liên kết của biến thể protein Env gp140 HIV tan được với kháng thể (mAb 17b) đã biết là liên kết với vị trí liên kết đồng thụ thể của protein vỏ HIV, mà chỉ được bộc lộ ra sau sự liên kết của CD4 (dữ liệu không được thể hiện). Cụ thể là, thụ thể tan được CD4 (sCD4) được sử dụng kết hợp với mAb 17 để đánh giá thay đổi hình dạng gây CD4. Sự liên kết của mAb 17b với biến thể protein Env gp140 HIV mà không có sự liên kết CD4 trước với protein vỏ là dấu hiệu của protein vỏ không gấp nếp một phần hoặc được khởi động trước (tức là, Env không ổn định mà tuân theo hình dạng “mở” khi không có sự liên kết CD4).

Đối với thử nghiệm AlphaLISA, cấu trúc Env gp140 HIV trong vector pcDNA2004 chứa chất liên kết sau đó là thẻ sortaza A sau đó là thẻ Flag sau đó là chất liên kết (G₄S)₇ linh hoạt và kết thúc bằng thẻ His, được điều chế (trình tự của thẻ, mà

được đặt ở đầu C của protein Env HIV, được đề xuất trong SEQ ID NO: 19). Cấu trúc Env gp140 HIV được biểu hiện trong tế bào HEK-Expi293, mà được nuôi cấy trong ba ngày trong tấm 96 giếng (200 μ L/giếng). Dịch nổi thô được pha loãng 120 lần trong chất đệm AlphaLISA (PBS + 0,05% Tween-20 + 0,5 mg/mL BSA). Đối với thử nghiệm trên cơ sở mAb 17b, dịch nổi được pha loãng 12 lần. Sau đó, 10 μ L mỗi dịch pha loãng được chuyển sang tấm 96 giếng và được trộn với 40 μ L hạt nhận, hạt cho, và một trong số các mAb nêu trên. Hạt cho liên hợp với ProtA (Cat#: AS102M, Lot#1831829, Perkin Elmer), mà liên kết với mAb. Hạt nhận liên hợp với kháng thể kháng His (Cat#: AL128M, Perkin Elmer), mà liên kết với thể His của cấu trúc. Để định lượng tổng sản lượng protein, bao gồm tất cả các dạng của protein vỏ, kết hợp của hạt cho liên hợp niken (Cat#: AS101M, Perkin Elmer) để dò thể His cùng với hạt nhận liên hợp với thể kháng Flag (Cat#: AL112R, Perkin Elmer) để dò thể Flag được sử dụng. Đối với thử nghiệm sử dụng mAb 17b kết hợp với sCD4-His, kết hợp của hạt cho ProtA và hạt nhận kháng Flag được sử dụng (dữ liệu không thể hiện). Một mẫu được trộn với hạt cho và hạt nhận để phát hiện sự tạo thành trime, và mẫu thứ hai của cùng biến thể Env được trộn với hạt cho liên hợp với niken và hạt nhận liên hợp kháng Flag để đo tổng lượng protein được biểu hiện (tức là, tổng hiệu suất protein).

Hỗn hợp của dịch nổi chứa protein Env gp140 của HIV được biểu hiện, mAb, hạt cho, và hạt nhận được ủ ở nhiệt độ phòng trong 2 giờ không lắc. Sau đó, tín hiệu phát quang được đo bằng thiết bị đọc tấm Synergy NEO (BioTek). Tín hiệu nền trung bình góp phần giả được tế bào chuyển nhiễm được trừ từ đếm AlphaLISA được đo cho mỗi biến thể Env gp140 HIV. Sau đó, toàn bộ dữ liệu được chia cho tín hiệu được đo đối với protein Env HIV có tín hiệu trình tự mạch chính ConC_SOSIP để chuẩn hóa tín hiệu cho mỗi trong số các biến thể Env gp140 HIV được thử cho mạch chính. Dữ liệu liên kết đối với mỗi trong số các biến thể Env gp140 HIV với mAb đặc hiệu trime PGT145 được dùng để xác định tỷ lệ phần trăm tạo thành trime và hiệu suất trime đối với mỗi trong số các biến thể. Liên kết với các mAb khác được sử dụng để đánh giá mẫu liên kết chung của các biến thể Env HIV với bNAb và không-bNAb (không thể hiện).

tỷ lệ phần trăm tạo thành trime đối với mỗi trong số các biến thể Env HIV được tính toán bằng cách chia tín hiệu phát quang chuẩn hóa thu được từ hỗn hợp mẫu của biến thể Env HIV, mAb PGT145, hạt cho liên hợp ProtA, và hạt nhận liên hợp kháng-

His, cho tín hiệu phát quang chuẩn hóa thu được từ hỗn hợp mẫu của biến thể Env HIV, hạt cho liên hợp kháng His và hạt nhận liên hợp kháng Flag.

Hiệu suất trime đối với mỗi trong số các biến thể Env HIV được xác định so với hiệu suất trime đối với protein Env HIV có trình tự mạch chính ConC_SOSIP mà không có bất kỳ các đột biến bổ sung. Tín hiệu phát quang chuẩn hóa thu được từ liên kết của mAb PGT145 với protein vỏ ConC_SOSIP được thiết lập ở 1, và tín hiệu phát quang chuẩn hóa thu được từ liên kết của mAb PGT145 với mỗi trong số các protein gp140 HIV được chuẩn hóa đến giá trị này.

Các kết quả của phân tích thử nghiệm AlphaLISA - Tỷ lệ phần trăm trime và Hiệu suất Trime

Tỷ lệ phần trăm tạo thành trime như được xác định bởi thử nghiệm AlphaLISA đối với một vài sự thể axit amin đơn, đôi, và ba từ danh sách (i)-(vii) trong Bảng 1 ở trên trong trình tự mạch chính ConC_SOSIP được thể hiện trong Hình 2A. Trong số khoảng 200 biến thể Env gp140 HIV chứa thể axit amin đơn mà được thử, bảy vị trí thể được nhận diện mà tỷ lệ phần trăm trime tạo thành tăng thêm ít nhất 25% so với tỷ lệ phần trăm của trime tạo thành đối với trình tự mạch chính ConC_SOSIP mà không có thêm bất kỳ sự thể axit amin.

Các kết quả được thể hiện trong Hình 2A chứng minh rằng bảy vị trí thể được ưu tiên mà tăng đáng kể về tỷ lệ phần trăm tạo thành trime được quan sát thấy bao gồm N651, K655, I535, D589, I573, A204, và E647 theo đánh số trong gp160 của thể phân lập HXB2 của HIV-1. Cụ thể là, thể axit amin đơn mà tạo ra tỷ lệ phần trăm tạo thành trime cải thiện nhất bao gồm N651F, K655I(/F/W) (mặc dù cngx có một thử nghiệm trong đó K655F không thấy tạo ra cải thiện), I535N, D589V(/A), I573F, A204I, E647F. Một số đột biến mà được thử trong kết hợp với một vài trong số các đột biến này, bao gồm K588Q/E, I556P và A558P, và chúng cải thiện hơn nữa tỷ lệ phần trăm trime của các thể đột biến với axit amin được ưu tiên ở các vị trí theo sáng chế ((i)-(vii) của Bảng 1) trong thử nghiệm này.

Tất cả thể đôi được thử trong thử nghiệm này có tỷ lệ phần trăm tạo thành trime cao hơn so với thể đơn tương ứng, và tất cả thể ba được thử có tỷ lệ phần trăm tạo thành trime cao hơn so với các đột biến đơn và đôi tương ứng (Hình 2A). Các kết quả không đoán được và ngạc nhiên này cho thấy rằng các đột biến này có thể thể hiện dạng hiệu trợ trong các thử nghiệm này về sự trime hóa của protein vỏ.

Ngoài ra để cải thiện tỷ lệ phần trăm tạo thành trime, hiệu suất trime tăng cũng có thể mong muốn. Do đó, hiệu suất trime của các biến thể gp140 HIV chứa các đột biến đơn, đôi, và ba trong trình tự mạch chính ConC_SOSIP cũng được xác định bằng thử nghiệm AlphaLISA. Các kết quả được thể hiện trong Hình 2B Hầu hết các biến thể gp140 HIV chứa các đột biến đơn (các loại trừ là I535N, D589A và D589I), có hiệu suất trime cao hơn so với protein vỏ ConC_SOSIP. Tuy nhiên, phân tích SEC-MALS chính xác hơn của thể đột biến I535N, như được mô tả dưới đây, thể hiện tăng hiệu suất trime. Hơn nữa, các đột biến bổ sung kết hợp với I535N, như D589V, tạo ra cùng hiệu suất trime được thấy đối với protein vỏ có thể bổ sung cụ thể này không có đột biến I535N. Hiệu suất trime của các biến thể với các đột biến đôi cũng tăng khi mỗi trong số các biến thể đột biến đơn có hiệu suất trime cao hơn so với protein vỏ ConC_SOSIP (Hình 2B

Tỷ lệ phần trăm tạo thành trime đối với các biến thể gp140 HIV với các đột biến đôi trong mạch chính ConC_SOSIP mà được mô tả trước đây trong tài liệu tham khảo cũng được thử, bao gồm thể đôi E64K, T316W được mô tả bởi (De Taeye et al., *supra*), và thể đôi disulfua I204C, A433C được mô tả bởi (Kwon et al., *trước đây*). Thể đôi E64K, T316W tạo ra tỷ lệ phần trăm tạo thành trime thấp hơn so với protein vỏ ConC_SOSIP, tức là, 15% (dữ liệu không thể hiện). Mặc dù thể đôi disulfua I204C, A433C làm tăng tỷ lệ phần trăm trime đến 43% (dữ liệu không thể hiện), thể đôi mới được mô tả ở đây, như I535N/K588E, K588Q/D589V, K655I/K588E, I535N/D589V, I535N/E647F, D589V/K655I, và I535N/K655I (Hình 2A) tạo ra tỷ lệ phần trăm tạo thành trime cao hơn nữa trong thử nghiệm AlphaLISA.

Các đột biến bổ sung (prolin ở gốc 558 và/hoặc 556) cũng được đưa vào mạch chính ConC_SOSIP, và tỷ lệ phần trăm tạo thành trime và hiệu suất trime được đo cho các protein Env gp140 HIV này. Cả hai sự thể đơn Pro ở vị trí 558 hoặc 556, và thể đôi của prolin ở cả vị trí 556 và 558 ngoài các đột biến SOSIP đã chứa trong mạch chính ConC_SOSIP (tức là, Cys ở các vị trí 501 và 605, và Pro ở vị trí 559) làm tăng tỷ lệ phần trăm tạo thành trime và hiệu suất trime (dữ liệu không thể hiện). Thật vậy, việc đưa một hoặc nhiều trong số thể làm ổn định axit amin mới theo sáng chế vào mạch chính ConC_SOSIP còn bao gồm gốc Pro ở các vị trí 558 và/hoặc 556 còn cải thiện tỷ lệ phần trăm tạo thành trime và/hoặc hiệu suất trime (ví dụ Hình 2A, ví dụ A558P/I535N, K655I/L556P, và một vài thể đột biến ba bao gồm đột biến A558P).

Dữ liệu liên kết của các biến thể Env gp140 HIV với các bNAb khác và không-bNAb chứng minh rằng hầu hết các đột biến đơn, đôi và ba được thử mà làm tăng hiệu suất trime và tỷ lệ phần trăm tạo thành trime, như loại được liệt kê trên các Hình 2A và 2B, cũng có liên kết tăng với bNAb, và liên kết giống hoặc giảm với không-bNAb so với lượng liên kết quan sát được với bNAb và không-bNAb đối với protein vỏ HIV có trình tự mạch chính ConC_SOSIP (dữ liệu không thể hiện). Để phát triển vắc-xin, tăng liên kết với bNAb và giảm liên kết với không-bNAb được ưu tiên. Do đó dữ liệu chứng minh rằng protein vỏ HIV bao gồm sự thể aiox amin ở các vị trí (i)-(vii) được nêu ở Bảng 1 ở trên có các đặc tính mong muốn về mẫu liên kết với các kháng thể trung hòa rộng và không trung hòa rộng.

Phân tích SEC-MALS

Phân tích SEC-MALS cũng được dùng để xác minh hiệu suất trime và tỷ lệ phần trăm tạo thành trime cho các biến thể gp140 HIV được sàng lọc sử dụng thử nghiệm AlphaLISA. Các biến thể gp140 HIV được biểu hiện trong nuôi cấy quy mô 30 mL và được tinh chế bằng cách áp dụng dịch nổi không tế bào lên 200 μ l hạt *Galanthus nivalis* lectin (Vectorlab Cat# AL-1243) trong cột dòng trọng lực Polyprep (Biorad Cat# 731-1550). Các hạt được rửa bằng 2 ml chất đệm liên kết (40 mM Tris, 500 mM NaCl pH 7,4). Protein được rửa giải bằng cách sử dụng 250-500 μ l 40 mM Tris, 500 mM NaCl, 1 M mannopyranosit pH 7,4. Hệ sắc ký lỏng hiệu năng cao (Agilent Technologies) và dụng cụ MiniDAWN TREOS (Wyatt) được liên hợp với Máy dò chỉ số khúc xạ Optilab T-rEX (Wyatt) được dùng để thực hiện thử nghiệm SEC-MALS. Tổng cộng, hoặc là 100 μ l dung dịch lectin hoặc xấp xỉ 30 μ g protein được áp dụng lên cột TSK-Gel G3000SWxl (Tosoh Bioscience) được cân bằng chất đệm chạy (150 mM natri phosphat, 50 mM NaCl, pH 7,0) ở 1 mL/phút. Dữ liệu được phân tích bằng cách sử dụng gói phần mềm Astra 6, và việc tính toán khối lượng phân tử được dẫn xuất từ tín hiệu chỉ số khúc xạ.

Sắc phổ SEC-MALS của protein vỏ ConC_SOSIP và các biến thể gp140 HIV chứa các đột biến đơn được thể hiện trong Hình 3. Nói chung, các kết quả thu được từ phân tích SEC-MALS có thể so sánh với và thống nhất với các kết quả thu được từ phân tích AlphaLISA. Sắc phổ của protein vỏ ConC_SOSIP có bốn đỉnh chính, với đỉnh thứ hai được rửa giải ở khoảng 7,3 phút là đỉnh trime. Protein vỏ ConC_SOSIP được xác định là khoảng 27% trime. Sự hình thành chất kết tụ và monome cho thấy rằng có một

số lỗi gấp nếp và không ổn định liên quan đến protein Env gp140 của HIV có trình tự phổ biến ConC_SOSIP. Như được chứng minh bởi sắc phổ được thể hiện trong Hình 3, tất cả các sự thể đơn tạo ra đỉnh trime tương đối cao hơn so với đỉnh trime cho protein vỏ ConC_SOSIP, chỉ ra rằng hiệu suất trime tăng đối với mỗi trong số các các biến thể gp140 HIV.

Lấy cùng nhau, các kết quả chứng minh rằng sự thể axit amin được nhận diện trong (i)-(vii) của Bảng 1 ở đây tạo cho protein Env của HIV tái tổ hợp tỷ lệ phần trăm tạo thành trime cải thiện và/hoặc hiệu suất trime cải thiện. Cụ thể là, các biến thể protein Env của HIV có nhiều sự thể ở vị trí được nhận diện của (i)-(vii) của Bảng 1, như kết hợp của hai hoặc nhiều các đột biến được nhận diện thường thể hiện hiệu suất trime và/hoặc tỷ lệ phần trăm tạo thành trime cải thiện hơn nữa so với các biến thể protein Env của HIV chỉ có đột biến đơn, mà thể hiện hiệu quả hiệp trợ có thể của các đột biến kết hợp (i)-(vii) của Bảng 1. Theo hiểu biết của tác giả, không có kết hợp nào trong số các kết hợp thể axit amin này đã được báo cáo trong các trình tự protein vỏ HIV có trong tự nhiên, và tất cả các kết hợp (trong khoảng (i)-(vii)) do đó được tin là kết hợp đột biến làm ổn định trime mới. HIV protein vỏ có tỷ lệ phần trăm tạo thành trime tăng, như protein vỏ HIV tái tổ hợp theo sáng chế, có lợi từ triển vọng sản xuất, như đối với vắc-xin, vì cần ít hơn tinh chế và loại bỏ protein vỏ có trong chế phẩm ở hình dạng không tự nhiên không mong muốn. Cũng vậy, tổng hiệu suất biểu hiện trime tăng là có lợi để sản xuất sản phẩm vắc-xin.

Ví dụ 4: Tính ổn định của protein vỏ HIV trime

Tính ổn định nhiệt của protein Env của HIV tái tổ hợp theo các phương án của sáng chế được thử AlphaLISA và phân tích nhiệt quét vi sai (DSC).

Đo tính ổn định nhiệt sử dụng AlphaLISA

Tính ổn định nhiệt được thử bằng cách đo sự mất trime nguyên vẹn khi xử lý nhiệt dựa trên liên kết với mAb PGT145 đặc hiệu trime. Dịch nổi thô (20 μ l) được gia nhiệt ở 60°C trong 1 giờ. Các mẫu sau đó được ly tâm ở tốc độ tối đa trong năm phút để loại chất kết tụ. Thử nghiệm AlphaLISA được tiến hành như được mô tả ở trên trong Ví dụ 3.

Các kết quả được thể hiện trong Hình 4, và dữ liệu được báo cáo là tỷ lệ phần trăm của trime vẫn nguyên vẹn sau xử lý nhiệt. Từ kết quả, ít có thể thấy là hầu hết trong số các protein Env gp140 HIV tái tổ hợp đột biến đơn theo sáng chế được thử có tính ổn

định nhiệt cao hơn so với protein vỏ ConC_SOSIP. Protein vỏ HIV có sự thể hai và ba làm ổn định trime được xác định ở đây mà được thử were cũng được thấy là có tính ổn định nhiệt cao hơn so với protein vỏ ConC_SOSIP.

Đo tính ổn định nhiệt bằng cách sử dụng DSC

Nhiệt độ nóng chảy (T_m) của các biến thể Env gp140 HIV được xác định bằng DSC sử dụng hệ DSC mao dẫn MicroCal. Mỗi phép đo được tiến hành với nhiệt độ khởi đầu bằng 20°C và nhiệt độ kết thúc bằng 110°C ở tốc độ quét bằng 100°C/giờ. Mẫu protein với nồng độ bằng 0,5 mg/mL (400 μ L) được dùng cho mỗi lần đo. Dữ liệu được phân tích bằng cách sử dụng phần mềm Origin J. (công cụ phân tích MicroCal VP).

Nhiệt độ nóng chảy (T_m) của protein vỏ ConC_SOSIP được đo bằng cách sử dụng DSC được xác định là 69,8°C và nhiệt độ bắt đầu nóng chảy là 60,1°C. T_m được đo cho protein vỏ ConC_SOSIP là cao hơn so với cho protein vỏ BG505_SOSIP (protein vỏ HIV của thể phân lập virus BG505 có cái gọi là đột biến SOSIP), mà được báo cáo là có T_m bằng 67,0°C (Kwon et al, 2015). Điều này chỉ ra rằng protein vỏ HIV có trình tự mạch chính ConC_SOSIP có các đặc tính có lợi hơn về tính ổn định nhiệt so với trình tự vỏ HIV đã biết khác với đột biến làm ổn định trime.

T_m của thể đột biến K655I của ConC_SOSIP được đo là 72,3°C và nhiệt độ bắt đầu nóng chảy là 63,7°C, mà cao hơn nữa so với T_m của protein vỏ ConC_SOSIP. T_m của thể đột biến A558P, N651F, I535N của ConC_SOSIP được đo là 77,29°C với nhiệt độ khởi đầu bằng 74,87°C. Các kết quả DSC do đó xác nhận các kết quả về tính ổn định nhiệt được tạo ra bằng thử nghiệm AlphaLISA.

Lấy cùng nhau, các kết quả chứng minh rằng protein Env của HIV bao gồm ít nhất một trong số các thể amin được mô tả ở đây thường có tính ổn định nhiệt cao hơn so với protein vỏ không có các đột biến này. Các kết quả còn chứng minh rằng tất cả các biến thể protein Env của HIV thể đôi có tính ổn định nhiệt cao hơn so với protein vỏ ConC_SOSIP. Các biến thể protein Env của HIV thể ba cũng ổn định hơn so với protein vỏ ConC_SOSIP.

Ví dụ 5: Các biến thể protein vỏ của HIV tái tổ hợp dựa trên trình tự phổ biến protein vỏ nhánh B

Protein Env của HIV tái tổ hợp theo các phương án của sáng chế bao gồm sự thể axit amin đơn (I535N, D589V, N651F hoặc K655I) được đưa vào trình tự phổ biến nhánh B ConB_SOSIP (SEQ ID NO: 5) được tạo ra và được tinh chế như được mô tả

trong Ví dụ 2. Hiệu suất trime và tỷ lệ phần trăm tạo thành trime được đo bởi thử nghiệm AlphaLISA như được mô tả trong Ví dụ 3.

Các kết quả được thể hiện trong Hình 5A (tỷ lệ phần trăm tạo thành trime) và Hình 5B (hiệu suất trime). Các giá trị được báo cáo là so với giá trị được đo cho protein vỏ ConB_SOSIP, mà được thiết lập ở 1 đối với cả tỷ lệ phần trăm tạo thành trime và hiệu suất trime. Các kết quả thể hiện rằng tất cả trong số các đột biến được thử làm tăng tỷ lệ phần trăm tạo thành trime. Hiệu suất trime gần như giống hoặc cải thiện so với protein vỏ ConB_SOSIP đối với tất cả trong số các đột biến được thử.

Các kết quả này chứng minh rằng các đột biến này cũng có hiệu quả làm ổn định lên protein vỏ, ví dụ, cải thiện hiệu suất trime, cải thiện tỷ lệ phần trăm tạo thành trime, v.v., khi được đưa vào trình tự protein vỏ HIV mạch chính khác, trong trường hợp này trình tự phổ biến có nguồn gốc từ Nhánh B.

Ví dụ 6: Các biến thể protein vỏ của HIV tái tổ hợp dựa trên trình tự protein vỏ tổng hợp

Protein Env của HIV tái tổ hợp theo các phương án của sáng chế bao gồm sự thay thế axit amin được đưa vào protein vỏ HIV tổng hợp (tên là 'DS_sC4_SOSIP_E166R') có trình tự được thể hiện trong SEQ ID NO: 7 được điều chế và được tinh chế như được mô tả trong Ví dụ 2. Protein vỏ HIV tổng hợp DS_sC4_SOSIP_E166R có cái gọi là đột biến SOSIP (Cys ở gốc 501 và 605, và Pro ở gốc 559), Cys ở gốc 201 và 433 tạo ra việc đưa vào cầu disulfua (DS), và Arg ở vị trí 166 để làm ổn định đỉnh. Ngoài ra, protein bị cắt đứt ở vị trí 655. Tỷ lệ phần trăm tạo thành trime và hiệu suất trime được đo bởi thử nghiệm AlphaLISA như được mô tả trong Ví dụ 3.

Các kết quả được thể hiện trong Hình 6, mà so sánh tỷ lệ phần trăm tạo thành trime đối với mỗi trong số các biến thể được thử về tỷ lệ phần trăm tạo thành trime (Hình 6A) và hiệu suất trime (Hình 6B) đối với mạch chính DS_sC4_SOSIP_E166R. Tỷ lệ phần trăm tạo thành trime lớn hơn được quan sát thấy đối với mỗi trong số các biến thể được thử so với trình tự mạch chính.

Bên cạnh E166R, một số axit amin hiếm xảy ra khác được thay đổi thành axit amin phổ biến hơn ở vị trí tương ứng trong tập hợp của protein Env của HIV kiểu đại (A114Q, E117K, T375S và I434M), để 'sửa' protein theo khung được giải thích cụ thể hơn trong Ví dụ 12 dưới đây và Hình 12. Trong protein 'được sửa này, đột biến làm ổn định A204I, và K655I cải thiện sC4_SOSIP hơn nữa (Hình 13).

Kết quả của Ví dụ này là thống nhất với kết quả của Ví dụ 5 trong việc chứng minh rằng các đột biến được mô tả ở đây cũng có hiệu quả làm ổn định lên protein vỏ, ví dụ, cải thiện tỷ lệ phần trăm tạo thành trime, và/hoặc cải thiện hiệu suất trime, v.v., khi được đưa vào trình tự protein vỏ của HIV mạch chính khác nhau, trong trường hợp này trình tự Env, tổng hợp, không phổ biến.

Ví dụ 7: Các kết hợp khác của đột biến Env của HIV

Protein Env của HIV tái tổ hợp theo các phương án của sáng chế bao gồm sự thể axit amin được đưa vào ConC_SOSIP (có trình tự được thể hiện trong SEQ ID NO: 3) được điều chế và được tinh chế như được mô tả trong Ví dụ 2. Tỷ lệ phần trăm tạo thành trime được đo bởi thử nghiệm AlphaLISA như được mô tả trong Ví dụ 3. Sau đó, việc lựa chọn kết hợp nhỏ hơn (một cái được thể hiện dưới đây in nghiêng và bổ sung K655I; I535N, D589V; I535N, K655I; D589V, K655I) được tinh chế sử dụng *Galanthus nivalis* lectin và hàm lượng trime được phân tích bằng cách sử dụng SEC-MALS như được mô tả trong Ví dụ 3. Độ ổn định được xác định như được mô tả trong Ví dụ 4.

Các thể đột biến sau đây được điều chế cho thử nghiệm này:

K655I, N651F;

K655I, N651F, E647F;

K655I, N651F, E647F, I535N;

K655I, N651F, I535N;

K655I, I573F;

K655I, D589V, I573F;

K655I, D589V, I573F, N651F;

K655I, D589V, I573F, K588E;

K655I, D589V, I573F, N651F, K588E;

K655I, D589V, I573F, N651F, K588E, I535N;

K655I, D589V, I573F, N651F, K588E, I535N, A204I;

K655I, D589V, I535N, L556P;

K655I, D589V, I573F, N651F, K588E, L556P;

K655I, D589V, A204I;

L556P, N651F;

L556P, N651F, K655I;

L556P, N651F, K655I, I535N;

L556P, N651F, K655I, I535N, I573F;

L556P, N651F, K655I, I535N, I573F, D589V;

L556P, N651F, K655I, I535N, I573F, D589V, A204I;

L556P, N651F, K655I, I535N, I573F, D589V, A204I, K588Q;

L556P, N651F, K655I, I535N, I573F, D589V, A204I, K588Q, E647F;

L556P, N651F, I535N;

L556P, N651F, I535N, I573F;

L556P, N651F, I535N, I573F, D589V;

L556P, N651F, I535N, I573F, D589V, A204I;

L556P, N651F, I535N, I573F, D589V, A204I, K588Q;

L556P, N651F, I535N, I573F, D589V, A204I, K588Q, E647F;

L556P, K655I, I535N;

L556P, K655I, I535N, I573F;

L556P, K655I, I535N, I573F, D589V;

L556P, K655I, I535N, I573F, D589V, A204I;

L556P, K655I, I535N, I573F, D589V, A204I, K588Q;

L556P, K655I, I535N, I573F, D589V, A204I, K588Q, E647F;

L556P, N651F, I535N, I573F, D589V, A204I, K588Q với đột biến SOS được loại ra.

Tất cả các kết hợp thể được thử trong mạch chính ConC_SOSIP thể hiện tỷ lệ phần trăm trime cao hơn, hiệu suất trime cao hơn và tính ổn định trime cao hơn ở 60°C so với mạch chính trong AlphaLISA (dữ liệu không thể hiện). SEC_MALS xác nhận tỷ lệ phần trăm trime cải thiện đối với tất cả đột biến được thử trong mạch chính (dữ liệu không thể hiện).

Đối với bộ bao gồm một đột biến tiếp một đột biến bổ sung đến chín đột biến, SEC-MALS thể hiện rằng bằng việc đưa vào mỗi đột biến bên cạnh tỷ lệ trime/monome tăng, vì chiều cao của đỉnh monome giảm, trong khi đó chiều cao của đỉnh trime giữ nguyên trong đồ thị SEC (Hình 7). Trong số tất cả các biến thể được thử trong SEC-MALS, biến thể với L556P, N651F, K655I, I535N, I573F, D589V, A204I, K588Q, E647F sự thể thể hiện tỷ lệ phần trăm trime cao nhất (monome gp140 ít nhất và monome gp120 ít nhất), tổng sản lượng protein cao nhất và một trong số tính ổn định nhiệt cao hơn. Điều này nghĩa là các đột biến này có thể được kết hợp mà không làm mất trime so với mạch chính. Ngoài ra, điều này đề xuất rằng, nói chung, việc bổ sung các đột biến

được mô tả trong (i)-(vii) của Bảng 1, kết hợp tùy ý với các đột biến được mô tả trong Bảng 2, tạo ra sự trime hóa cải thiện hơn.

Cấu trúc với đột biến L556P, N651F, I535N, I573F, D589V, A204I, K588Q trong đó ‘đột biến SOS’ được loại đi (tức là hai gốc xystein ở các vị trí 501 và 605 được chuyển lại vào gốc axit amin mà có mặt ban đầu trong trình tự nhánh C phổ biến), cũng được thử. Mặc dù tính ổn định nhiệt của nó là thấp, thể đột biến cos tỷ lệ phần trăm và sản lượng trime so sánh được vì thể đột biến tương ứng của mà đã bao gồm đột biến SOS. Thể đột biến trong đó đột biến SOS được loại ra thậm chí có lợi ích là nó liên kết ít hơn các không-bNAb so với đối trọng chứa SOS (có đột biến L556P, N651F, I535N, I573F, D589V, A204I, K588Q). Điều này chứng minh rằng các đặc tính có lợi của sáng chế, như tỷ lệ phần trăm trime hóa cao, có thể thu được ở protein Env của HIV mà không có tất cả các đột biến SOSIP.

Một thể đột biến (được thử trong mạch chính ConC_SOSIP), dựa trên kết hợp của các đặc điểm có lợi về mức biểu hiện, sự tạo thành trime và liên kết với kháng thể trung hòa rộng PGT151, có các đột biến sau đây: L556P, N651F, I535N, I573F, D589V, A204I, K588Q.

Trong nền ConC_SOSIP, 9 sự thể thành công nhất là L556P, E647F, N651F, K655I, I535N, D589V, I573F, và K588E ở gp41 và A204I ở gp120. Kết hợp của tất cả 9 sự thể này dẫn đến tính ổn định, hàm lượng trime và hiệu suất trime tăng. Do việc bổ sung L556P trong biến thể này với 9 sự thể có hiệu quả tương đối hạn chế về tỷ lệ phần trăm trime cải thiện, và sự thể E647F trong bối cảnh này xuất hiện để cản trở liên kết PGT151, hai đột biến này không thường xuyên được sử dụng trong các biến thể khác, và biến thể với 7 sự thể (tên là ConC_SOSIP_7mut, đôi khi ở đây còn gọi là ‘ConC_SOSIP được làm ổn định’ hoặc ‘ConC_cơ sở’; bao gồm N651F, K655I, I535N, D589V, I573F, K588E, và A204I) được thấy là ổn định hơn chút (nhiệt độ nóng chảy tăng) so với biến thể với 9 sự thể nêu trên. Trình tự hoàn thiện của biến thể này (ConC_SOSIP Env, HIV 160544 được làm ổn định) được đề xuất trong SEQ ID NO: 20.

Vào lúc này, đột biến được đặc biệt ưu tiên [được thử trong mạch chính ConC_SOSIP với các đột biến bổ sung sau đây: (a) D279N, A281V, A362Q (tăng sự tương tự với virus thiết lập được truyền, như được mô tả bởi người khác); (b) Del139-152 (loại bỏ vòng biến đổi để làm giảm cơ hội tạo ra kháng thể cho vòng này); và (c)

V295N (đưa vào vị trí glycan mà có mặt trong đa số chủng HIV)], dựa trên kết hợp của các đặc điểm có lợi về mức biểu hiện, sự tạo thành trime và liên kết với kháng thể trung hòa rộng, có đột biến làm ổn định sau đây theo sáng chế: N651F, K655I, I535N, I573F, D589V, A204I, K588E. Trình tự hoàn thiện của biến thể này (ConC_SOSIP.v3 Env được làm ổn định (HIV170654, ConC_SOSIP.v3)) được đề xuất trong SEQ ID NO: 28.

Trong biến thể khác, đột biến K658V được bổ sung vào cấu trúc này (cũng xem Ví dụ 15 dưới đây), mà cải thiện hơn các kết quả.

Ví dụ 8: Các hạt tự lắp ghép thể hiện protein Env HIV được làm ổn định

Các hạt tự lắp ghép feritin và DPS được điều chế mà thể hiện protein Env được làm ổn định theo cách tương tự như được mô tả trong (He et al, 2016). Để làm cho protein gp140 này được dung hợp vào đầu N của hạt thông qua liên kết axit amin ngắn (ví dụ GSG hoặc AAAGS, nhưng các liên kết khác cũng có thể được sử dụng, xem ví dụ He et al, 2016) ở mức ADN và biểu hiện protein dung hợp trong tế bào Expi293F. Một ví dụ của hạt mà được điều chế theo cách này là dựa trên feritin được dung hợp vào protein Env HIV ConC_SOSIP (SEQ ID NO: 3) với các đột biến sau đây: I535N, A558P, D589V, K655I. Các hạt feritin với protein Env này có đột biến V570D bổ sung, mà đã được báo cáo để cải thiện sự trime hóa (Kesavardhana et al, 2014), cũng được điều chế, nhưng quan sát thấy là đột biến này dẫn đến tăng mạnh liên kết của kháng thể không trung hòa (17b), mà là không mong muốn. Env với năm các đột biến này còn được dung hợp vào hai loại hạt DPS, từ *Helicobacter pylori* và từ *Mycobacterium smegmatis* (xem ví dụ WO2011/082087 để điều chế hạt DPS). Env với năm các đột biến này và ngoài ra cầu disulfua đưa vào đột biến đôi I201C-A433C cũng được dung hợp vào feritin.

Các hạt được tinh chế từ dịch nổi không có tế bào vvowis hạt ái lực PGDM140 và các hạt được phân tích bằng cách sử dụng SEC-MALS với cột TSKgel G6000PWCL. SEC-MALS, cũng như Native PAGE (3-12%), xác nhận rằng các hạt với xấp xỉ kích thước mong muốn được tạo thành.

Theo cách tương tự, các hạt nano tự lắp ghép feritin và DPS thể hiện Env HIV có trình tự ConC_SOSIP với kết hợp của các đột biến sau đây: (L556P, N651F, I535N, I573F, D589V, A204I, K588Q), cũng được điều chế.

Các liposome và/hoặc các hạt nano tự lắp ghép khác thể hiện các biến thể Env HIV khác được mô tả ở đây, ví dụ Env HIV có SEQ ID NO: 20, 22, 24, 26, 27, 28, 29, 30, 31, hoặc 32, cũng được điều chế.

Ví dụ 9: Các biến thể protein vỏ của HIV tái tổ hợp dựa trên trình tự protein vỏ nhánh A

Protein Env của HIV tái tổ hợp theo các phương án của sáng chế bao gồm thể axit amin đơn (I535N, D589V, N651F, K655I, I573F, A204I hoặc E647F) được đưa vào protein vỏ của HIV nhánh A kiểu đại với cải biến SOSIP (tên là 'BG505_SOSIP') như được mô tả trong Ví dụ 2. Protein vỏ của HIV BG505_SOSIP có cái gọi là đột biến SOSIP (Cys ở gốc 501 và 605, và Pro ở gốc 559), cũng như Cys khác ở gốc 201 và 433 tạo ra việc đưa vào cầu disulfua (DS), và vị trí N-glycosyl hóa tiềm năng trên vị trí 332 (T332N đột biến). Protein bị cắt đứt ở vị trí 664. Trình tự của BG505_SOSIP được thể hiện trong SEQ ID NO: 21.

Tỷ lệ phần trăm tạo thành trime và hiệu suất trime được đo bởi thử nghiệm AlphaLISA như được mô tả trong Ví dụ 3. Tỷ lệ phần trăm tạo thành trime và hiệu suất trime đối với mỗi trong số các biến thể được thử được so sánh với BG505_SOSIP. Tỷ lệ phần trăm tạo thành trime cao hơn được quan sát thấy đối với sự thể M535N, D589V, N651F hoặc K655I so với trình tự mạch chính (ví dụ Hình 8A). Kết hợp của ví dụ L556P, K655I và M535N thể hiện hiệu suất và tỷ lệ phần trăm trime tăng hơn nữa (ví dụ Hình 8A và 8B). Kết hợp của N651F và D589V cải thiện hiệu suất và tỷ lệ phần trăm trime cao hơn nữa (dữ liệu không thể hiện). Kết quả của Ví dụ này đối với virut nhánh A là thống nhất với kết quả của các ví dụ 10 và 11 (nhánh C) dưới đây và Ví dụ 5 (nhánh B), trong đó các đột biến I535N, D589V, N651F và K655I cũng thể hiện hiệu quả làm ổn định lên protein vỏ có nguồn gốc từ chủng kiểu đại, ví dụ, cải thiện tỷ lệ phần trăm tạo thành trime, và/hoặc cải thiện hiệu suất trime. Rõ ràng là, các đột biến này theo sáng chế cũng cải thiện sự trime hóa của Env HIV có nguồn gốc từ chủng nhánh A kiểu đại.

Vào lúc này, thể đột biến được đặc biệt ưu tiên (được thử trong mạch chính BG505_SOSIP, dựa trên kết hợp của các đặc điểm có lợi về mức biểu hiện, sự tạo thành trime và liên kết với kháng thể trung hòa rộng, là thể đột biến có các đột biến sau đây: L556P, K655I, M535N, N651F, D589V, (xem ví dụ Hình 9, thể hiện cải thiện mạnh sự tạo thành trime của thể đột biến này trong phân tích SEC-MALS, và Hình 13, thể hiện liên kết cải thiện rõ ràng của các kháng thể trung hòa rộng của thể đột biến này). Trình

tự của Env BG505_SOSIP được làm ổn định này (HIV170863) được thể hiện trong SEQ ID NO: 22.

Việc bổ sung đột biến Q658V tạo ra cải thiện khác nhỏ hơn.

Cấu trúc được ưu tiên khác chứa đột biến L556P, K655I, M535N, N651F, D589V, cũng như các đột biến 'DS' (Cys ở các vị trí 201 và 433 tạo ra việc đưa vào cầu disulfua), R588E, và Q658V. Trình tự của biến thể (BG505_SOSIP.v2 Env, HIV171814) được đề xuất trong SEQ ID NO: 29.

Ví dụ 10: Các biến thể protein vỏ của HIV tái tổ hợp dựa trên trình tự protein vỏ kiểu đại nhánh C

Protein Env của HIV tái tổ hợp theo các phương án của sáng chế bao gồm sự thể axit amin đơn T651F, sự thể axit amin đôi T651F, M535N được đưa vào trình tự Env WT C97ZA_SOSIP (SEQ ID NO: 23) với sự thể bổ sung L556P (C97ZA_SOSIP_L556P) được tạo ra và được biểu hiện như được mô tả trong Ví dụ 2. Hiệu suất trime và tỷ lệ phần trăm tạo thành trime được đo bởi thử nghiệm AlphaLISA như được mô tả trong Ví dụ 3.

Các kết quả được thể hiện trong Hình 10A và B. Hiệu suất trime của C97ZA_SOSIP_L556P_T651F_M535N cao hơn năm lần so với của mạch chính C97ZA_SOSIP.

Do đó thể L556P, T651F và M535N tạo ra sự cải thiện lớn của C97ZA_SOSIP, nhưng liên kết với bNAb và tỷ lệ phần trăm trime đối với biến thể dẫn xuất từ kiểu đại nhánh C này vẫn thấp hơn nhiều so với mạch chính ConC_SOSIP. Vì Env kiểu đại có thể được làm cho thích hợp với vật chủ của nó, có thể làm giảm sự phù hợp chung của nó, và do đó sự gấp có thể bị hỏng, trình tự Env được 'sửa' theo khung khái niệm được mô tả trong Ví dụ 12 và trong Hình 12. Tổng số 21 gốc đã thay đổi, để sửa trình tự, và ba vị trí N-glycosyl hóa tiềm năng (PNGS) được thêm để lấp đầy cái gọi là "lỗ glycan" (các vị trí trong đó ít nhất 50% protein Env của chủng HIV kiểu đại có mặt vị trí N-glycosyl hóa tiềm năng). Các đột biến được đưa vào bằng cách tuân theo nguyên tắc này đối với C97ZA_SOSIP được nêu ở Bảng 3 trong cột 'đột biến sửa'. Việc bổ sung đột biến làm ổn định K655I được bộc lộ ở đây làm tăng tỷ lệ phần trăm và hiệu suất trime hơn nữa, như D589V, A204I và K588E.

Các kết quả này chứng minh rằng các đột biến T651F, M535N và K655I, D589V, A204I và K588E được mô tả ở đây cũng có hiệu quả làm ổn định lên protein vỏ, ví dụ,

cải thiện hiệu suất trime, cải thiện tỷ lệ phần trăm tạo thành trime khi được đưa vào C97ZA_SOSIP (có nguồn gốc từ protein Env chủng kiểu đại nhánh C) và các biến thể của chúng.

Vào lúc này, biến thể được đặc biệt ưu tiên (được thử trong mạch chính C97ZA_SOSIP, dựa trên kết hợp của các đặc điểm có lợi về mức biểu hiện, sự tạo thành trime và liên kết với kháng thể trung hòa rộng, là thể đột biến có các đột biến sau đây: Q567K (được mô tả bởi người khác trước đây); A198T, S243N, K236T, V295N (để lắp lỗ glycan); M34L, T46K, T58A, Q171K, G172V, P179L, L183Q, I192R, N209T, M307I, Q350R, N352H, Y353F, D412N, G429E, V455T, I489V, L491I, G500K, S547G, T578A, T651N (để sửa trình tự); V505N, E507T, T663N (vị trí N-glycosyl hóa tiềm năng được thêm ở nền phân tử); và A204I, M535N, L556P, K588E, D589V, T651F, K655I (đột biến làm ổn định theo sáng chế). Dữ liệu cho biến thể này ví dụ được thể hiện trong Hình 13, xem cụ thể là ‘C97ZA được sửa và được làm ổn định’ ở đây), thể hiện tăng mạnh liên kết kháng thể trung hòa rộng so với phân tử Env C97ZA kiểu đại gốc. Trình tự của biến thể này (C97ZA_SOSIP Env được làm ổn định và được sửa (HIV170690)) được đề xuất trong SEQ ID NO: 24.

Việc bổ sung đột biến K658V làm ổn định protein này hơn nữa.

Biến thể được ưu tiên khác bao gồm đột biến ‘DS’ và K658V, và trình tự của biến thể này (C97ZA_SOSIP.v2 Env, HIV171810) được đề xuất trong SEQ ID NO: 30.

Ví dụ 11: Các biến thể protein vỏ của HIV tái tổ hợp dựa trên trình tự protein vỏ kiểu đại nhánh C khác

Trong protein Env từ chủng nhánh C strain Du422, SOSIP các đột biến được đưa vào và hai lỗ glycan được lắp đầy ở vị trí 295 và 386 bằng các đột biến K295N và D386N. Ngoài ra, một số gốc được sửa theo khung khái niệm được mô tả trong Ví dụ 12 và Hình 12 (V272I, W456R, G466E và F643Y), và sự thể làm ổn định L556P, I535N, N651F và D589V được đưa vào. Thể bổ sung tạo ra hiệu suất trime và tỷ lệ phần trăm trime cao hơn (ví dụ Hình 11).

Trong biến thể được thử cụ thể với bốn đột biến làm ổn định này (SEQ ID NO: 25), sự thể K655I bổ sung còn làm tăng hiệu suất trime và tỷ lệ phần trăm trime thêm hệ số 1,3 và 1,4 tương ứng (dữ liệu không thể hiện).

Vào lúc này, biến thể Env Du422_SOSIP được đặc biệt ưu tiên, dựa trên kết hợp của các đặc điểm có lợi về mức biểu hiện, sự tạo thành trimer và liên kết với kháng thể trung hòa rộng, là thể đột biến có các đột biến sau đây: L556P, K655I, M535N, N651F, D589V, K588E, I201C, A433C, V272I, W456R, G466E, F643Y, D386N, và K295N. Trình tự của biến thể này (Du422_SOSIP Env được làm ổn định và được sửa (HIV170859) được đề xuất trong SEQ ID NO: 26. Dữ liệu cho biến thể này ví dụ được thể hiện trong Hình 13, xem Du422 được sửa và được làm ổn định ở đây), thể hiện tăng mạnh liên kết kháng thể trung hòa rộng so với phân tử Env Du422 kiểu đại gốc.

Biến thể được ưu tiên khác còn bao gồm đột biến 'DS' và K658V, và trình tự của biến thể này (Du422_SOSIP.v1 Env, HIV171812) được đề xuất trong SEQ ID NO: 31. Ví dụ 12: Sửa và làm ổn định các trình tự HIV-1 Env khác nhau

Vì trình tự kiểu đại từ virus được phân lập từ bệnh nhân lây nhiễm có thể nhận các đột biến làm mất ổn định mà cản trở sự gấp nếp đúng, trình tự Env kiểu đại của nhánh C C97ZA, DU422 và sC4 khám được sửa đầu tiên.

Để tra cứu cho các đột biến không tối ưu trong trình tự kiểu đại sự sắp hàng tất cả trình tự HIV-1 Env trong cơ sở dữ liệu UniProt và cơ sở dữ liệu Los Alamos HIV (~90.000 trình tự) được thực hiện và sự phân bố axit amin được tính toán đối với mỗi axit amin. Nói chung, số axit amin tương đối hiếm có trong tự nhiên trong trình tự Env kiểu đại được thể thành axit amin thông thường hơn (dựa trên tần suất trong cơ sở dữ liệu ở vị trí tương ứng) theo khung khái niệm được mô tả trong Hình 12.

Ngoài ra, hai thể bổ sung Y353F và Q171K ở đỉnh của C97ZA_SOSIP được đưa vào để cải thiện có thể liên kết của đỉnh nhắm đích kháng thể, và vị trí glycan khác được đưa vào bằng cách thể D411N, K236T và V295N vì các vị trí N-glycosyl hóa tiềm năng này (PNGS) được bảo toàn >50%. Tiếp theo, thể làm ổn định được mô tả trong các ví dụ trước được chuyển sang trình tự được sửa.

ConC_SOSIP được làm ổn định chứa thể A204I, I535N, I573F, K588E, D589V, N651F và K655I (ConC_SOSIP được làm ổn định). Trình tự hoàn thiện của ConC_SOSIP được làm ổn định được nêu trong SEQ ID NO: 20.

Tổng quan của một số trong các protein Env biến thể và các đột biến của chúng được đề xuất trong Bảng 3.

Bảng 3. Các biến thể protein Env của HIV.

Protein	Các đột biến từ tài liệu tham khảo	PNGS được thêm	Các trình tự dẫn (SEQ ID NO:)	Đột biến sửa	Đột biến làm ổn định	Đột biến khác	Đầu tận cùng
ConC_SOSIP		V295N	11				664
ConC_SOSIP được làm ổn định		V295N	11		A204I, I535N, I573F, K588E, D589V, N651F, K655I		664
BG505_SOSIP		T332N	34				664
BG505_SOSIP được làm ổn định		T332N	34		M535N, L556P, D589V, N651F, K655I		664
C97ZA_SOSIP	L535M, Q567K		Tự nhiên				664
C97ZA_SOSIP được sửa	L535M, Q567K	A198T, S243N, K236T, V295N	11	M34L, T46K, T58A, Q171K, G172V, P179L, L183Q, I192R, N209T, M307I, Q350R, N352H, Y353F, D412N, G429E, V455T, I489V, L491I, G500K, S547G, T578A, T651N		V505N, E507T, T663N	664
C97ZA_SOSIP được sửa và được làm ổn định	Q567K	A198T, S243N, K236T, V295N	11	M34L, T46K, T58A, Q171K, G172V, P179L, L183Q, I192R, N209T, M307I,	A204I, M535N, L556P, K588E, D589V, T651F, K655I	V505N, E507T, T663N	664

				Q350R, N352H, Y353F, D412N, G429E, V455T, I489V, L491I, G500K, S547G, T578A, T651N			
Du422_SOSIP		D386N , K295N	11				664
Du422_SOSIP được sửa		D386N , K295N	11	V272I, W456R, G466E, F643Y			664
Du422_SOSIP được sửa và được làm ổn định		D386N , K295N	11	V272I, W456R, G466E, F643Y	M535N, L556P, K588E, D589V, N651F, K655I		664
DS_sC4_SOSI P	I201C- A433C	V295N	33				655
DS_sC4_SOSI P được sửa	I201C- A433C	V295N	33	A114Q, E117K, E166R, T375S, I434M			655
DS_sC4_SOSI P được sửa và được làm ổn định	I201C- A433C	V295N	33	A114Q, E117K, E166R, T375S, I434M	A204I, I535N, L556P, Q588E, D589V, N651F, K655I	delta13 8-152 (SSNG TYNII HNET YK), delta19 1 (SEKS SENSSE), delta 463 (GVP)	655

Bảng 3. Một vài trong số các biến thể protein Env của HIV được mô tả ở đây. Cột ‘các đột biến từ tài liệu tham khảo’ mô tả các đột biến mà được dùng trong các cấu trúc này và được mô tả trước đây bởi người khác. Cột ‘PNGS bổ sung’ mô tả các đột biến mà thêm vị trí N-glycosyl hóa tiềm năng (ở các vị trí mà nhiều protein Env kiểu đại bao gồm vị trí này). Cột ‘trình tự dẫn’ mô tả mà trình tự dẫn được dùng để biểu hiện nếu không phải là trình tự dẫn gốc (tự nhiên). Cột ‘đột biến sửa’ mô tả các đột biến mà cải thiện sự gấp và tính ổn định (được đo là hiệu suất và tỷ lệ phần trăm trime, dựa trên liên kết với bNAb) của một số trong các protein Env kiểu đại, như được mô tả trong Ví dụ 12 và Hình 12. Cột ‘đột biến làm ổn định’ mô tả các đột biến theo sáng chế mà làm ổn định protein và cải thiện sự trime hóa như được bộc lộ ở đây. Cột ‘đột biến khác’ mô tả các đột biến bổ sung được thực hiện cho một số cấu trúc. Cột ‘đầu’ mô tả vị trí của axit amin cuối (đánh số trong toàn bộ Bảng là so với trình tự Env HXB2).

Dịch nổi của tế bào được chuyển nhiễm tạm thời với các biến thể Env kiểu đại (wt), được sửa, và được làm ổn định được thử đối với liên kết với một vài các kháng thể trung hòa rộng đặc hiệu trime nhắm đến đỉnh. Sự thể sửa và đặc biệt là thể làm ổn định có tác động mạnh lên hàm lượng trime (các Hình 13 và 14), được xác định bằng AlphaLISA (Hình 13) và SEC-MALS (Hình 14).

Trình tự của biến thể được ưu tiên của protein Env DS_sC4 được sửa và được làm ổn định (DS_sC4_SOSIP Env được sửa và được làm ổn định (HIV170686)) được đề xuất trong SEQ ID NO: 27.

Trình tự IgG1 được lấy làm ví dụ được đề cập trong SEQ ID NO: 32 (sC4_SOSIP.v4 Env. được sửa và được làm ổn định

Ví dụ 13: Đột biến làm ổn định theo sáng chế đóng vai trò trong sự vắng mặt của các đột biến SOSIP

Như được thể hiện trong các ví dụ trước, 7 đột biến (A204I, I535N, I573F, K588E, D589V, N651F và K655I) cải thiện hiệu suất và tỷ lệ phần trăm trime trong ConC_SOSIP (tạo ra ‘ConC_cơ sở’ hoặc ‘ConC_SOSIP được làm ổn định’ hoặc ‘ConC_SOSIP 7mut’) (ví dụ các Hình 13 và 15).

Ví dụ này chứng minh rằng các đột biến SOSIP khác nhau (tức là đột biến ‘SOS’: 2 sự thể bởi gốc Cys ở các vị trí 501 và 605; và sự thể ‘đột biến IP’: thể bởi gốc Pro ở vị trí 559) góp phần làm ổn định thêm, nhưng không cần thu được lợi ích từ các đột biến theo sáng chế.

7 đột biến được thể hiện cũng cải thiện hiệu suất trime trong cái gọi là ConC_SOS, mà không chứa đột biến I559P làm ổn định (đột biến 'IP'), như được thể hiện trong Hình 15 (so sánh ConC_SOS với ConC_SOS, 7mut). Do đó, đột biến 'IP' không cần thiết để thu được lợi ích từ các đột biến được mô tả ở đây. Việc bổ sung đột biến I559P tạo ra sự tăng lớn, thể hiện rằng đột biến 'IP' có lợi trong cấu trúc này ngoài 7 đột biến theo sáng chế. Đột biến IP làm ổn định (I559P) có thể cũng được thay thế bởi A558P hoặc L556P, cả hai trong số đó cũng tạo ra sự tăng mạnh so với biến thể không có đột biến I559P.

Tương tự ConC_IP, 7 mut, mà chứa 7 đột biến theo sáng chế được mô tả ở trên, nhưng không có đột biến 'SOS', còn thể hiện hiệu suất trime rất cao, chứng minh rằng các đột biến 'SOS' cũng không cần thu được các lợi ích từ các đột biến được mô tả ở đây (ví dụ so sánh ConC_SOSIP với ConC_IP, 7 mut), phù hợp với các quan sát trong ví dụ 7. Việc bổ sung đột biến 'SOS' không làm tăng thêm hiệu suất trime.

Do đó, trong khi các trime Env chứa đột biến làm ổn định được mô tả ở đây có thể có lợi từ việc làm ổn định hơn với các đột biến SOSIP, không có đột biến nào trong số 3 đột biến SOSIP cần để thu được lợi ích (ví dụ cải thiện hiệu suất trime) của đột biến làm ổn định được mô tả ở đây.

Ví dụ 14: Thể methionin ở các vị trí 647, 651 hoặc 655 cải thiện chất lượng trime

Tiếp theo các đột biến được mô tả trong Ví dụ 2, các vị trí 589, 647, 651 và 655 được thể riêng bằng gốc Met trong mạch chính ConC_SOSIP (SEQ ID NO: 3), và được thử về tỷ lệ phần trăm và hiệu suất trime hóa bằng cách sử dụng phương pháp như được mô tả ở trên. Thể hiện rằng Met ở các vị trí 647, 651, hoặc 655, giống các đột biến được mô tả trong Ví dụ 2, cải thiện chất lượng của trime (tỷ lệ phần trăm và hiệu suất trime cao hơn, tăng liên kết bNAb), như có thể được thấy trên Hình 16.

Do đó, ngoài việc thể bằng Phe, Ala, hoặc Trp ở vị trí 651, thể bằng Met ở vị trí 651 còn cải thiện sự tạo thành trime; ngoài việc thể bằng Phe, Ile, hoặc Trp ở vị trí 655, thể bằng Met ở vị trí 655 còn cải thiện sự tạo thành trime; và ngoài việc thể bằng Phe, hoặc Ile ở vị trí 647, thể bằng Met ở vị trí 647 còn cải thiện sự tạo thành trime.

Ví dụ 15: Protein Env HIV với đột biến làm ổn định trime ở vị trí 658.

Protein Env của HIV tái tổ hợp với các đột biến thể ở vị trí 658 (đánh số theo gp160 của thể phân lập HXB2 của HIV-1) được điều chế, trong mạch chính ConC_SOSIP (SEQ ID NO: 3). K658 được đột biến thành Val, Ile, Phe, Leu, Met, hoặc

Ala. Ngoài ra, một số thể đột biến đôi được thực hiện trong đó các đột biến này được kết hợp với một trong số các đột biến làm ổn định được mô tả ở trên, K655I. Tỷ lệ phần trăm tạo thành trime được xác định bằng thử nghiệm AlphaLISA như được mô tả trong Ví dụ 3.

Các kết quả được thể hiện trong Hình 17A và B (tỷ lệ phần trăm trime, được đo trong các thử nghiệm khác nhau, do đó hai bảng) và Hình 17C và D (hiệu suất trime, được đo trong các thử nghiệm khác nhau, do đó hai bảng). Các kết quả này chứng minh rằng sự thế ở vị trí 658 by Ile, Phe, Met, Leu, Ala, hoặc Val tạo ra cải thiện tỷ lệ phần trăm tạo thành trime và cải thiện hiệu suất trime. Việc thế bằng Ile ở vị trí 658 tạo ra sự tăng mà gần giống khoảng như K655I đột biến (Hình 17A, C), mà là thể đột biến hiệu suất tốt nhất từ các đột biến (i)-(vii) in Bảng 1 được mô tả ở trên (xem ví dụ Hình 2A). Sự thế bằng Val ở vị trí 658 tạo ra sự cải thiện cao hơn nữa (Hình 17A, C).

Các kết quả cũng chứng minh rằng sự thế ở vị trí 658 bằng Ile hoặc Val có thể được kết hợp với đột biến K655I mà được mô tả ở trên, và điều này tạo ra cải thiện hơn so với mỗi trong số các thể đột biến đơn tương ứng (Hình 17A, C).

Thể đột biến K658V còn được thử bằng cách sử dụng SEC-MALS. Nuôi cấy 96 giếng được nuôi trong ba ngày như được làm đối với AlphaLISA. Dịch nổi được nạp trực tiếp lên cột SEC-MALS. Sắc phổ thu được cho dịch nổi giả được (với biểu hiện furin) được trừ từ sắc phổ của dịch nổi với protein Env. Protein trime được tách rửa từ cột nằm trong khoảng từ 7 đến 8 phút. Các kết quả được thể hiện trong Hình 18, và xác nhận rằng thể đột biến K658V thể hiện cải thiện sự trime hóa so với protein Env cơ sở, và so với protein Env thể đột biến K655I.

Ví dụ này chứng minh rằng việc thế của axit amin ở vị trí 658 trong protein Env của HIV bằng Val, Ile, Phe, Met, Leu, hoặc Ala tạo ra tỷ lệ phần trăm trime và hiệu suất trime cải thiện.

Các thử nghiệm khác để đo sự tạo thành trime của các biến thể bằng cách sử dụng AlphaLISA và/hoặc SEC-MALS được thực hiện trong các biến thể Env của HIV trong đó đột biến K658V có mặt kết hợp với các đột biến khác từ các Bảng 1 và/hoặc 2 như được mô tả ở đây, cũng như ở các chủng HIV từ nhánh A và B. Ví dụ, đột biến 658V đã thể hiện để cải thiện biến thể ConC_SOSIP_{7mut} như được mô tả ở trên (Ví dụ 7), cũng như BG505_SOSIP với L556P, K655I, M535N, N651F, D589V, K588E (Ví dụ 9), cũng như C97ZA_SOSIP được sửa và được làm ổn định (Ví dụ 10).

Dựa trên các kết quả được mô tả ở trên, kỳ vọng rằng đột biến của axit amin ở vị trí 658 thành Valin, Isoleuxin, Phenylalanin, Leucine, Methionin hoặc Alanin, tốt hơn là thành Valin, gốc sẽ cải thiện sự tạo thành trime và/hoặc hiệu suất trime trong các protein Env của HIV cơ sở khác nhau.

Ví dụ 16. Chủng ngừa bằng protein Env của HIV được làm ổn định

Nghiên cứu chủng ngừa thử được tiến hành với protein Env tan được và với protein Env được liên hợp với liposome. Đoạn môi được thực hiện với ConC_SOSIP.v3 được làm ổn định (SEQ ID NO: 28), và sau đó là bốn liều tăng cường, mỗi liều với một protein khác, tức là với 1) sC4_SOSIP.v4 được sửa và được làm ổn định (SEQ ID NO: 32); 2) C97ZA_SOSIP.v2 được sửa và được làm ổn định (SEQ ID NO: 30); 3) Du422_SOSIP.v1 được sửa và được làm ổn định (SEQ ID NO: 31); và 4) BG505_SOSIP.v2 được làm ổn định (SEQ ID NO: 29).

Huyết thanh được phân lập sau các chủng ngừa liên tiếp, và được phân tích đối với kháng thể được gây mà liên kết đặc biệt với hình dạng ổn định, đóng kín, trước dung hợp của Env (bằng cách sử dụng ELISA), cũng như đối với việc gây bNAb (bằng cách sử dụng thử nghiệm trung hòa virus).

Các ví dụ trên chứng minh rằng sáng chế tạo ra tiếp cận phổ quát để tối ưu hóa sự gấp nếp và tính ổn định của protein trime vỏ HIV đóng trước dung hợp.

Cần hiểu rằng các ví dụ và Phương án được mô tả ở đây chỉ nhằm mục đích minh họa, và các thay đổi có thể được thực hiện đối với các Phương án được mô tả ở trên mà không vượt ra khỏi phạm vi sáng tạo rộng lớn của chúng. Do đó, cần hiểu rằng sáng chế không bị giới hạn ở các Phương án cụ thể được bộc lộ, mà nó bao hàm các cải biến trong tinh thần và phạm vi của sáng chế như được xác định bởi các yêu cầu bảo hộ kèm theo.

DANH MỤC TRÌNH TỰ

SEQ ID NO: 1 gp160 của thể phân lập HXB2 của HIV-1 (trình tự tín hiệu viết nghiêng; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPVWKEATTT
 LFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLVNVTFNFMWKNDMV
 EQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSSGRMIMEKGEI
 KNCSFNISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKVSF
 EPIPIHYCAPAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSL
 AEEEVVIRSVNFTDNAKTIIVQLNTSVEINCTRPNNNTRKRIRIQRGPGRAFTI
 GKIGNMRQAHCNISRAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTH
 SFNCGGEFFYCNSTQLFNSTWFNSTWSTEGSNNTGSDTITLPCRIKQIINMWQ
 KVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESEIFRPGGGDMRDNWR
 ELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVGIGALFLGFLGAAGSTMGA
 A SMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE
 RY LKDQQLGIWGC SGKLICTTAVPWNASWSNKSLEQIWNHTTWMEWDREINN
 YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVG
 GLVGLRIVFAVLSIVNRVRQGYSPFSFQTHLPTPRGPDRPEGIEEEGERDRDR
 SIRLVNGSLALIWD DLRLSLCLFSYHRLRDLLIVTRIVELLGRRGWEALKYWW
 NLLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQG
 LERILL

SEQ ID NO: 2 Nhánh C phổ biến của Env HIV (chỉ có trình tự phổ biến, không bao gồm bất kỳ trình tự tín hiệu, vùng xuyên màng (664 là axit amin cuối), đột biến các SOSIP, và/hoặc đột biến ở vị trí phân cắt furin; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

NLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNWATHACVPTDPNPQ
 EMVLENTENFNMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLNCTN
 VNVNTNTNNNNMKEEMKNCSFNNTTTEIRDKKQKEYALFYRLDIVPLNENSSEY
 RLINCNTSTITQACPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTGPCNNVSTVQ
 CTHGIKPVVSTQLLLNGSLAEEIIIIRSENLTDNAKTIIVHLNESVEINCTRPNNN
 TRKSIRIGPGQTFYATGDIIGDIRQAHCNISEAKWNKTLQRVKKKLKEHFPNKT
 IKFAPSSGGDLEITTHSFNCRGEFFYCNTSKLFNSTYNNNTTSNSTITLPCRIKQIIN
 MWQEVGRAMYAPPIAGNITCKSNITGLLLTRDGGNNNNNTETFRPGGGDMRD
 NWRSELYKYKVVEIKPLGIAPTKAKRRVVEREKRRAVGIGAVFLGFLGAAGST
 MGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQARVL
 AIERYLKDQQLGIWGC SGKLICTTAVPWNSSWSNKSQEDIWDNMTWMQWD
 REISNYTDTIYRLLEESQNQQEKNEKDLLALD

SEQ ID NO: 3 ConC_SOSIP (trình tự phổ biến nhánh C trưởng thành với đột biến SOSIP và vị trí phân cắt furin (viết nghiêng), và cắt cụt đầu C; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

NLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNWATHACVPTDPNPQ
 EMVLENTENFNMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLNCTN

VNVTNTNNNNMKEEMKNCSFNNTTTEIRDKKQKEYALFYRLDIVPLNENSSEY
RLINCNTSTITQACPKVSFDPIPIHYCAPAGYAILKCNKTFNGTGPCNNVSTVQ
CTHGIKPVVSTQLLLNGSLAEEIIIIRSENLTDNAKTIIVHLNESVEINCTRPNNN
TRKSIRIGPGQTFYATGDIIGDIRQAHCNISEAKWNKTLQRVKKKLKEHFPNKT
IKFAPSSGGDLEITTHSFNCRGEFFYCNTSKLFNSTYNNTTSNSTITLPCRIKQIIN
MWQEVGRAMYAPPIAGNITCKSNITGLLLTRDGGNNNNNTETFRPGGGDMRD
NWRSELYKYKVVEIKPLGIPTKCKRRVVERRRRRRAVGIGAVFLGFLGAAGS
TMGAASITLTVQARQLLSGIVQQQSNLLRAPEAQQHMLQLTVWGKQLQARV
LAIRYLKDQQLGIWGCSGKLICCTAVPWNSSWSNKSQEDIWDNMTWMQW
DREISNYTDTIYRLLEESQNQQEKNEKDLLALD

SEQ ID NO: 4 Nhánh B phổ biến của Env HIV (chỉ có trình tự phổ biến, không bao gồm bất kỳ trình tự tín hiệu, vùng xuyên màng (664 là axit amin cuối), đột biến các SOSIP, và/hoặc đột biến ở vị trí phân cắt furin; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

AEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNWVATHACVPTDPN
PQEVVLENTENFNMWKNNMVEQMHEIISLWDQSLKPCVKLTPLCVTLNCT
DLNNNTTNNNSSSEKMEKGEIKNCSFNITTSIRDKVQKEYALFYKLDVVPIDN
NNTSYRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGTGPCTN
VSTVQCTHGIRPVVSTQLLLNGSLAEEVVIRSENFTDNAKTIIVQLNESVEINC
TRPNNNTRKSIHIGPGRAFAYATGDIIGDIRQAHCNISRTKWNNTLKQIVKKLRE
QFGNKTIVFNQSSGGDPEIVMHSFNCGGEFFYCNTTQLFNSTWNSNGTWNNTT
GNDTITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGNNNN
NTTETFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTCKCKRRVVQRRRRRR
AVGIGAMFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQNNLLRAPEAQQ
HLLQLTVWGKQLQARVLAVERYLKDQQLGIWGCSGKLICCTAVPWNTSWS
NKSLDEIWDNMTWMQWEREIDNYTGLIYTLIEESQNQQEKNEQELLELD

SEQ ID NO: 5 ConB SOSIP (trình tự phổ biến nhánh B trưởng thành với đột biến SOSIP và vị trí phân cắt furin (viết nghiêng), và cắt cụt đầu C; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

AEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNWVATHACVPTDPN
PQEVVLENTENFNMWKNNMVEQMHEIISLWDQSLKPCVKLTPLCVTLNCT
DLNNNTTNNNSSSEKMEKGEIKNCSFNITTSIRDKVQKEYALFYKLDVVPIDN
NNTSYRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGTGPCTN
VSTVQCTHGIRPVVSTQLLLNGSLAEEVVIRSENFTDNAKTIIVQLNESVEINC
TRPNNNTRKSIHIGPGRAFAYATGDIIGDIRQAHCNISRTKWNNTLKQIVKKLRE
QFGNKTIVFNQSSGGDPEIVMHSFNCGGEFFYCNTTQLFNSTWNSNGTWNNTT
GNDTITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGNNNN
NTTETFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTCKCKRRVVQRRRRRRRA
VGIGAMFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQNNLLRAPEAQH
LLQLTVWGKQLQARVLAVERYLKDQQLGIWGCSGKLICCTAVPWNTSWSN
KSLDEIWDNMTWMQWEREIDNYTGLIYTLIEESQNQQEKNEQELLELD

SEQ ID NO: 6 Đoạn Mos2S Env C4 của protein vỏ HIV tổng hợp; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

MGNLWVTVYYGVPVWKDAKTTLFCASDAKAYEKEVHNWVATHACVPTDP
NPQEIVLGNVTENFNMWKNDMVDQMHEIISLWDASLEPCVKLTPLCVTLNC

RNVRNVSSNGTYNIIHNETYKEMKNCSFNATTVVEDRKQKVHALFYRLDIVPL
DENNSSEKSSSENSSEYYRLINCNTSAITQACPKVSFDPIPIHYCAPAGYAILKCN
NKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLNGSLAEEEEIIRSENLTNNAKTI
IVHLNETVNITCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHCNLSRDGWN
KTLQGVKKKLAEHFPNKTIKFAPHSGGDLEITHTFNCRGEFFYCNTSNLFNES
NIERNDSIITLPCRIKQIINMWQEVGRAIYAPPIAGNITCRSNITGLLLTRDGGSN
NGVPNDTETFRPGGGDMRNNWRSELYKYKVVEVKPLGVAPTEAKRRVVERE
KRAVGIGAVFLGILGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQ
QHMLQLTVWGIKQLQTRVLAIERYLQDQQLLGLWGCSGKLICTTAVPWNTS
WSNKSQTDIWDNMTWMQWDKEIGNYTGEIYRLLEESQNQQEK

SEQ ID NO: 7 (Trình tự DS_sC4_SOSIP_E166R; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

MGNLWVTVYYGVPVWKDAKTTLFCASDAKAYEKEVHNWVATHACVPTDP
NPQEIVLGNVTENFNMWKNDMVDQMHEIISLWDASLEPCVKLTPLCVTLNC
RNVRNVSSNGTYNIIHNETYKEMKNCSFNATTVVRDRKQKVHALFYRLDIVP
LDENNSSEKSSSENSSEYYRLINCNTSACTQACPKVSFDPIPIHYCAPAGYAILKC
NNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLNGSLAEEEEIIRSENLTNNAK
TIIVHLNETVNINCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHCNLSRDGW
NKTLQGVKKKLAEHFPNKTIKFAPHSGGDLEITHTFNCRGEFFYCNTSNLFNE
SNIERNDSIITLPCRIKQIINMWQEVGRICIYAPPIAGNITCRSNITGLLLTRDGG
NNGVPNDTETFRPGGGDMRNNWRSELYKYKVVEVKPLGVAPTECKRRVVER
RRRRRAVGIGAVFLGILGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAP
EAQQHMLQLTVWGIKQLQTRVLAIERYLQDQQLLGLWGCSGKLICTTAVPW
NTSWSNKSQTDIWDNMTWMQWDKEIGNYTGEIYRLLEESQNQQEK

SEQ ID NO: 8 (Mos1.Env, Trình tự protein vỏ HIV khảm; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

AGKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNWVATHACVPTDPN
PQEVVLENTENFNMWKNNMVEQMHEIISLWDQSLKPCVKLTPLCVTLNCT
DDVRNVNTNATNTNSSWGEPMEKGEIKNCSFNITTSIRNKVQKQYALFYKLD
VVPIDNDSNNTNYRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKF
NGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEEVVIRSENFTNNAKTIMV
QLNVSVEINCTRPNNNTRKSIHIGPGRAFYTAGDIIGDIRQAHCNISRANWNNT
LRQIVEKLGKQFGNNKTIVFNHSSGGDPEIVMHSFNCGGEFFYCNSTKLFNST
WTWNNSTWNNTKRSNDTEEHITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCS
SNITGLLLTRDGGNDTSGTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPT
KAKRRVVQSEKSAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQQ
QNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLI
CTTTVPWNASWSNKSLDKIWNNMTWMEWEREINNYTSLIYTLIEESQNQQEK

SEQ ID NO: 9 (Mos2.Env, trình tự protein vỏ HIV khảm; axit amin ở các vị trí (i)-(vii) đối với các đột biến theo sáng chế được thể hiện bằng cách bôi màu xám)

MGNLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNWVATHACVPTDPN
PQEMVLENTENFNMWKNDMVDQMHEIIRLWDQSLKPCVKLTPLCVTLEC
RNVRNVSSNGTYNIIHNETYKEMKNCSFNATTVVEDRKQKVHALFYRLDIVPL
DENNSSEKSSSENSSEYYRLINCNTSAITQACPKVSFDPIPIHYCAPAGYAILKCN

NKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLNGSLAEEIIIIRSENLTNNAKTI
IVHLNETVNITCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHCNLSRDGWN
KTLQGVKKKLAEHFPNKTINFSSSGGDLEITTHSFNCRGEFFYCNTSGLFNGT
YMPNGTNSNSSSNITLPCRIKQIINMWQEVGRAMYAPPIAGNITCRSNITGLLLT
RDGGSNNGVPNDTETFRPGGGDMRNNWRSELYKYKVVEVKPLGVAPTEAKR
RVVEREKRAVGIGAVFLGILGAAGSTMGAASITLTVQARQLLSGIVQQQSNLL
RAIEAQQHMLQLTVWGIKQLQTRVLAIERYLQDQQLLGLWGCSGKLICTTAV
PWNTSWSNKSQTDIWDNMTWMQWDKEIGNYTGEIYRLLLEESQNQQEK

SEQ ID NO: 10 (trình tự thể đột biến vị trí phân cắt furin)
RRRRRR

SEQ ID NO: 11 (ví dụ của trình tự tín hiệu (ví dụ được dùng cho ConC_SOSIP, và một số biến thể có nguồn gốc kiểu đại))
MRVRGILRNWQQWWIWGILGFWMLMICNVVG (lưu ý: VG cuối có thể là bắt đầu của protein trưởng thành hoặc kết thúc của trình tự tín hiệu)

SEQ ID NO: 12 (ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1)
NPDWLPDM

SEQ ID NO: 13 (ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1)
GSGSGSGS

SEQ ID NO: 14 (ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1)
DDVHPDWD

SEQ ID NO: 15 (ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1)
RDTFALMM

SEQ ID NO: 16 (ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1)
DEEKVMDF

SEQ ID NO: 17 (ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1)
DEDPHWDP

SEQ ID NO: 18 (ví dụ của trình tự tín hiệu (ví dụ được dùng cho ConB_SOSIP))
MRVKGIRKNYQHLWRWGTMLLGMLMICA

SEQ ID NO: 19 (thẻ được dùng cho cấu trúc gp140 HIV trong thử nghiệm AlphaLISA)
AAALPETGGGSDYKDDDDKPGGGSGGGGSGGGGSGGGGSGGGGSGGGGS
GGGSHHHHHH

SEQ ID NO: 20 (ConC_SOSIP được làm ổn định, 'ConC_SOSIP_7mut' (HIV160544))
 NLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNWATHACVPTDPNPQ
 EMVLENTNFNMWKNMVDQMHEDIISLWDQSLKPCVKLTPLCVTLNCTN
 VNVNTNTNNNMKEEMKNCSFNNTTEIRDKKQKEYALFYRLDIVPLNENSSEY
 RLINCNTSTITQICPKVSFDPIPIHYCAPAGYAILKCNKTFNGTGPCNNVSTVQ
 CTHGIKPVVSTOLLNGLAEIEIIRSENLTDAKTIIVHLNESVEINCTRPNNN

TRKSIRIGPGQTFYATGDIIGDIRQAHCNISEAKWNKTLQRVKKKLKEHFPNKT
IKFAPSSGGDLEITTHSFNCRGEFFYCNTSKLFNSTYNNNTTSNSTITLPCRICKQIIN
MWQEVGRAMYAPPIAGNITCKSNITGLLLTRDGGNNNNNTETFRPGGGDMRD
NWRSELYKYKVVEIKPLGIAPTKCKRRVVERRRRRRRAVGIGAVFLGFLGAAGS
TMGAASnTLTVQARQLLSGIVQQQSNLLRAPEAQQHMLQLTVWGfKQLQARV
LAIERYLEvQQLLGIWGC SGK LICCTAVPWNSSWSNKSQEDIWDNMTWMQW
DREISNYTDTIYRLLEESQfQQEiNEKDLLALD

SEQ ID NO: 21 (protein Env BG505_SOSIP (HIV150673))

AENLWVTVYYGVPVWKDAETTLFCASDAKAYETEKHNWATHACVPTDPN
PQEIHLENVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCT
NVTNNITDDMRGELKNCSFNMTTEL RDKKQKVYSLFYRLDVVQINENQGNRS
NNSNKEYRLINCNTSAITQACPKVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPC
PSVSTVQCTHGKIPVVSTQLLLNGSLAE E E VMIRSENITNNAKNILVQFNTPVQI
NCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHCNVSKATWNETLGKVVKQ
LRKHFGNNTIIRFANSSGGDLEVTTTHSFNCGGEFFYCNTSGLFNSTWISNTSVQ
GSNSTGSNDSITLPCRICKQIINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDG
GSTNSTTETFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTRCKRRVVGRRR
RRRAVGIGAVFLGFLGAAGSTMGAASnTLTVQARNLLSGIVQQQSNLLRAPE
AQQHLLKLTWVGKQLQARVLAVERYLRDQQLLGIWGC SGK LICCTNVPWNS
WSNRNLSEIWDNMTWLQWDKEISNYTQIIYGLLEESQNQQEKNEQDLLALD

SEQ ID NO: 22 (protein Env BG505_SOSIP được làm ổn định (HIV170863))

AENLWVTVYYGVPVWKDAETTLFCASDAKAYETEKHNWATHACVPTDPN
PQEIHLENVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCT
NVTNNITDDMRGELKNCSFNMTTEL RDKKQKVYSLFYRLDVVQINENQGNRS
NNSNKEYRLINCNTSAITQACPKVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPC
PSVSTVQCTHGKIPVVSTQLLLNGSLAE E E VMIRSENITNNAKNILVQFNTPVQI
NCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHCNVSKATWNETLGKVVKQ
LRKHFGNNTIIRFANSSGGDLEVTTTHSFNCGGEFFYCNTSGLFNSTWISNTSVQ
GSNSTGSNDSITLPCRICKQIINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDG
GSTNSTTETFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTRCKRRVVGRRR
RRRAVGIGAVFLGFLGAAGSTMGAASnTLTVQARNLLSGIVQQQSNLpRAPEA
QQHLLKLTWVGKQLQARVLAVERYLRvQQLLGIWGC SGK LICCTNVPWNS
WSNRNLSEIWDNMTWLQWDKEISNYTQIIYGLLEESQfQQEiNEQDLLALD

SEQ ID NO: 23 (protein Env C97ZA_SOSIP kiểu đại với L535M và Q567K (HIV150673))

NMWVTVYYGVPVWTDAKTTLFCASDTKAYDREVHNWATHACVPTDPNPQ
EIVLENVTENFNMWKNDMVDQM HEDIISLWDQSLKPCVKLTPLCVTLHCTNA
TFKNNVTNDMNKEIRNCSFNNTTTEIRDKKQQGYALFYRPDIVLLKENRNNSNN
SEYILINCNASTITQACPKVNFDPIPIHYCAPAGYAILKCNNKTFSGKGPCNNVS
TVQCTHGKIPVVSTQLLLNGSLAEKEIIRSENLT DN VKTIIVHLNKSVEIVCTRP
NNNTRKSMRIGPGQTFYATGDIIGDIRQAYCNISGSKWNETLKR VKEKLQENY
NNNKTIKFAPSSGGDLEITTHSFNCRGEFFYCNTTRLFN NN NATEDETITLPCRICK
QIINMWQGVGRAMYAPPIAGNITCKSNITGLLLVRDGGEDNKTEEIFRPGGGN
MKDNWRSELYKYKVIELKPLGIAPT Gc KRRVVERrrrrRAVGIGAVFLGFLGAA

GSTMGAASmTLTVQARQLLSSIVQQQSNLLRApEAQQHMLkLTVWGIKQLQT
RVLAIERYLKDQQLLGIWGC SGK LICcTNVPWNSSWSNKSQTDIWN NM TWME
WDREISNYTDTIYRLLED SQTQQEKNEKD L LALD

SEQ ID NO: 24 (protein Env C97ZA_SOSIP được sửa và được làm ổn định (HIV170690))

NIWVTVYYYGVPVWkDAKTTLFCASDaKAYDREVHN VWATHACVPTDPNPQEI
VLENTENFNMWKNDMVDQM HEDIISLWDQSLKPCVKLTPLCVTLHCTNAT
FKNNVTNDMNKEIRNCSFN TTTTEIRDKKQkvYALFYRIDIVqLKENRNN SNSE
YrLINCNTstITQ iCPK VtFDPIPIHYCAPAGYAILKCNNKTFnGtGPCNNVSTVQC
THGIKPVVSTQ LLLNGSLAEKEIIIRSENLT DN VKTIIVHLNKSVEInCTRPNNNT
RKS iRIGPGQTFYATGDIIGDIRQAYCNISGSKWNETLKR VKEKLrEhfNNNKTIK
FAPSSGGDLEITTHSFNCRGEFFYCNTTRLFN NNATE nETITLPCRIKQIINMWQe
VGRAMYAPPIAGNITCKSNITGLLLtRDGGEDNKTEEIFRPGGGNMKDNWRSE
LYKYKVvEiKPLGIAPTkcKRRnVtRrrrrRAVGIGAVFLGFLGAAGSTMGAASnTL
TVQARQLLSgIVQQQSNLpRApEAQQHMLkLTVWGIKQLQaRVLAIERYL evQQ
LLGIWGC SGK LICcTNVPWNSSWSNKSQTDIWN NM TWMEWDREISNYTDTIY
RLLED SQTQQEiNEKD L LAnD

SEQ ID NO: 25 (biến thể của cấu trúc Du422 được sửa và được làm ổn định (HIV161818))

NLWVTVYYYGVPVWKEAKTTLFCASDAKAYDKEVHN VWATHACVPTDPNPQ
EIVLENTENFNMWKNDMVDQM HEDIISLWDQSLKPCVKLTPLCVTLNCKNV
NISANANATATLNSSMNGEIKNCSFN TTTTEL RDKKQKVYALFYKPDVVPLNG
GEHNETGEYILINCNSSTcTQACPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTG
PCNNVSTVQCTHGIKPVVSTQ LLLNGSLAE EEEIIIRSENLTNNIKTIIVHLNKSVE
INCTRPNNNTRKSVRIGPGQTFYATGEIIGDIREAHCNISRETWNSTLIQVKEKL
REHYNKTIKFEPSSGGDLEVTTTHSFNCRGEFFYCNTTKLFNETKLFNESEYVDN
KTIILPCRIKQIINMWQEVGRcMYAPPIEGNITCKSNITGLLLTRDGGENSTEEVF
RPGGGNMKDNWRSEL YKYKVVEIKPLGVAPTKCKRK nVtRRRRRRRAVGLGA
VLLGFLGAAGSTMGAASnTLTVQARQLLSGIVQQQSNLpRAPEAQQHLLQLTV
WGIKQLQTRVLAIERYLKvQQLLGLWGCSGKLICCTAVPWNSSWSNKS LGDI
WDNMTWMQWDREISNYTNTIYRLLED SQTQQEKNEKD L LAnD

SEQ ID NO: 26 (Du422_SOSIP được sửa và được làm ổn định (HIV170859))

NLWVTVYYYGVPVWKEAKTTLFCASDAKAYDKEVHN VWATHACVPTDPNPQ
EIVLENTENFNMWKNDMVDQM HEDIISLWDQSLKPCVKLTPLCVTLNCKNV
NISANANATATLNSSMNGEIKNCSFN TTTTEL RDKKQKVYALFYKPDVVPLNG
GEHNETGEYILINCNSSTITQACPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTG
PCNNVSTVQCTHGIKPVVSTQ LLLNGSLAE EEEIIIRSENLTNNIKTIIVHLNKSVE
INCTRPNNNTRKSVRIGPGQTFYATGEIIGDIREAHCNISRETWNSTLIQVKEKL
REHYNKTIKFEPSSGGDLEVTTTHSFNCRGEFFYCNTTKLFNETKLFNESEYVDN
KTIILPCRIKQIINMWQEVGRAMYAPPIEGNITCKSNITGLLLTRDGGENSTEEV
FRPGGGNMKDNWRSEL YKYKVVEIKPLGVAPTKCKRK VVGRRRRRRAVGLG
AVLLGFLGAAGSTMGAASnTLTVQARQLLSGIVQQQSNLpRAPEAQQHLLQLT
VWGIKQLQTRVLAIERYL evQQLLGLWGCSGKLICCTAVPWNSSWSNKS LGDI
WDNMTWMQWDREISNYTNTIYRLLED SQTQQEiNEKD L LALD

SEQ ID NO: 27 (DS_sC4_SOSIP được sửa và được làm ổn định (HIV170686))

MGNLWVTVYYGVPVWKDAKTTLFCASDAKAYEKEVHNVWATHACVPTDP
NPQEIVLGNVTENFNMWKNDMVDQMHEDIISLWDqSLkPCVKLTPLCVTLNC
RNVVRNVEMKNCSFNATTVVrDRKQKVHALFYRLDIVPLDENNSSYRLINCNTS
AcTQiCPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTGPCNNVSTVQCTHGIKP
VSTQLLNGSLAEEEEIIIRSENLTNNAKTIIVHLNETVNINCTRPNNNTRKSIRIGP
GQTFYATGDIIGDIRQAHCNLSRDGWNKTLQGVKKKLAEHFPNKTIKFAPHSG
GDLEITTHsFNCRGEFFYCNTSNLNFESNIERNDSIITLPCRIKQIINMWQEVGRc
mYAPPIAGNITCRSNITGLLLTRDGGSSNNNDTETFRPGGGDMRNNWRSELYKY
KVVEVKPLGVAPTECKRRVVERRRRRRRAVGIGAVFLGILGAAGSTMGAASnT
LTVQARQLLSGIVQQQSNLpRAPEAQQHMLQLTVWGIKQLQTRVLAIERYLev
QQLLGLWGCSGKLICTAVPWNTSWSNKSQTDIWDNMTWMQWDKEIGNYT
GEIYRLLEESQfQQEi

SEQ ID NO: 28 (ConC_SOSIP được làm ổn định.v3 (HIV170654))

NLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVPTDPNPQ
EMVLENVTENFNMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLNCTN
VNVTEMKNCSFNNTTTEIRDKKQKEYALFYRLDIVPLNENSSEYRLINCNTSTIT
QICPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTGPCNNVSTVQCTHGIKPVVS
TQLLNGSLAEEEEIIIRSENLTNNVKTIIVHLNESVEIVCTRPNNNTRKSIRIGPGQ
TFYATGDIIGDIRQAHCNISEAKWNKTLQRVKKKLKEHFPNKTIKFQPSGGDL
EITTHSFNCRGEFFYCNTSKLFNSTYNNTTSNSTITLPCRIKQIINMWQEVGRAM
YAPPIAGNITCKSNITGLLLTRDGGNNNNNTETFRPGGGDMRDNRSELYKY
KVVEIKPLGIAPTKCKRRVVERRRRRRRAVGIGAVFLGFLGAAGSTMGAASNTL
TVQARQLLSGIVQQQSNLLRAPEAQQHMLQLTVWGFKQLQARVLAIERYLEV
QQLLGIWGCSGKLICTAVPWNSSWSNKSQEDIWDNMTWMQWDREISNYTD
TIYRLLEESQFQQEINEKDLLALD

SEQ ID NO: 29 (BG505_SOSIP.v2 được làm ổn định (HIV171814))

AENLWVTVYYGVPVWKAETTLFCASDAKAYETEKHNVWATHACVPTDPN
PQEIHLNVTEEFNMWKNNMVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCT
NVTNNITDDMRGELKNCSFNMTTEL RDKKQKVYSLFYRLDVVQINENQGNRS
NNSNKEYRLINCNTSAcTQACPKVSFEPIPIHYCAPAGFAILKCKDKKFNGTGP
CPSVSTVQCTHGIKPVVSTQLLNGSLAEEVMIRSENITNNAKNILVQFNTPV
QINCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHCNVSKATWNETLGKVV
KQLRKHFGNNTIIRFANSSGGDLEVTTHSFNCGGEFFYCNTSGLFNSTWISNTS
VQGSNSTGSNDSITLPCRIKQIINMWQRIGQcMYAPPIQGVIRCVSNITGLILTRD
GGSTNSTTETFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTRCKRRVVGR
RRRRRAVGIGAVFLGFLGAAGSTMGAASnTLTVQARNLLSGIVQQQSNLpRAPE
AQQHLLKLTVWGIKQLQARVLAVERYLevQQLLGIWGCSGKLICTNVPWNS
SWSNRNLSEIWDNMTWLQWDKEISNYTQIIYGLLEESQfQQEiNEvDLLALD

SEQ ID NO: 30 (C97ZA_SOSIP.v2 được sửa và được làm ổn định (HIV171810))

NIWVTVYYGVPVWkDAKTTLFCASDaKAYDREHNVWATHACVPTDPNPQE
VLENVTENFNMWKNDMVDQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNAT
FKNNVTNDMNKEIRNCSFNNTTTEIRDKKQkvYALFYRIDIVqLKENRNNNSN
SEYrLINCNTStTcTQiCPKVtFDPIPIHYCAPAGYAILKCNNKTFnGtGPCNNVSTVQ

THGIKPVVSTQLLNGSLAEKEIIIRSENLTDNVKTIIVHLNKSVEInCTRPNNNT
 RKSIRIGPGQTFYATGDIIGDIRQAYCNISGSKWNETLKRVEKELrEhfNNNKTIK
 FAPSSGGDLEITTHSFNCRGEFFYCNTTRLFNNNATE_nETITLPCRICKQIINMWQ_e
 VGR_cMYAPPIAGNITCKSNITGLLLtRDGGEDNKTEEIFRPGGGNMKDNWRSEL
 YKYKV_vEiKPLGIAPT_kcKRR_nVtRrrrrRAVGIGAVFLGFLGAAGSTMGAAS_nTLT
 VQARQLLSgIVQQQSNL_pRApEAQQHMLkLTVWGIKQLQaRVLAIERYLevQQL
 LGIWGCSGKLIC_cTNVPWNSSWSNKSQTDIWNMTWMEWDREISNYTDTIYR
 LLEDsqfQQEiNEvDLLAnD

SEQ ID NO: 31 (Du422_SOSIP.v1 được sửa và được làm ổn định (HIV171812))

NLWVTVYYGVPVWKEAKTTLFCASDAKAYDKEVHNWVATHACVPTDPNPQ
 EIVLENTENFNMWKNDMVDQMHEIISLWDQSLKPCVKLTPLCVTLNCKNV
 NISANANATATLNSSMNGEIKNCSFNNTTEL RDKKQKVYALFYKPDVVPLNG
 GEHNETGEYILINCNSST_cTQACPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTG
 PCNNVSTVQCTHGIKPVVSTQLLNGSLAEKEIIIRSENLTNNIKTIIVHLNKSVE
 INCTRPNNNTRKSVRIGPGQTFYATGEIIGDIREAHCNISRETWNSTLIQVKEKL
 REHYNKTIKFEPSSGGDLEVTTTHSFNCRGEFFYCNTTKLFNETKLFNESEYVDN
 KTIILPCRICKQIINMWQEVGR_cMYAPPIEGNITCKSNITGLLLTRDGGENSTEEVF
 RPGGGNMKDNWRSELKYKYVVEIKPLGVAPTKCKRKVVGRRRRRRRAVGLG
 AVLLGFLGAAGSTMGAAS_nTLTVQARQLLSGIVQQQSNL_pRAPEAQQHLLQLT
 VWGIKQLQTRVLAIERYLEvQQLLGLWGCSGKLICCTAVPWNSSWSNKS LGDI
 WDNMTWMQWDREISNYTNTIYRLLEDsqfQQEiNEvDLLALD

SEQ ID NO: 32 (sC4_SOSIP.v4 được sửa và được làm ổn định)

MGNLWVTVYYGVPVWKDAKTTLFCASDAKAYEKEVHNWVATHACVPTDP
 NPQEIVLGNVTENFNMWKNDMVDQMHEIISLWD_qSLkPCVKLTPLCVTLNC
 RNVNRNVMKNCSFNATTVV_rDRKQKVHALFYRLDIVPLDENNSSYRLINCNTS
 AcTQiCPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTGPCNNVSTVQCTHGIKPV
 VSTQLLNGSLAEKEIIIRSENLTNNAKTIIVHLNETVNIVCTRPNNNTRKSIRIGP
 GQTFYATGDIIGDIRQAHCNLSRDGWNKTLQGVKKKLAEHFPNKTIKFAPHSG
 GDLEITTHsFNCRGEFFYCNTSNLFNESNIERNDSIITLPCRICKQIINMWQEVGR_c
 mYAPPIAGNITCRSNITGLLLTRDGGSSNNNDTETFRPGGGDMRNNWRSELYKY
 KVVEVKPLGVAPTECKRR_nVtRRRRRRRAVGIGAVFLGILGAAGSTMGAAS_nTL
 TVQARQLLSGIVQQQSNL_pRAPEAQQHMLQLTVWGIKQLQTRVLAIERYLEdQ
 QLLGLWGCSGKLICCTAVPWNTSWSNKSQTDIWDNMTWMQWDKEIGNYTG
 EIYRLLEESQfQQEi

SEQ ID NO: 33 (ví dụ của trình tự tín hiệu (ví dụ được dùng cho các biến thể
 DS_sC4_SOSIP)

MRVVRGMLRNWQQWWIWSSLGFWMLMIYSV

SEQ ID NO: 34 (ví dụ của trình tự tín hiệu (ví dụ được dùng cho các biến thể
 BG505_SOSIP)

MRVMGIQRNCQHILFRWGTMLGMIHICSA

TÀI LIỆU THAM KHẢO

1. Sanders et al. J. Virol. (2002) 76(17), 8875-89
2. Sanders et al. Science (2015) 349(6224), 139-140

3. Julien et al. *Proc. Nat. Acad. Sci.* (2015) 112(38), 11947-52
4. de Taeye et al. *Cell* (2015) 163(7), 1702-15
5. Kwon et al. (2015) *Nat. Struct. Mol. Biol.* 22(7) 522-31
6. Eglen et al. *Curr. Chem. Genomics*, (2008) 25(1), 2-10
7. Kong et al, *Nat Commun.* 2016 Jun 28;7:12040. doi: 10.1038/ncomms12040
8. Julien et al. *Proc. Natl. Acad. Sci.* (2015) 112(38) 11947-52
9. Barouch et al, *Nat Med* 2010, 16: 319-323
10. WO 2010/059732
11. Đơn sáng chế Châu Âu số EP15200138.4
12. Sharma SK, et al. *Cell Rep.* (2015) 11(4):539-50. doi: 10.1016/j.celrep.2015.03.047.
13. Georgiev IS, et al. *J Virol.* (2015) 89(10):5318-29. doi: 10.1128/JVI.03451-14.
14. López-Sagaseta J, et al (2016) *Computational và Struct Biotechnol J* 14: 58-68.
15. Zhao L, et al (2014) *Vaccine* 32: 327-337
16. He L, et al (2016) *Nat Commun.* 2016 Jun 28;7:12041. doi: 10.1038/ncomms12041
17. WO 2011/082087
18. Kesavardhana A và Varadarajan R (2014) *J Virol* 88: 9590-9604
19. Guenaga J, et al (2015) *Immunity* 46: 792-803
20. Bale S, et al (2017) *J. Virol.* doi:10.1128/JVI.00443-17
21. Abbink et al (2007) *Virol.* 81(9): 4654-64
22. Altschul SF et al (1997) *Nucleic Acid Res.* 25: 3389-3402
23. Harris et al (2011) *PNAS* 108 (28): 11440-11445
24. Kushnir et al (2012) *Vaccine* (31): 58-83
25. WO 2007/104792
26. WO 2014/124301
27. US 2016/0122392

YÊU CẦU BẢO HỘ

1. Protein vỏ (Env) của virus gây suy giảm miễn dịch tái tổ hợp ở người (Human Immunodeficiency Virus - HIV), bao gồm hai hoặc nhiều trong số các gốc axit amin sau đây:

- (i) Phe, Leu, Met, hoặc Trp, tốt hơn là Phe, ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp, tốt hơn là Ile, ở vị trí 655;
- (iii) Asn hoặc Gln, tốt hơn là Asn, ở vị trí 535;
- (iv) Val, Ile hoặc Ala, tốt hơn là Val, ở vị trí 589;
- (v) Phe hoặc Trp, tốt hơn là Phe, ở vị trí 573;
- (vi) Ile ở vị trí 204; và/hoặc
- (vii) Phe, Met, hoặc Ile, tốt hơn là Phe, ở vị trí 647,

trong đó việc đánh số các vị trí là theo đánh số trong gp160 của HXB2 thể phân lập HIV-1.

2. Protein Env của HIV tái tổ hợp, bao gồm một hoặc nhiều trong số các gốc axit amin sau đây:

- (i) Phe, Leu, Met, hoặc Trp, tốt hơn là Phe, ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp, tốt hơn là Ile, ở vị trí 655;
- (iii) Asn hoặc Gln, tốt hơn là Asn, ở vị trí 535;
- (iv) Val, Ile hoặc Ala, tốt hơn là Val, ở vị trí 589;
- (v) Phe hoặc Trp, tốt hơn là Phe, ở vị trí 573;
- (vi) Ile ở vị trí 204; và/hoặc
- (vii) Phe, Met, hoặc Ile, tốt hơn là Phe, ở vị trí 647,

trong đó protein Env HIV được chọn từ nhóm bao gồm:

- (1) trình tự đồng nhất Env HIV, ví dụ từ nhánh C ví dụ bao gồm trình tự axit amin nêu ở SEQ ID NO: 2 hoặc 3 hoặc ví dụ từ nhánh B ví dụ bao gồm trình tự axit amin nêu ở SEQ ID NO: 4 hoặc 5;

(2) protein Env HIV tổng hợp, ví dụ bao gồm trình tự axit amin của (a): SEQ ID NO: 6; hoặc (b): SEQ ID NO: 6 với đột biến của Glu thành Arg ở vị trí 166; hoặc (c): (a) hoặc (b) với đột biến của axit amin ở các vị trí 501 và 605 vào gốc Cys và đột biến của axit amin ở vị trí 559 vào gốc Pro; hoặc (d): (a), (b) hoặc (c) có đột biến vị trí phân cắt furin nữa, ví dụ thay thế ở các vị trí 508-511 bằng RRRRRR (SEQ ID NO: 10); hoặc (e) SEQ ID NO: 7; hoặc (f) SEQ ID NO: 8 hoặc SEQ ID NO: 9; và

(3) protein Env HIV bố mẹ mà tốt hơn là protein Env HIV kiểu đại, tốt hơn là của nhánh C, bao gồm ít nhất một đột biến cặp ở gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất ít hơn 7,5%, tốt hơn là ít hơn 2%, của trình tự Env HIV trong tập hợp của ít nhất 1000, tốt hơn là ít nhất 10000, trình tự Env HIV kiểu đại, trong đó đột biến cặp là sự thế bằng gốc axit amin mà có mặt ở vị trí tương ứng ở tần suất bằng ít nhất 10% của trình tự Env HIV trong tập hợp này và tốt hơn là đột biến cặp là sự thế bằng gốc axit amin mà có mặt ở vị trí tương ứng thường xuyên nhất trong tập hợp này;
và

trong đó việc đánh số các vị trí là theo đánh số trong gp160 của HXB2 thể phân lập HIV-1.

3. Protein Env của HIV tái tổ hợp, bao gồm một hoặc nhiều trong số các gốc axit amin sau đây:

- (i) Phe, Leu, Met, hoặc Trp, tốt hơn là Phe, ở vị trí 651;
- (ii) Phe, Ile, Met, hoặc Trp, tốt hơn là Ile, ở vị trí 655;
- (iii) Asn hoặc Gln, tốt hơn là Asn, ở vị trí 535;
- (iv) Val, Ile hoặc Ala, tốt hơn là Val, ở vị trí 589;
- (v) Phe hoặc Trp, tốt hơn là Phe, ở vị trí 573;
- (vi) Ile ở vị trí 204; và/hoặc
- (vii) Phe, Met, hoặc Ile, tốt hơn là Phe, ở vị trí 647,

trong đó protein Env HIV là protein Env HIV bao gồm ít nhất một của các gốc sau đây:

(a) Cys ở các vị trí 501 và 605;

(b) Pro ở vị trí 559;

(c) Cys ở các vị trí 501 và 605 và Pro ở vị trí 559; và

việc đánh số các vị trí là theo đánh số trong gp160 của HXB2 thể phân lập HIV-1.

4. Protein Env HIV tái tổ hợp theo điểm 2 hoặc 3, bao gồm hai hoặc nhiều trong số gốc axit amin được nêu ở (i) đến (vii).
5. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1, 2, và 4, bao gồm Cys ở các vị trí 501 và 605 hoặc Pro ở vị trí 559, tốt hơn là Cys ở các vị trí 501 và 605 và Pro ở vị trí 559.
6. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 5, bao gồm ba hoặc nhiều trong số gốc axit amin được nêu ở (i) đến (vii).
7. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 5, bao gồm bốn hoặc nhiều trong số gốc axit amin được nêu ở (i) đến (vii).
8. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 7, bao gồm Phe ở vị trí 651, Ile ở vị trí 655, Asn ở vị trí 535, và Val ở vị trí 589.
9. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 8, bao gồm năm hoặc nhiều trong số gốc axit amin được nêu ở (i) đến (vii).
10. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 9, còn bao gồm một hoặc nhiều trong số các gốc sau đây:

(viii) Gln, Glu, Ile, Met, Val, Trp, hoặc Phe, tốt hơn là Gln hoặc Glu, ở vị trí 588;

(ix) Lys ở vị trí 64 hoặc Arg ở vị trí 66 hoặc Lys ở vị trí 64 và Arg ở vị trí 66;

(x) Trp ở vị trí 316;

(xi) Cys ở cả vị trí 201 và 433;

(xii) Pro ở vị trí 556 hoặc 558 hoặc ở cả vị trí 556 và 558;

(xiii) thay thế vòng ở vị trí axit amin 548-568 (vòng HR1) bằng vòng có 7-10 axit amin, tốt hơn là vòng có 8 axit amin, ví dụ có trình tự được chọn từ một trình tự bất kỳ trong số (SEQ ID NO: 12-17);

(xiv) Gly ở vị trí 568, hoặc Gly ở vị trí 569, hoặc Gly ở vị trí 636, hoặc Gly ở cả vị trí 568 và 636, hoặc Gly ở cả vị trí 569 và 636; và/hoặc
 (xv) Tyr ở vị trí 302, hoặc Arg ở vị trí 519, hoặc Arg ở vị trí 520, hoặc Tyr ở vị trí 302 và Arg ở vị trí 519, hoặc Tyr ở vị trí 302 và Arg ở vị trí 520, hoặc Tyr ở vị trí 302 và Arg ở cả vị trí 519 và 520.

11. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 10, còn bao gồm đột biến trong trình tự phân cắt furin của protein Env HIV, tốt hơn là thay thế ở các vị trí 508-511 bằng RRRRRR (SEQ ID NO: 10).
12. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 11, bao gồm trình tự axit amin mà là ít nhất 95% tương đồng với bất kỳ trong số các SEQ ID NO: 3, 5, 20, 22, 24, 26, 27, 28, 29, 30, 31, hoặc 32.
13. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 12, còn bao gồm: (xvi) gốc axit amin được chọn từ Val, Ile, Phe, Met, Ala, hoặc Leu, tốt hơn là Val hoặc Ile, tốt nhất là Val, ở vị trí 658.
14. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 13, là gp140 hoặc gp160 protein.
15. Protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 14, mà là từ nhánh C hoặc nhánh A, tốt hơn là từ nhánh C.
16. Phức trimer bao gồm oligome không cộng hóa trị của ba trong số các protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 15.
17. Hạt, tốt hơn là liposom hoặc hạt nano, thể hiện trên protein Env của HIV tái tổ hợp bề mặt của nó theo một điểm bất kỳ trong số các điểm từ 1 đến 15 hoặc phức trimer theo điểm 16.
18. Phân tử axit nucleic phân lập mã hóa protein Env của HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 15.
19. Vector chứa phân tử axit nucleic phân lập theo điểm 18 liên kết hoạt động với trình tự khởi đầu.
20. Vector theo điểm 19, mà là vector adenovirut.
21. Tế bào chủ chứa phân tử axit nucleic phân lập theo điểm 18 hoặc vector theo điểm 19 hoặc 20.
22. Phương pháp tạo ra protein Env của HIV tái tổ hợp, bao gồm nuôi tế bào chủ theo điểm 21 dưới các điều kiện thích hợp để tạo ra protein Env HIV tái tổ hợp.

23. Chế phẩm bao gồm protein Env HIV tái tổ hợp theo một điểm bất kỳ trong số các điểm từ 1 đến 15, phức trime theo điểm 16, hạt theo điểm 17, phân tử axit nucleic phân lập theo điểm 18, hoặc vector theo điểm 19 hoặc 20, và chất mang được dùng.

24. Phương pháp cải thiện sự tạo thành trime của protein Env HIV, phương pháp bao gồm thế một hoặc nhiều gốc axit amin trong protein Env HIV bố mẹ, trong đó một hoặc nhiều sự thế tạo ra một hoặc nhiều trong số các axit amin sau đây:

(i) Phe, Leu, Met, hoặc Trp, tốt hơn là Phe, ở vị trí 651;

(ii) Phe, Ile, Met, hoặc Trp, tốt hơn là Ile, ở vị trí 655;

(iii) Asn hoặc Gln, tốt hơn là Asn, ở vị trí 535;

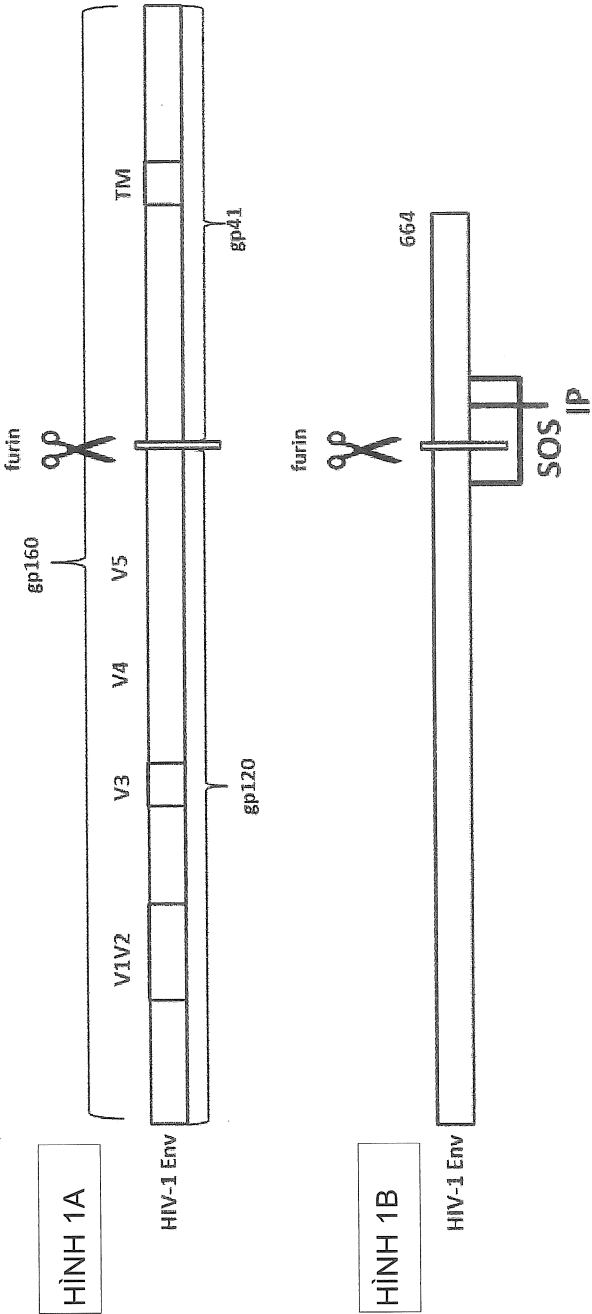
(iv) Val, Ile hoặc Ala, tốt hơn là Val, ở vị trí 589;

(v) Phe hoặc Trp, tốt hơn là Phe, ở vị trí 573;

(vi) Ile ở vị trí 204; và/hoặc

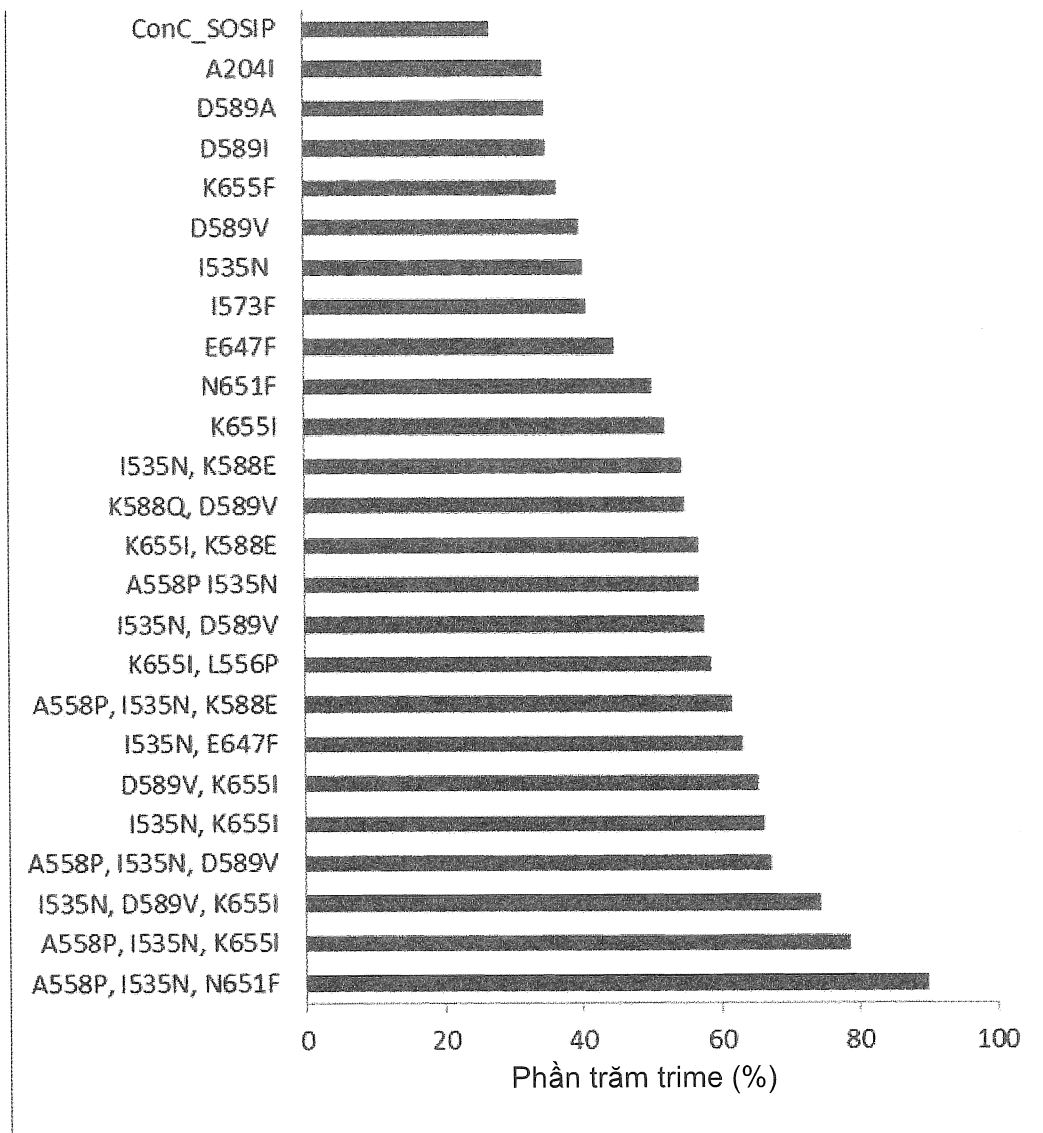
(vii) Phe, Met, hoặc Ile, tốt hơn là Phe, ở vị trí 647,

trong đó việc đánh số các vị trí là theo đánh số trong gp160 của HXB2 thể phân lập HIV-1.



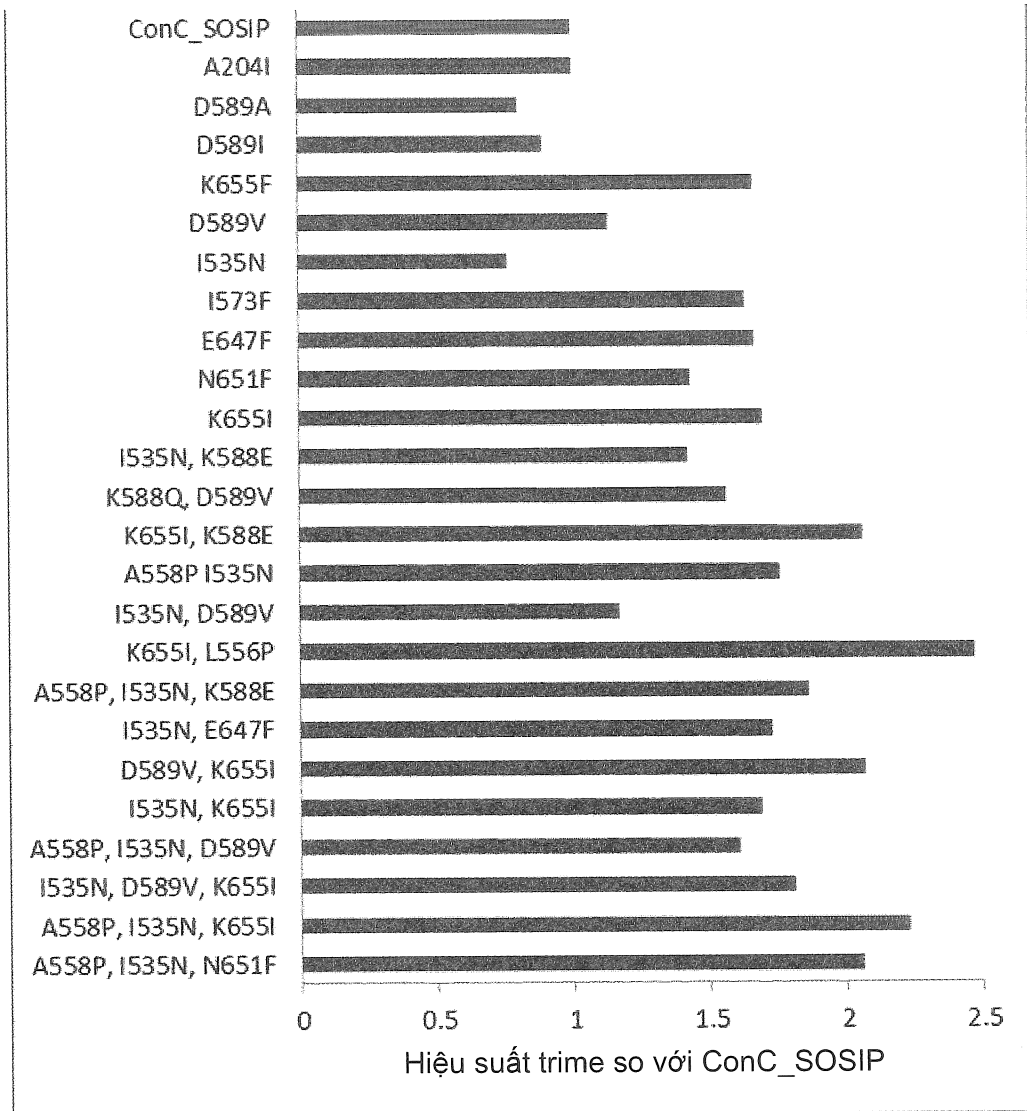
HÌNH 1

2/23



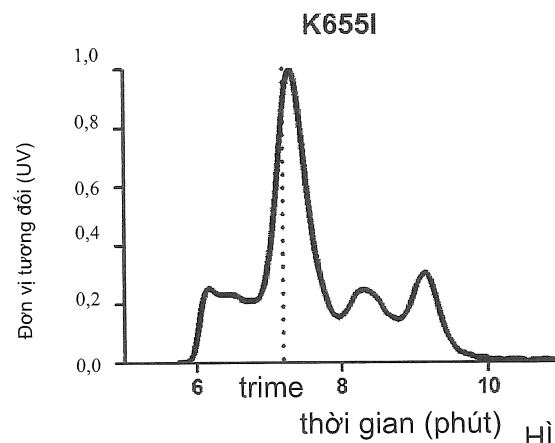
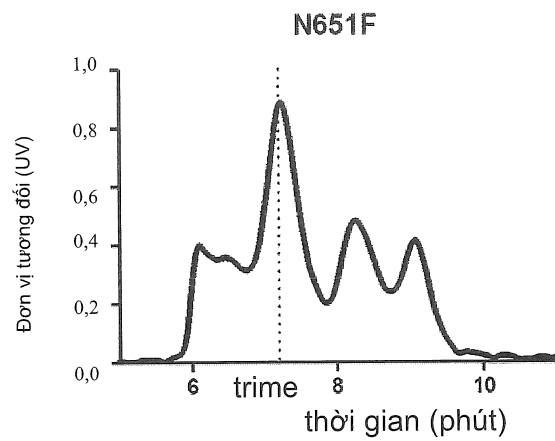
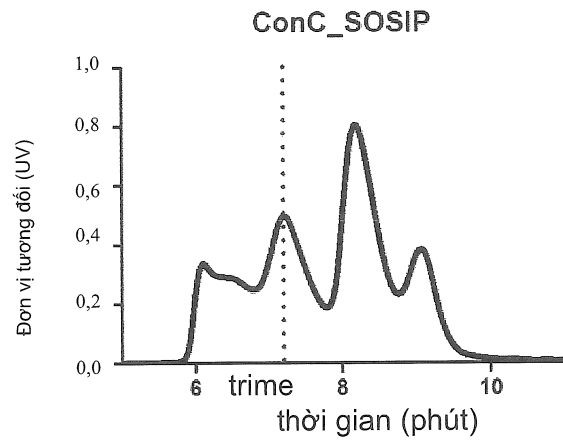
HÌNH 2A

3/23



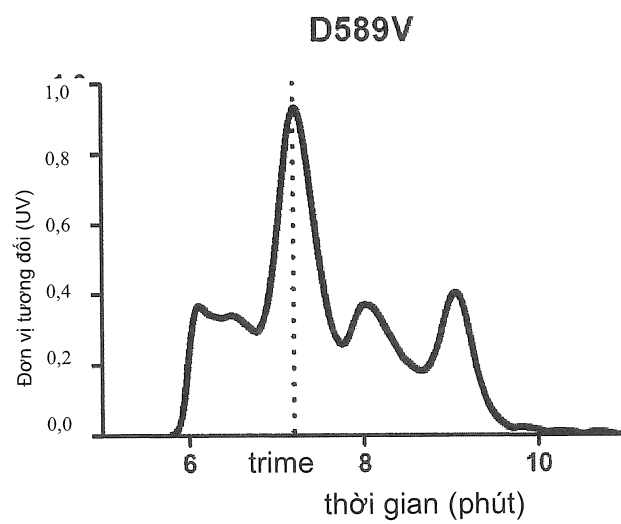
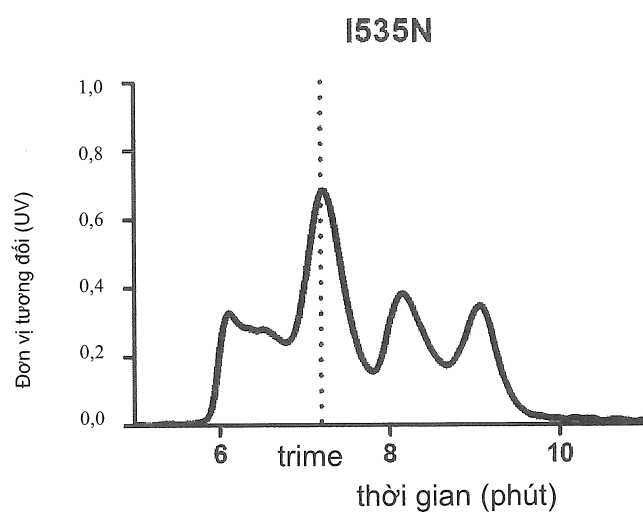
HÌNH 2B

4/23



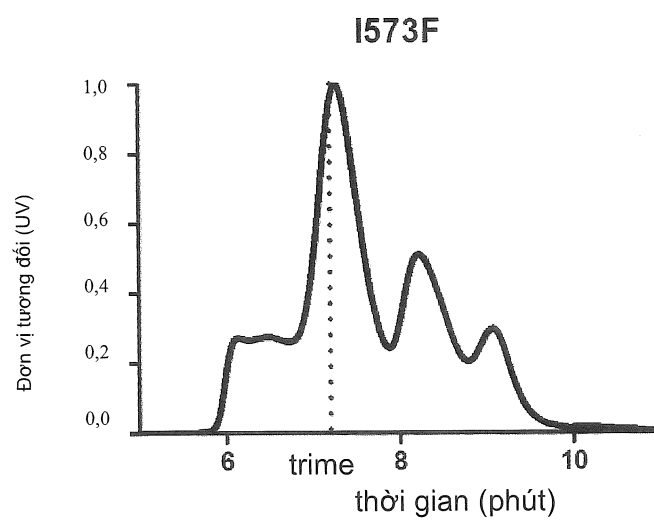
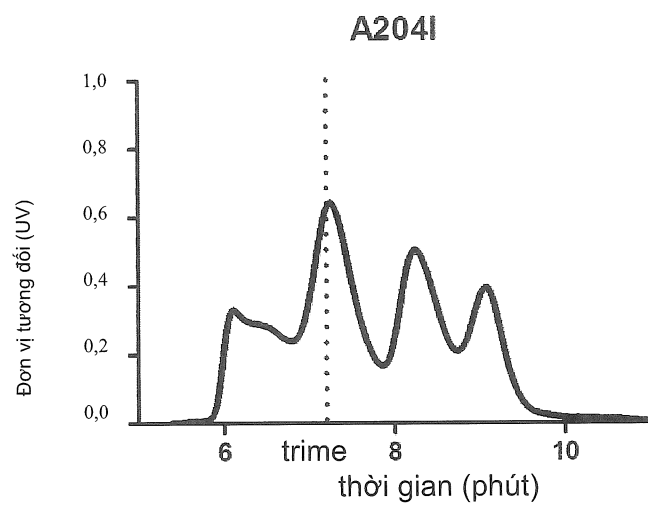
HÌNH 3

5/23

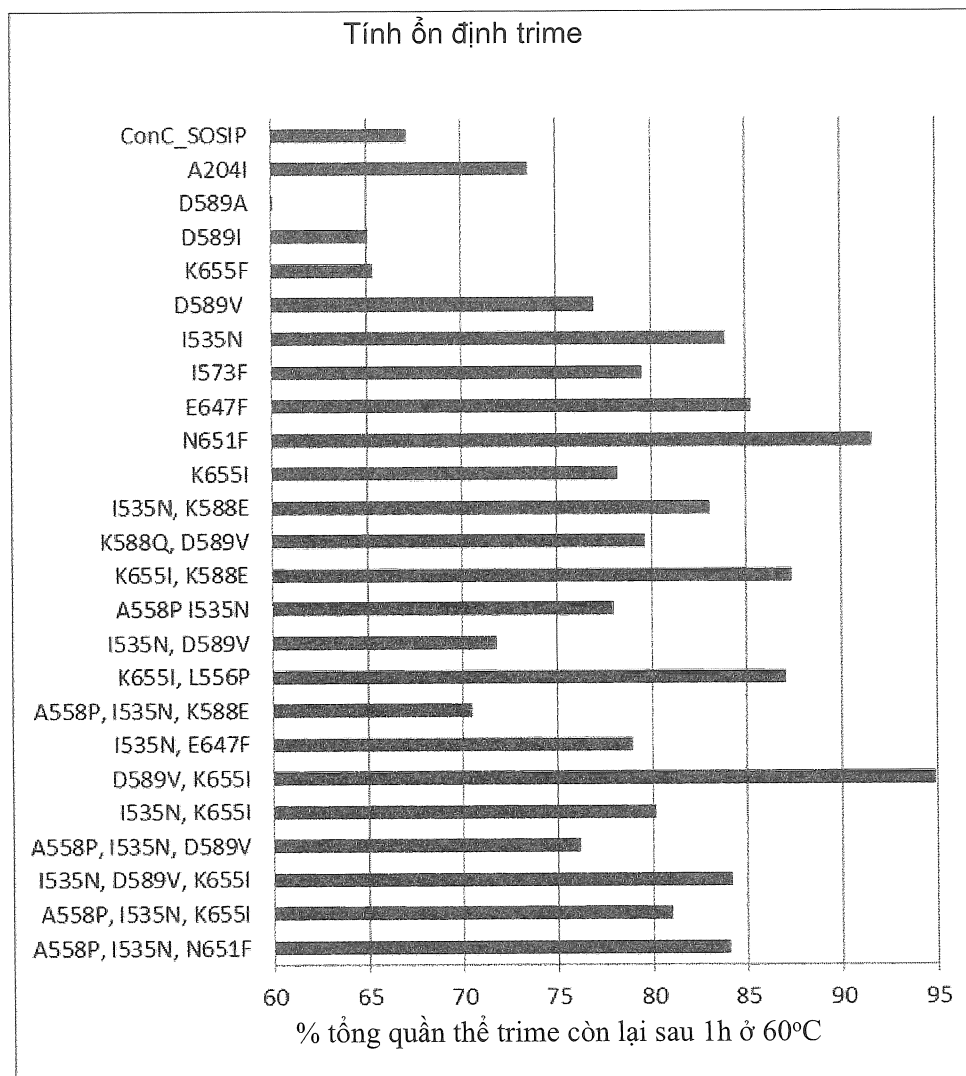


HÌNH 3 tiếp

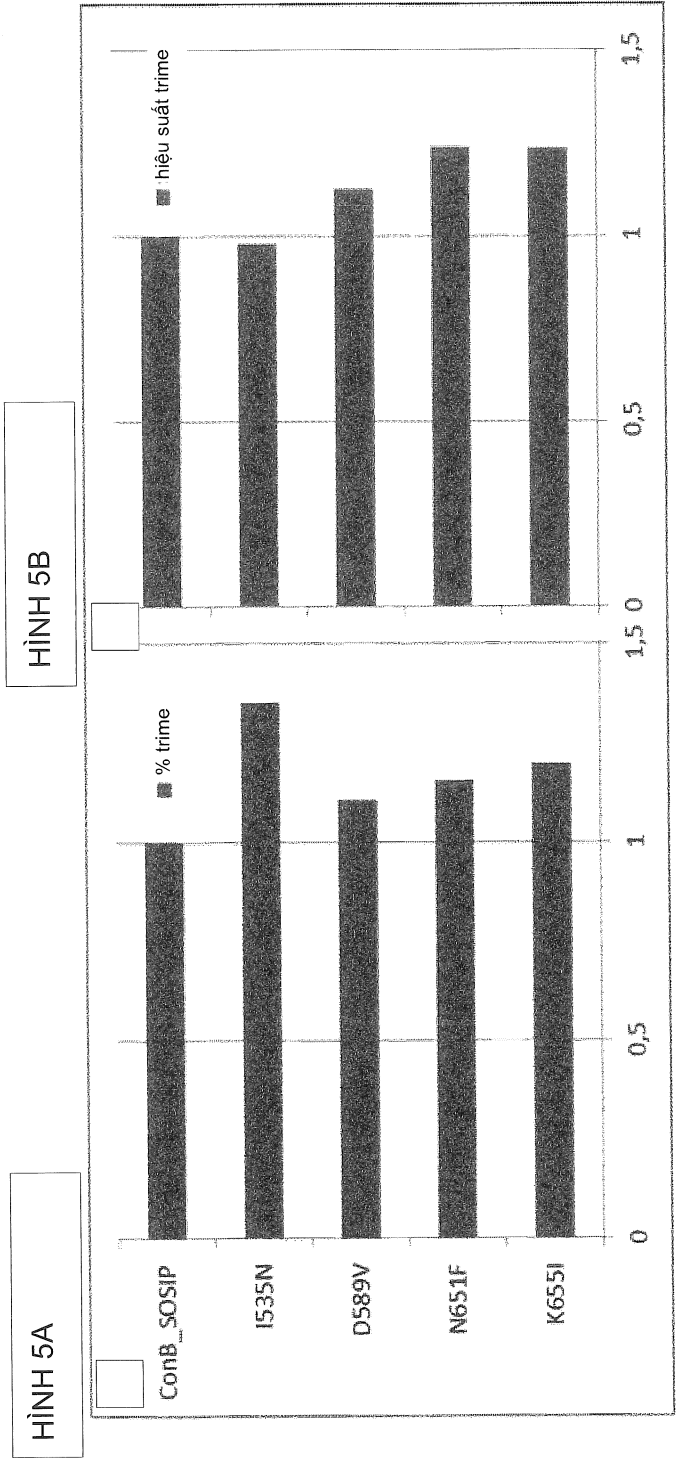
6/23



HÌNH 3 tiếp

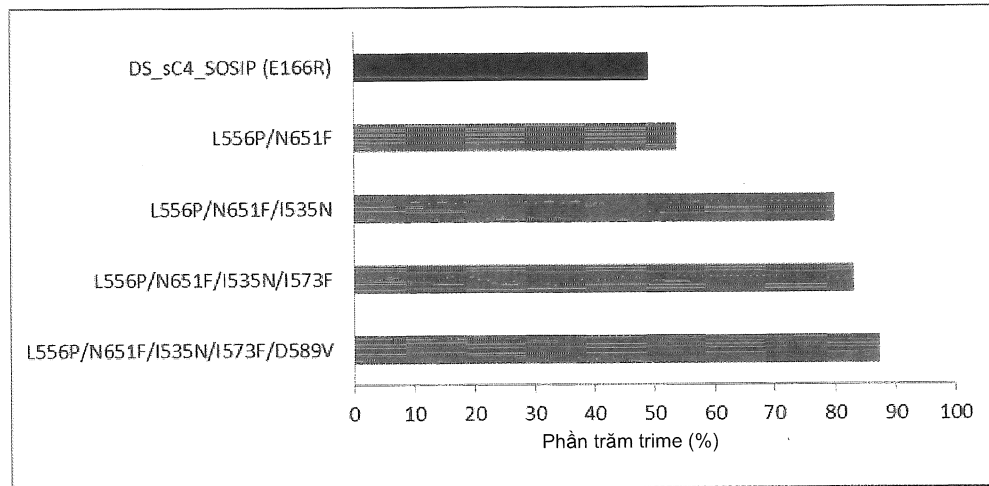


HÌNH 4

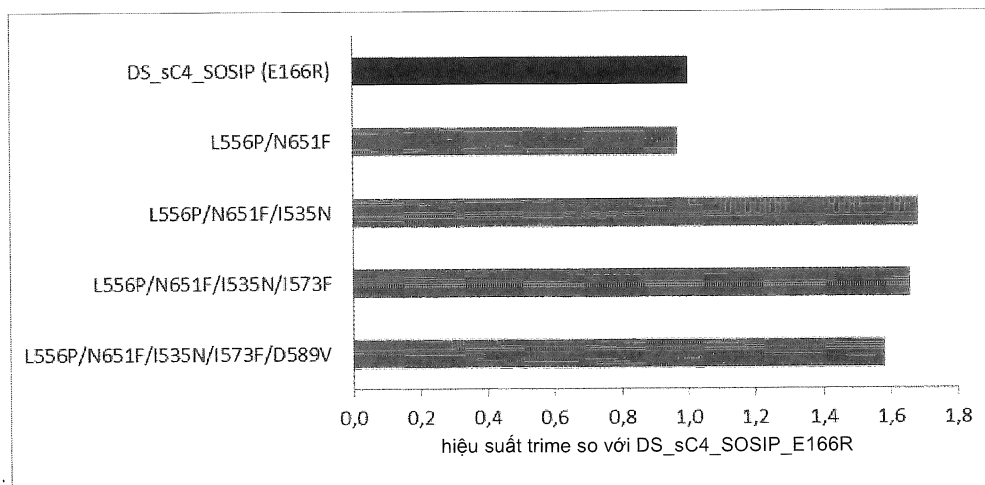


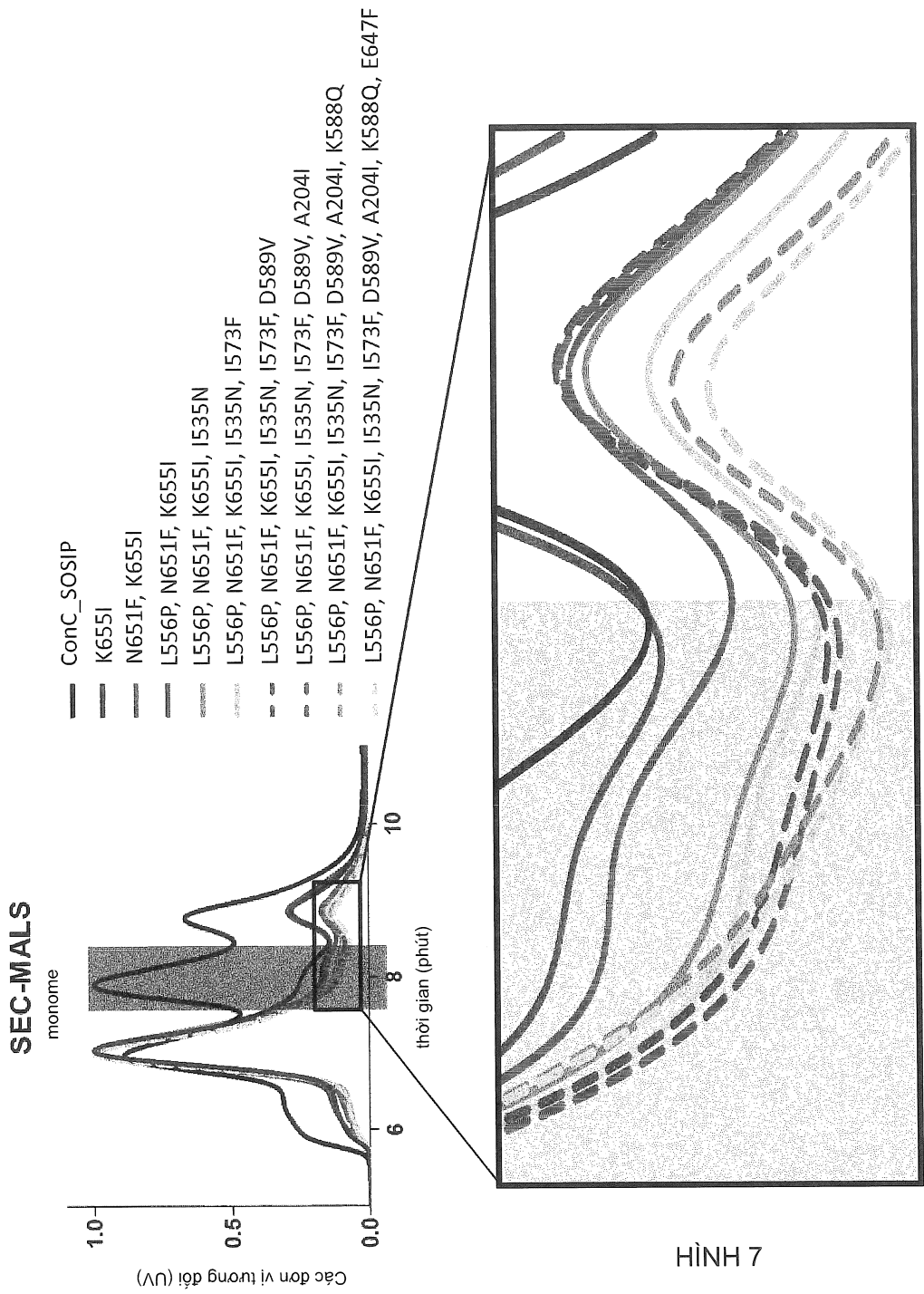
9/23

HÌNH 6A



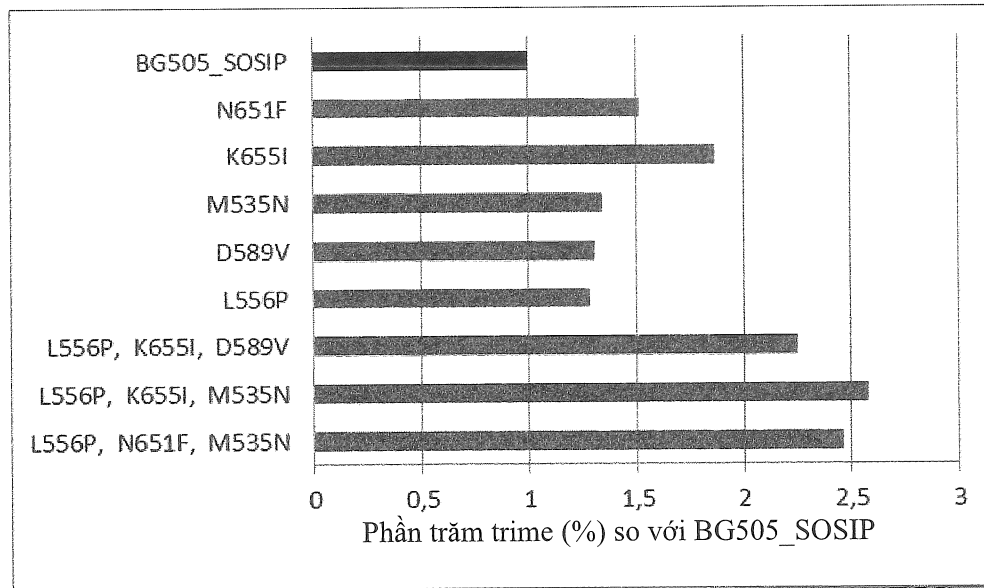
HÌNH 6B



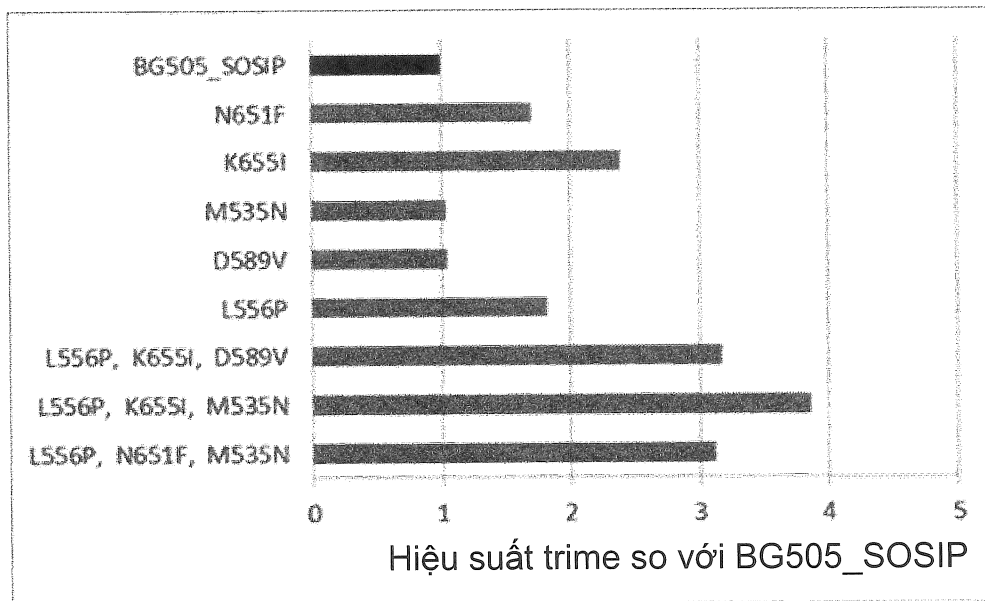


11/23

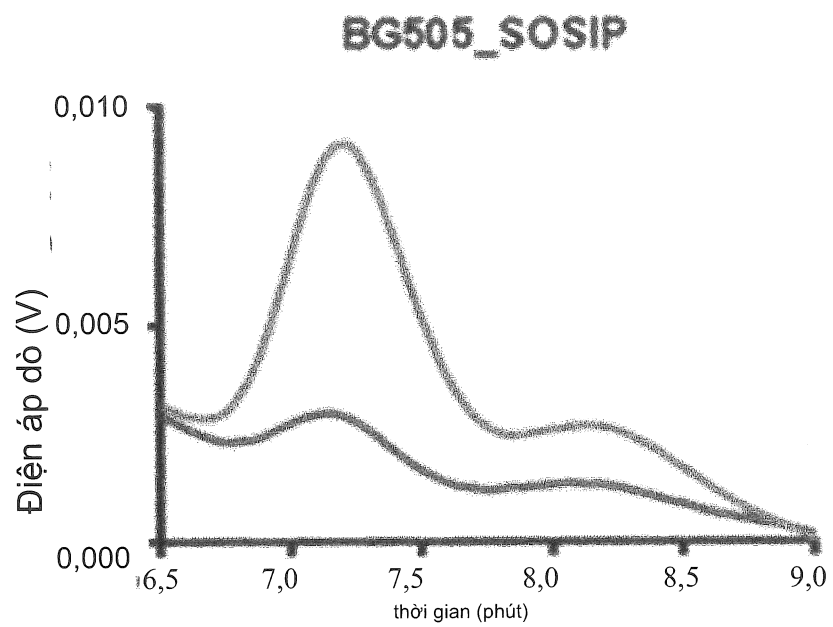
HÌNH 8A



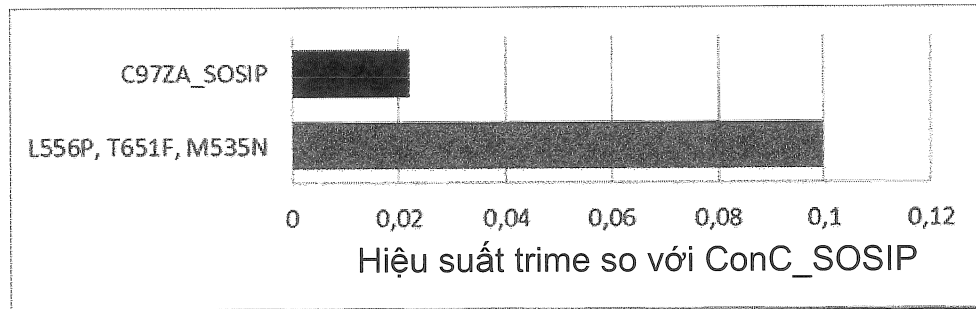
HÌNH 8B



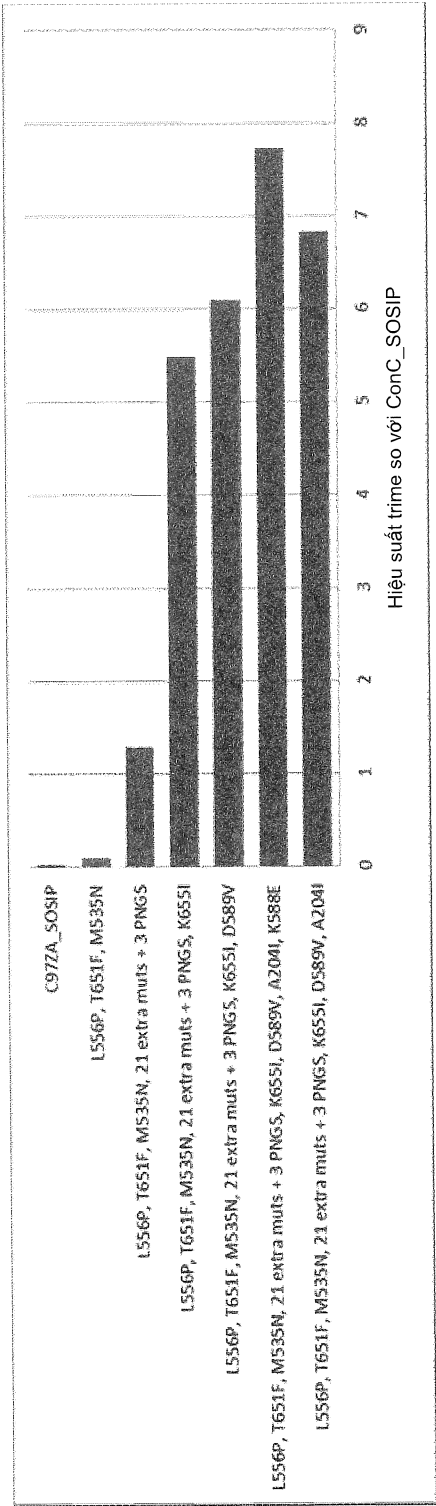
HÌNH 9



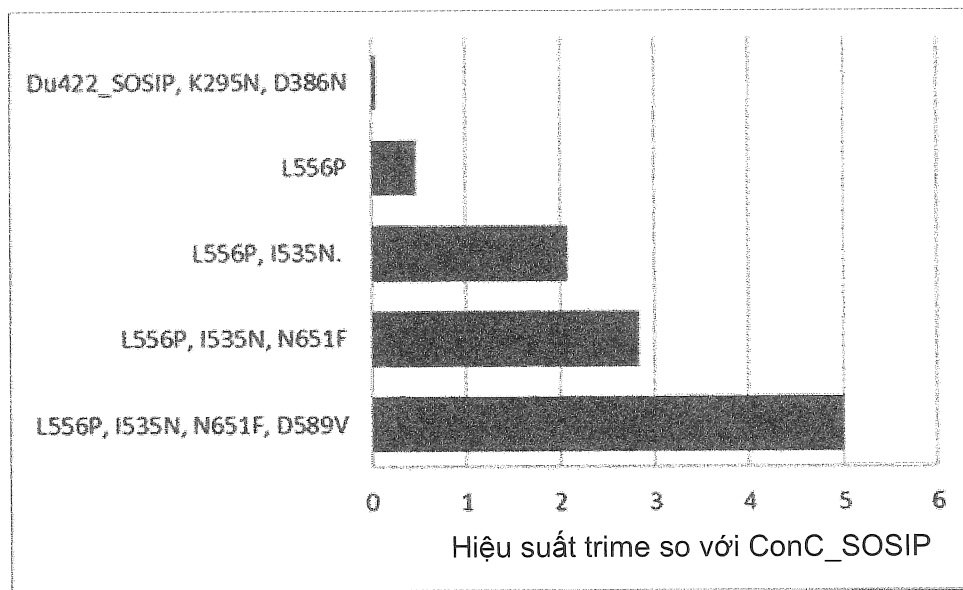
HÌNH 10A



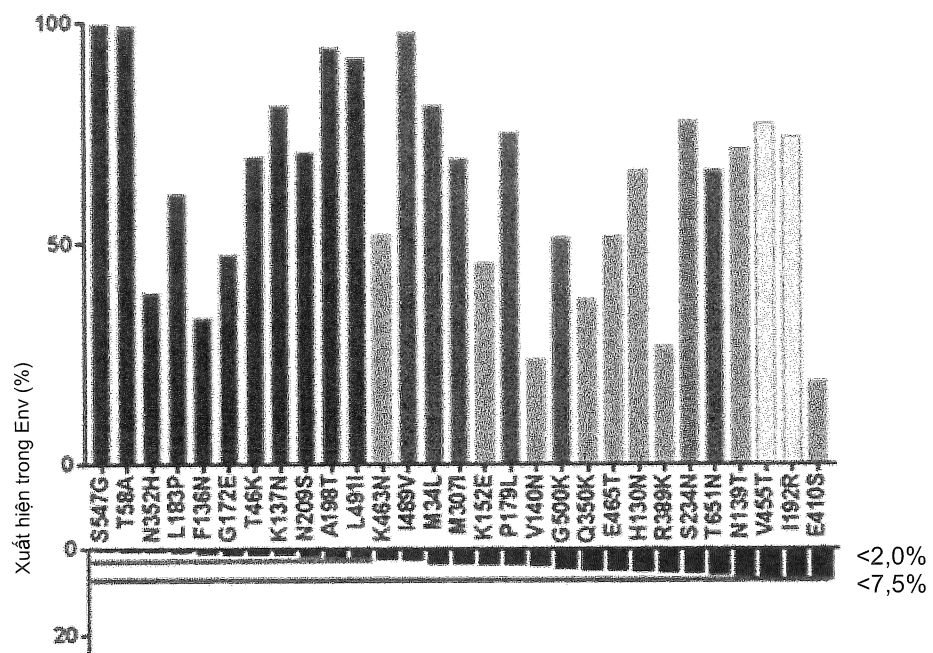
HÌNH 10B



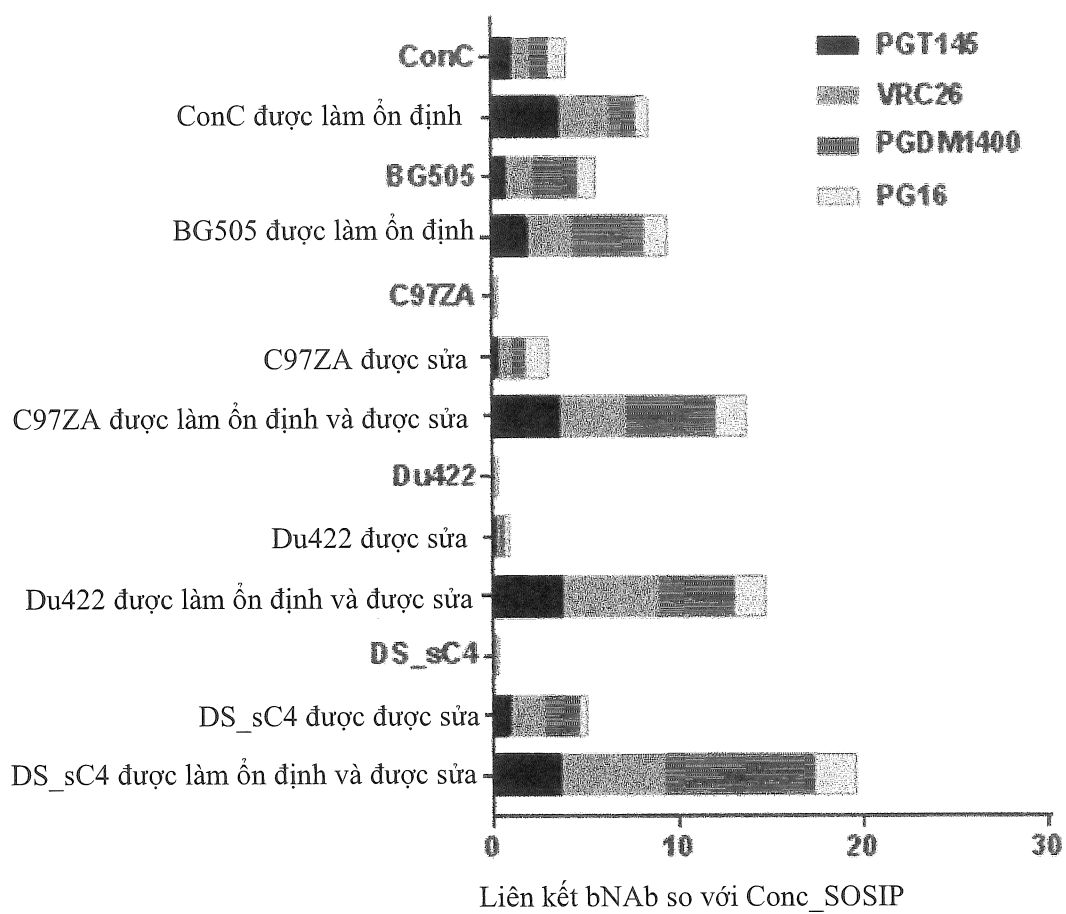
HÌNH 11



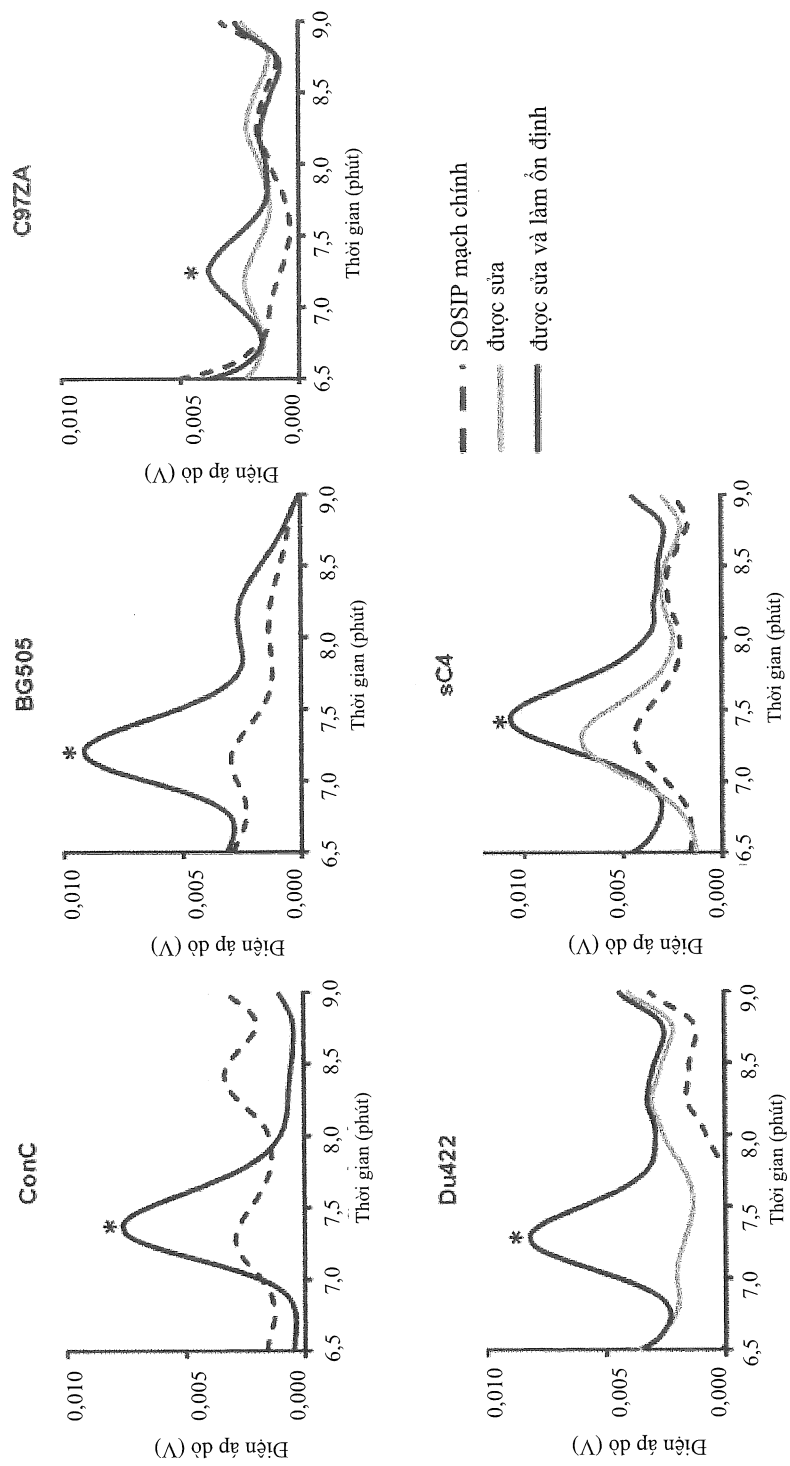
HÌNH 12



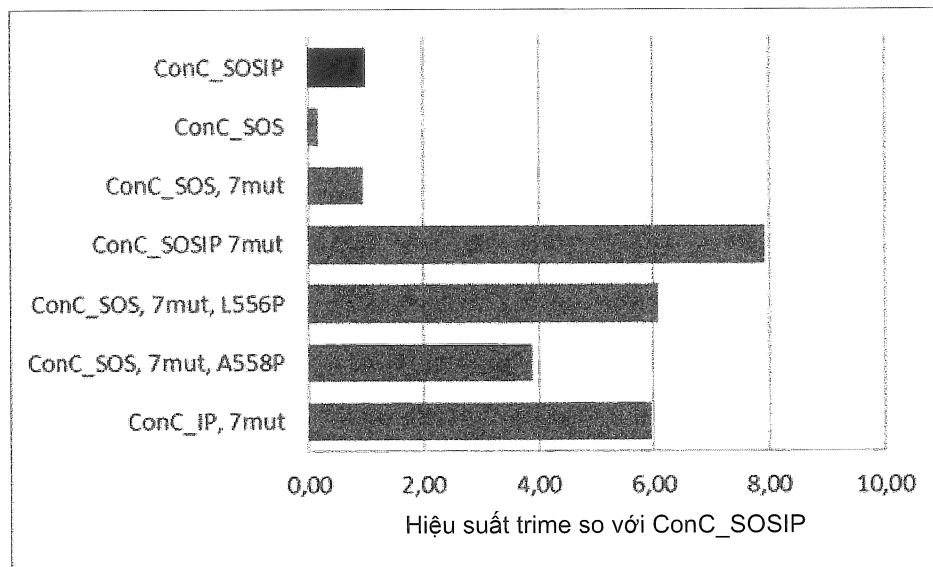
HÌNH 13



HÌNH 14

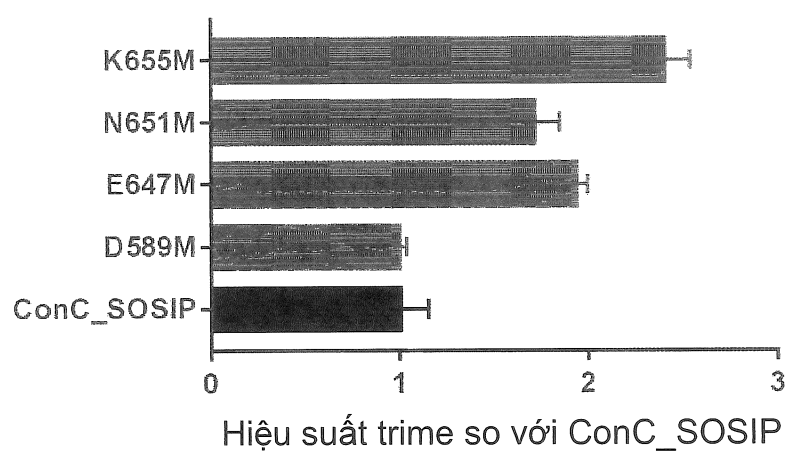


HÌNH 15

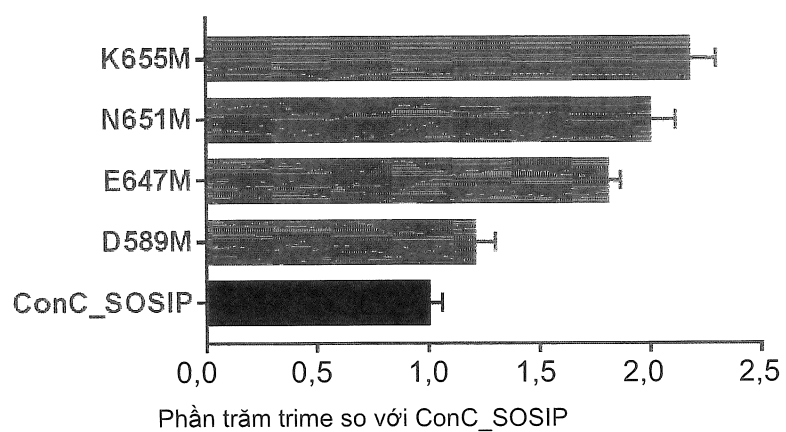


20/23

HÌNH 16A

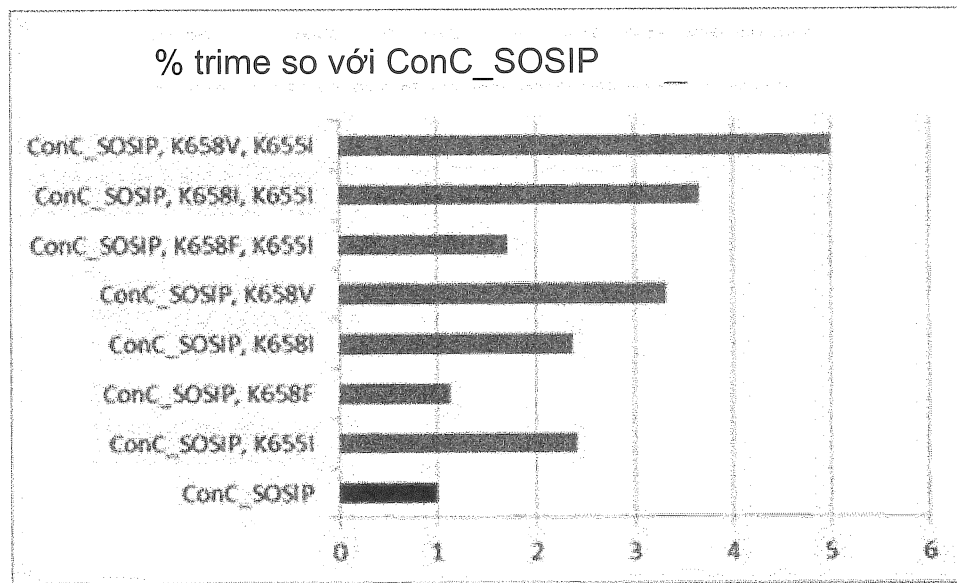


HÌNH 16B

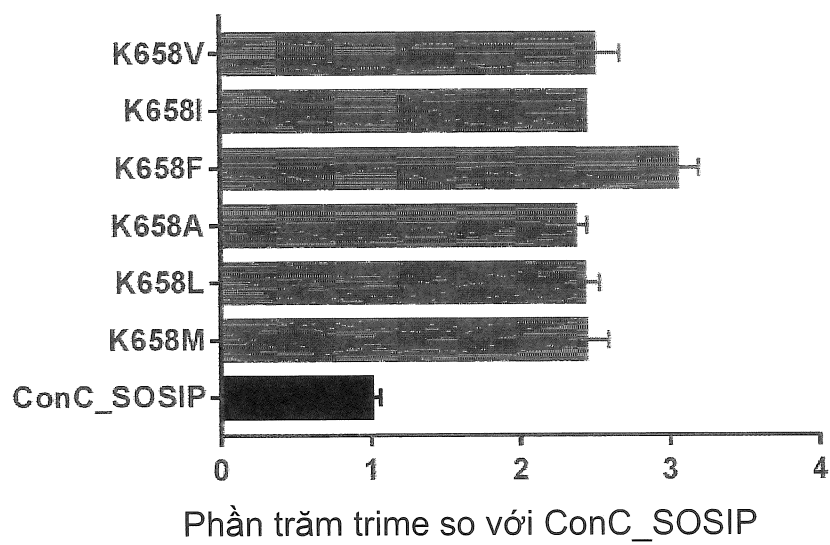


21/23

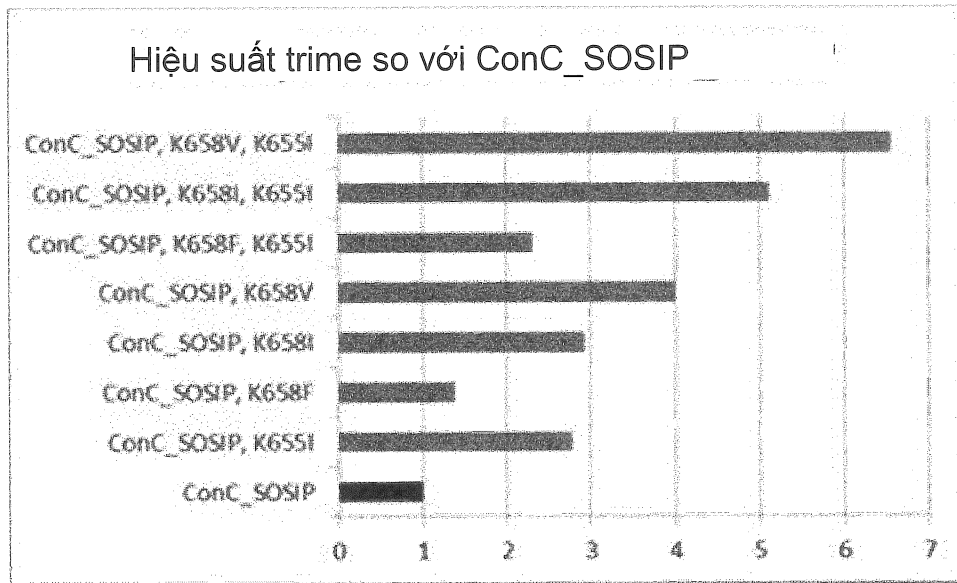
HÌNH 17A



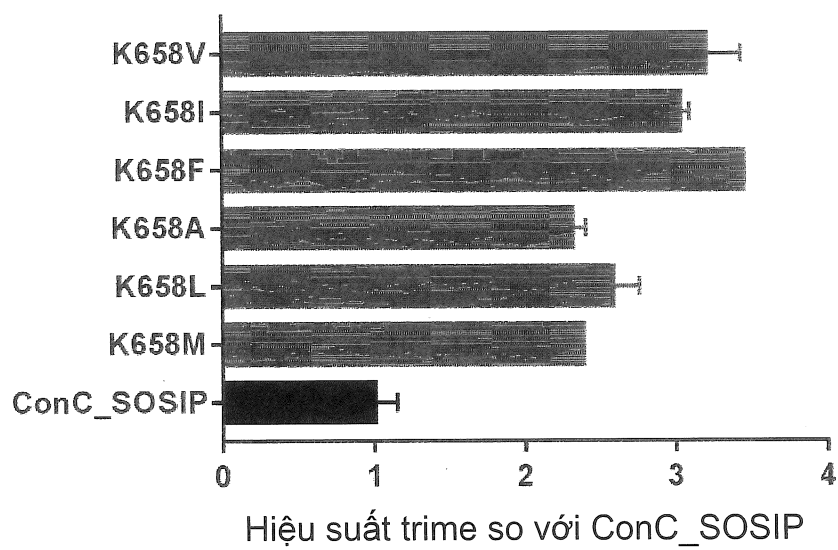
HÌNH 17B

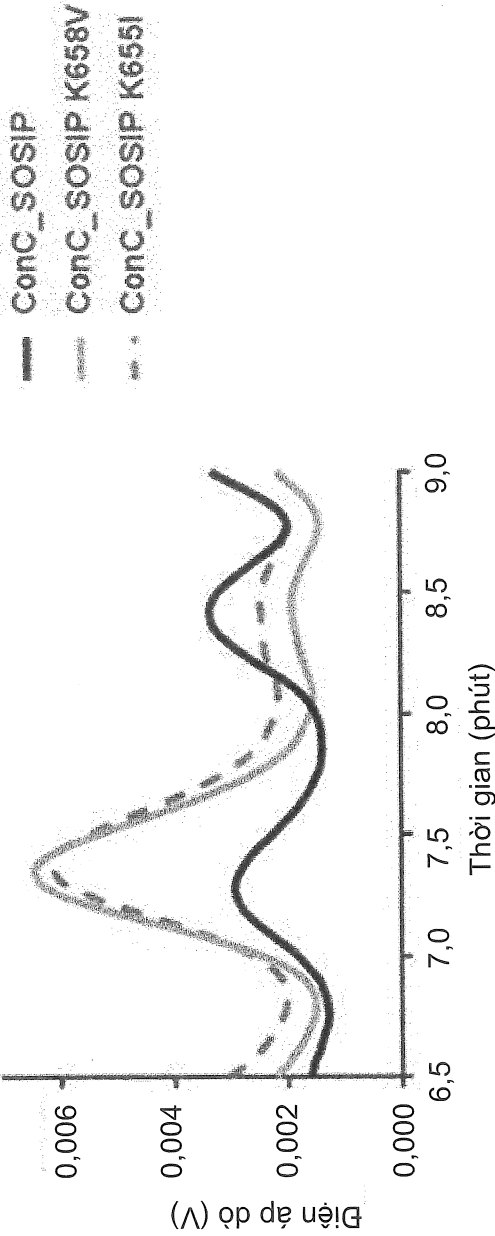


HÌNH 17C



HÌNH 17D





HÌNH 18

DANH MỤC TRÌNH TỰ

<110> Janssen Vaccines & Prevention B.V.

Rutten, Lucy

Truan, Daphné

Strokappe, Nika M.

Langedijk, Johannes P.M.

<120> Protein vỏ (env) của virus gây suy giảm miễn dịch tái tổ hợp ở người, phương pháp cải thiện sự tạo thành trimer của protein HIV

<130> 0276 WO 00 ORD

<150> EP16188866.4

<151> 2016-09-15

<160> 34

<170> PatentIn phiên bản 3.5

<210> 1

<211> 856

<212> PRT

<213> HIV-1

<220>

<221> đặc điểm_hỗn tạp

<223> gp160 của HXB2 thể phân lập HIV-1

<400> 1

Met Arg Val Lys Glu Lys Tyr Gln His Leu Trp Arg Trp Gly Trp Arg
 1 5 10 15

Trp Gly Thr Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala Thr Glu
 20 25 30

Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
 35 40 45

Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu
 50 55 60

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 65 70 75 80

Pro Gln Glu Val Val Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp
 85 90 95

Lys Asn Asp Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp
 100 105 110

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Ser
 115 120 125

Leu Lys Cys Thr Asp Leu Lys Asn Asp Thr Asn Thr Asn Ser Ser Ser
 130 135 140

Gly Arg Met Ile Met Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn
 145 150 155 160

Ile Ser Thr Ser Ile Arg Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe
 165 170 175

Tyr Lys Leu Asp Ile Ile Pro Ile Asp Asn Asp Thr Thr Ser Tyr Lys
 180 185 190

Leu Thr Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys Val
 195 200 205

Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala
 210 215 220

Ile Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr
 225 230 235 240

Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val Ser
 245 250 255

Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val Val Ile
 260 265 270

Arg Ser Val Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln Leu
 275 280 285

Asn Thr Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg
 290 295 300

Lys Arg Ile Arg Ile Gln Arg Gly Pro Gly Arg Ala Phe Val Thr Ile
 305 310 315 320

Gly Lys Ile Gly Asn Met Arg Gln Ala His Cys Asn Ile Ser Arg Ala
 325 330 335

Lys Trp Asn Asn Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln
 340 345 350

Phe Gly Asn Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp
 355 360 365

Pro Glu Ile Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr
 370 375 380

Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp
 385 390 395 400

Ser Thr Glu Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu
 405 410 415

Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Lys Val Gly Lys
 420 425 430

Ala Met Tyr Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn
 435 440 445

Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn Ser Asn Asn Glu
 450 455 460

Ser Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg
 465 470 475 480

Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val
 485 490 495

Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala
 500 505 510

Val Gly Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser
 515 520 525

Thr Met Gly Ala Ala Ser Met Thr Leu Thr Val Gln Ala Arg Gln Leu
 530 535 540

Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu
 545 550 555 560

Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu
 565 570 575

Gln Ala Arg Ile Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu
 580 585 590

Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val
 595 600 605

Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Glu Gln Ile Trp Asn
 610 615 620

His Thr Thr Trp Met Glu Trp Asp Arg Glu Ile Asn Asn Tyr Thr Ser
 625 630 635 640

Leu Ile His Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn
 645 650 655

Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp
 660 665 670

Phe Asn Ile Thr Asn Trp Leu Trp Tyr Ile Lys Leu Phe Ile Met Ile
 675 680 685

Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile
 690 695 700

Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His
 705 710 715 720

Leu Pro Thr Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu Glu
 725 730 735

Gly Gly Glu Arg Asp Arg Asp Arg Ser Ile Arg Leu Val Asn Gly Ser
 740 745 750

Leu Ala Leu Ile Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr
 755 760 765

His Arg Leu Arg Asp Leu Leu Leu Ile Val Thr Arg Ile Val Glu Leu
 770 775 780

Leu Gly Arg Arg Gly Trp Glu Ala Leu Lys Tyr Trp Trp Asn Leu Leu
 785 790 795 800

Gln Tyr Trp Ser Gln Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn
 805 810 815

Ala Thr Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val
 820 825 830

Val Gln Gly Ala Cys Arg Ala Ile Arg His Ile Pro Arg Arg Ile Arg
 835 840 845

Gln Gly Leu Glu Arg Ile Leu Leu
 850 855

<210> 2

<211> 615

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Trình tự nhánh C đồng nhất Env HIV

<400> 2

Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
 1 5 10 15

Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Lys Glu
 20 25 30

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 35 40 45

Pro Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp
 50 55 60

Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp
 65 70 75 80

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 85 90 95

Leu Asn Cys Thr Asn Val Asn Val Thr Asn Thr Asn Asn Asn Asn Met
 100 105 110

Lys Glu Glu Met Lys Asn Cys Ser Phe Asn Thr Thr Thr Glu Ile Arg
 115 120 125

Asp Lys Lys Gln Lys Glu Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val
 130 135 140

Pro Leu Asn Glu Asn Ser Ser Glu Tyr Arg Leu Ile Asn Cys Asn Thr
 145 150 155 160

Ser Thr Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro
 165 170 175

Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn
 180 185 190

Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln
 195 200 205

Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn
 210 215 220

Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr
 225 230 235 240

Asp Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Ser Val Glu Ile
 245 250 255

Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly
 260 265 270

Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg
 275 280 285

Gln Ala His Cys Asn Ile Ser Glu Ala Lys Trp Asn Lys Thr Leu Gln
 290 295 300

Arg Val Lys Lys Lys Leu Lys Glu His Phe Pro Asn Lys Thr Ile Lys
 305 310 315 320

Phe Ala Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser Phe
 325 330 335

Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Lys Leu Phe Asn
 340 345 350

Ser Thr Tyr Asn Asn Thr Thr Ser Asn Ser Thr Ile Thr Leu Pro Cys
 355 360 365

Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Met
 370 375 380

Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Lys Ser Asn Ile Thr
 385 390 395 400

Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Asn Thr Glu
 405 410 415

Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu
420 425 430

Leu Tyr Lys Tyr Lys Val Val Glu Ile Lys Pro Leu Gly Ile Ala Pro
435 440 445

Thr Lys Ala Lys Arg Arg Val Val Glu Arg Glu Lys Arg Arg Ala Val
450 455 460

Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr
465 470 475 480

Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg Gln Leu Leu
485 490 495

Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala Ile Glu Ala
500 505 510

Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln
515 520 525

Ala Arg Val Leu Ala Ile Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu
 530 535 540

Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro
 545 550 555 560

Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Glu Asp Ile Trp Asp Asn
 565 570 575

Met Thr Trp Met Gln Trp Asp Arg Glu Ile Ser Asn Tyr Thr Asp Thr
 580 585 590

Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu
 595 600 605

Lys Asp Leu Leu Ala Leu Asp
 610 615

<210> 3

<211> 616

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Trình tự ConC_SOSIP

<400> 3

Asn	Leu	Trp	Val	Thr	Val	Tyr	Tyr	Gly	Val	Pro	Val	Trp	Lys	Glu	Ala
1				5					10					15	

Lys	Thr	Thr	Leu	Phe	Cys	Ala	Ser	Asp	Ala	Lys	Ala	Tyr	Glu	Lys	Glu
			20					25					30		

Val	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp	Pro	Asn
		35					40						45		

Pro	Gln	Glu	Met	Val	Leu	Glu	Asn	Val	Thr	Glu	Asn	Phe	Asn	Met	Trp
		50				55					60				

Lys	Asn	Asp	Met	Val	Asp	Gln	Met	His	Glu	Asp	Ile	Ile	Ser	Leu	Trp
65						70				75				80	

Asp	Gln	Ser	Leu	Lys	Pro	Cys	Val	Lys	Leu	Thr	Pro	Leu	Cys	Val	Thr
				85					90					95	

Leu	Asn	Cys	Thr	Asn	Val	Asn	Val	Thr	Asn	Thr	Asn	Asn	Asn	Asn	Met
					100				105				110		

Lys Glu Glu Met Lys Asn Cys Ser Phe Asn Thr Thr Thr Glu Ile Arg
 115 120 125

Asp Lys Lys Gln Lys Glu Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val
 130 135 140

Pro Leu Asn Glu Asn Ser Ser Glu Tyr Arg Leu Ile Asn Cys Asn Thr
 145 150 155 160

Ser Thr Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro Ile Pro
 165 170 175

Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn
 180 185 190

Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln
 195 200 205

Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn
 210 215 220

Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr
 225 230 235 240

Asp Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Ser Val Glu Ile
 245 250 255

Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly
 260 265 270

Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg
 275 280 285

Gln Ala His Cys Asn Ile Ser Glu Ala Lys Trp Asn Lys Thr Leu Gln
 290 295 300

Arg Val Lys Lys Lys Leu Lys Glu His Phe Pro Asn Lys Thr Ile Lys
 305 310 315 320

Phe Ala Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser Phe
 325 330 335

Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Lys Leu Phe Asn
 340 345 350

Ser Thr Tyr Asn Asn Thr Thr Ser Asn Ser Thr Ile Thr Leu Pro Cys
 355 360 365

Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Met
 370 375 380

Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Lys Ser Asn Ile Thr
 385 390 395 400

Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Asn Thr Glu
 405 410 415

Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu
 420 425 430

Leu Tyr Lys Tyr Lys Val Val Glu Ile Lys Pro Leu Gly Ile Ala Pro
 435 440 445

Thr Lys Cys Lys Arg Arg Val Val Glu Arg Arg Arg Arg Arg Arg Ala
 450 455 460

Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser
 465 470 475 480

Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg Gln Leu
 485 490 495

Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala Pro Glu
 500 505 510

Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu
 515 520 525

Gln Ala Arg Val Leu Ala Ile Glu Arg Tyr Leu Lys Asp Gln Gln Leu
 530 535 540

Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr Ala Val
 545 550 555 560

Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Glu Asp Ile Trp Asp
 565 570 575

Asn Met Thr Trp Met Gln Trp Asp Arg Glu Ile Ser Asn Tyr Thr Asp
 580 585 590

Thr Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn
 595 600 605

Glu Lys Asp Leu Leu Ala Leu Asp
 610 615

<210> 4

<211> 630

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Trình tự nhánh B đồng nhất Env HIV

<400> 4

Ala Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys
 1 5 10 15

Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp
 20 25 30

Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp
 35 40 45

Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn
 50 55 60

Met Trp Lys Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser
 65 70 75 80

Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys
 85 90 95

Val Thr Leu Asn Cys Thr Asp Leu Asn Asn Asn Thr Thr Asn Asn Asn
 100 105 110

Ser Ser Ser Glu Lys Met Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe
 115 120 125

Asn Ile Thr Thr Ser Ile Arg Asp Lys Val Gln Lys Glu Tyr Ala Leu
 130 135 140

Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr Ser Tyr
 145 150 155 160

Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys
 165 170 175

Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe
 180 185 190

Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Thr Gly Pro Cys
 195 200 205

Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val
 210 215 220

Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val Val
 225 230 235 240

Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln
 245 250 255

Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr
 260 265 270

Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr Ala Thr Gly
 275 280 285

Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser Arg Thr
 290 295 300

Lys Trp Asn Asn Thr Leu Lys Gln Ile Val Lys Lys Leu Arg Glu Gln
 305 310 315 320

Phe Gly Asn Lys Thr Ile Val Phe Asn Gln Ser Ser Gly Gly Asp Pro
 325 330 335

Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys
 340 345 350

Asn Thr Thr Gln Leu Phe Asn Ser Thr Trp Asn Ser Asn Gly Thr Trp
 355 360 365

Asn Asn Thr Thr Gly Asn Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys
 370 375 380

Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro
 385 390 395 400

Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu
 405 410 415

Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Asn Thr Thr Glu Thr Phe
 420 425 430

Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr
 435 440 445

Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys
 450 455 460

Cys Lys Arg Arg Val Val Gln Arg Arg Arg Arg Arg Arg Ala Val Gly
 465 470 475 480

Ile Gly Ala Met Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met
 485 490 495

Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser
 500 505 510

Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Pro Glu Ala Gln
 515 520 525

Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala
 530 535 540

Arg Val Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly
 545 550 555 560

Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr Ala Val Pro Trp
 565 570 575

Asn Thr Ser Trp Ser Asn Lys Ser Leu Asp Glu Ile Trp Asp Asn Met
 580 585 590

Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr Thr Gly Leu Ile
 595 600 605

Tyr Thr Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln
 610 615 620

Glu Leu Leu Glu Leu Asp
 625 630

<210> 5

<211> 630

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Trình tự ConB_SOSIP

<400> 5

Ala	Glu	Lys	Leu	Trp	Val	Thr	Val	Tyr	Tyr	Gly	Val	Pro	Val	Trp	Lys
1				5					10					15	

Glu	Ala	Thr	Thr	Thr	Leu	Phe	Cys	Ala	Ser	Asp	Ala	Lys	Ala	Tyr	Asp
			20					25					30		

Thr	Glu	Val	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp
		35					40					45			

Pro	Asn	Pro	Gln	Glu	Val	Val	Leu	Glu	Asn	Val	Thr	Glu	Asn	Phe	Asn
	50						55					60			

Met	Trp	Lys	Asn	Asn	Met	Val	Glu	Gln	Met	His	Glu	Asp	Ile	Ile	Ser
65						70				75				80	

Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys
 85 90 95

Val Thr Leu Asn Cys Thr Asp Leu Asn Asn Asn Thr Thr Asn Asn Asn
 100 105 110

Ser Ser Ser Glu Lys Met Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe
 115 120 125

Asn Ile Thr Thr Ser Ile Arg Asp Lys Val Gln Lys Glu Tyr Ala Leu
 130 135 140

Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr Ser Tyr
 145 150 155 160

Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys
 165 170 175

Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe
 180 185 190

Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Thr Gly Pro Cys
 195 200 205

Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val
 210 215 220

Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val Val
 225 230 235 240

Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln
 245 250 255

Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr
 260 265 270

Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr Ala Thr Gly
 275 280 285

Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser Arg Thr
 290 295 300

Lys Trp Asn Asn Thr Leu Lys Gln Ile Val Lys Lys Leu Arg Glu Gln
 305 310 315 320

Phe Gly Asn Lys Thr Ile Val Phe Asn Gln Ser Ser Gly Gly Asp Pro
 325 330 335

Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys
 340 345 350

Asn Thr Thr Gln Leu Phe Asn Ser Thr Trp Asn Ser Asn Gly Thr Trp
 355 360 365

Asn Asn Thr Thr Gly Asn Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys
 370 375 380

Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro
 385 390 395 400

Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu
 405 410 415

Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Asn Thr Thr Glu Thr Phe
 420 425 430

Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr
 435 440 445

Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys
 450 455 460

Cys Lys Arg Arg Val Val Gln Arg Arg Arg Arg Arg Arg Ala Val Gly
 465 470 475 480

Ile Gly Ala Met Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met
 485 490 495

Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser
 500 505 510

Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Pro Glu Ala Gln
 515 520 525

Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala
 530 535 540

Arg Val Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly
 545 550 555 560

Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr Ala Val Pro Trp
 565 570 575

Asn Thr Ser Trp Ser Asn Lys Ser Leu Asp Glu Ile Trp Asp Asn Met
 580 585 590

Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr Thr Gly Leu Ile
 595 600 605

Tyr Thr Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln
 610 615 620

Glu Leu Leu Glu Leu Asp
 625 630

<210> 6

<211> 625

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Đoạn C4 Env Mos2S protein vỏ HIV tổng hợp

<400> 6

Met Gly Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys
 1 5 10 15

Asp Ala Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu
 20 25 30

Lys Glu Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp
 35 40 45

Pro Asn Pro Gln Glu Ile Val Leu Gly Asn Val Thr Glu Asn Phe Asn
 50 55 60

Met Trp Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser
 65 70 75 80

Leu Trp Asp Ala Ser Leu Glu Pro Cys Val Lys Leu Thr Pro Leu Cys
 85 90 95

Val Thr Leu Asn Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr
 100 105 110

Tyr Asn Ile Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser
 115 120 125

Phe Asn Ala Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala
 130 135 140

Leu Phe Tyr Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser
 145 150 155 160

Glu Lys Ser Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys
 165 170 175

Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro
 180 185 190

Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys
 195 200 205

Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr
 210 215 220

Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu
 225 230 235 240

Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn
 245 250 255

Leu Thr Asn Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val
 260 265 270

Asn Ile Thr Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg
 275 280 285

Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp
 290 295 300

Ile Arg Gln Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr
 305 310 315 320

Leu Gln Gly Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr
 325 330 335

Ile Lys Phe Ala Pro His Ser Gly Gly Asp Leu Glu Ile Thr Thr His
 340 345 350

Thr Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Asn Leu
 355 360 365

Phe Asn Glu Ser Asn Ile Glu Arg Asn Asp Ser Ile Ile Thr Leu Pro
 370 375 380

Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala
 385 390 395 400

Ile Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser Asn Ile
 405 410 415

Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly Val Pro
 420 425 430

Asn Asp Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asn Asn
 435 440 445

Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys Pro Leu
 450 455 460

Gly Val Ala Pro Thr Glu Ala Lys Arg Arg Val Val Glu Arg Glu Lys
 465 470 475 480

Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly Ala Ala
 485 490 495

Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg
 500 505 510

Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala
 515 520 525

Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys
 530 535 540

Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln Asp Gln
 545 550 555 560

Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
 565 570 575

Ala Val Pro Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr Asp Ile
 580 585 590

Trp Asp Asn Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly Asn Tyr
 595 600 605

Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu
 610 615 620

Lys
 625

<210> 7

<211> 627

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> trình tự DS_sC4_SOSIP_E166R

<400> 7

Met Gly Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys
 1 5 10 15

Asp Ala Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu
 20 25 30

Lys Glu Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp
 35 40 45

Pro Asn Pro Gln Glu Ile Val Leu Gly Asn Val Thr Glu Asn Phe Asn
 50 55 60

Met Trp Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser
 65 70 75 80

Leu Trp Asp Ala Ser Leu Glu Pro Cys Val Lys Leu Thr Pro Leu Cys
 85 90 95

Val Thr Leu Asn Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr
 100 105 110

Tyr Asn Ile Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser
 115 120 125

Phe Asn Ala Thr Thr Val Val Arg Asp Arg Lys Gln Lys Val His Ala
 130 135 140

Leu Phe Tyr Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser
 145 150 155 160

Glu Lys Ser Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys
 165 170 175

Asn Thr Ser Ala Cys Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro
 180 185 190

Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys
 195 200 205

Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr
 210 215 220

Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu
 225 230 235 240

Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn
 245 250 255

Leu Thr Asn Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val
 260 265 270

Asn Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg
 275 280 285

Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp
 290 295 300

Ile Arg Gln Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr
 305 310 315 320

Leu Gln Gly Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr
 325 330 335

Ile Lys Phe Ala Pro His Ser Gly Gly Asp Leu Glu Ile Thr Thr His
 340 345 350

Thr Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Asn Leu
 355 360 365

Phe Asn Glu Ser Asn Ile Glu Arg Asn Asp Ser Ile Ile Thr Leu Pro
 370 375 380

Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Cys
 385 390 395 400

Ile Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Arg Ser Asn Ile
 405 410 415

Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser Asn Asn Gly Val Pro
 420 425 430

Asn Asp Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asn Asn
 435 440 445

Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Val Lys Pro Leu
 450 455 460

Gly Val Ala Pro Thr Glu Cys Lys Arg Arg Val Val Glu Arg Arg Arg
 465 470 475 480

Arg Arg Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Ile Leu Gly
 485 490 495

Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln
 500 505 510

Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu
 515 520 525

Arg Ala Pro Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly
 530 535 540

Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Gln
 545 550 555 560

Asp Gln Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly Lys Leu Ile Cys
 565 570 575

Cys Thr Ala Val Pro Trp Asn Thr Ser Trp Ser Asn Lys Ser Gln Thr
 580 585 590

Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp Lys Glu Ile Gly
 595 600 605

Asn Tyr Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln
 610 615 620

Gln Glu Lys
 625

<210> 8

<211> 631

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> trình tự Mos1.Env

<400> 8

Ala	Gly	Lys	Leu	Trp	Val	Thr	Val	Tyr	Tyr	Gly	Val	Pro	Val	Trp	Lys
1				5					10					15	

Glu	Ala	Thr	Thr	Thr	Leu	Phe	Cys	Ala	Ser	Asp	Ala	Lys	Ala	Tyr	Asp
				20				25					30		

Thr	Glu	Val	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp
		35					40						45		

Pro	Asn	Pro	Gln	Glu	Val	Val	Leu	Glu	Asn	Val	Thr	Glu	Asn	Phe	Asn
	50						55					60			

Met	Trp	Lys	Asn	Asn	Met	Val	Glu	Gln	Met	His	Glu	Asp	Ile	Ile	Ser
65						70				75					80

36291

Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys

85 90 95

Val Thr Leu Asn Cys Thr Asp Asp Val Arg Asn Val Thr Asn Asn Ala
100 105 110

Thr Asn Thr Asn Ser Ser Trp Gly Glu Pro Met Glu Lys Gly Glu Ile
115 120 125

Lys Asn Cys Ser Phe Asn Ile Thr Thr Ser Ile Arg Asn Lys Val Gln
130 135 140

Lys Gln Tyr Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn
145 150 155 160

Asp Ser Asn Asn Thr Asn Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val
165 170 175

Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His
180 185 190

Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys
195 200 205

Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr
 210 215 220

His Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser
 225 230 235 240

Leu Ala Glu Glu Glu Val Val Ile Arg Ser Glu Asn Phe Thr Asn Asn
 245 250 255

Ala Lys Thr Ile Met Val Gln Leu Asn Val Ser Val Glu Ile Asn Cys
 260 265 270

Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly
 275 280 285

Arg Ala Phe Tyr Thr Ala Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala
 290 295 300

His Cys Asn Ile Ser Arg Ala Asn Trp Asn Asn Thr Leu Arg Gln Ile
 305 310 315 320

Val Glu Lys Leu Gly Lys Gln Phe Gly Asn Asn Lys Thr Ile Val Phe
 325 330 335

Asn His Ser Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn
 340 345 350

Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser Thr Lys Leu Phe Asn Ser
 355 360 365

Thr Trp Thr Trp Asn Asn Ser Thr Trp Asn Asn Thr Lys Arg Ser Asn
 370 375 380

Asp Thr Glu Glu His Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile
 385 390 395 400

Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Arg
 405 410 415

Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg
 420 425 430

Asp Gly Gly Asn Asp Thr Ser Gly Thr Glu Ile Phe Arg Pro Gly Gly
 435 440 445

Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val
 450 455 460

Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg
 465 470 475 480

Val Val Gln Ser Glu Lys Ser Ala Val Gly Ile Gly Ala Val Phe Leu
 485 490 495

Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Met Thr
 500 505 510

Leu Thr Val Gln Ala Arg Leu Leu Leu Ser Gly Ile Val Gln Gln Gln
 515 520 525

Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu
 530 535 540

Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu
 545 550 555 560

Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly
565 570 575

Lys Leu Ile Cys Thr Thr Thr Val Pro Trp Asn Ala Ser Trp Ser Asn
580 585 590

Lys Ser Leu Asp Lys Ile Trp Asn Asn Met Thr Trp Met Glu Trp Glu
595 600 605

Arg Glu Ile Asn Asn Tyr Thr Ser Leu Ile Tyr Thr Leu Ile Glu Glu
610 615 620

Ser Gln Asn Gln Gln Glu Lys
625 630

<210> 9

<211> 630

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> trình tự Mos2.Env

<400> 9

Met Gly Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys
 1 5 10 15

Glu Ala Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu
 20 25 30

Lys Glu Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp
 35 40 45

Pro Asn Pro Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn
 50 55 60

Met Trp Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Arg
 65 70 75 80

Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys
 85 90 95

Val Thr Leu Glu Cys Arg Asn Val Arg Asn Val Ser Ser Asn Gly Thr
 100 105 110

Tyr Asn Ile Ile His Asn Glu Thr Tyr Lys Glu Met Lys Asn Cys Ser
 115 120 125

Phe Asn Ala Thr Thr Val Val Glu Asp Arg Lys Gln Lys Val His Ala
 130 135 140

Leu Phe Tyr Arg Leu Asp Ile Val Pro Leu Asp Glu Asn Asn Ser Ser
 145 150 155 160

Glu Lys Ser Ser Glu Asn Ser Ser Glu Tyr Tyr Arg Leu Ile Asn Cys
 165 170 175

Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro
 180 185 190

Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys
 195 200 205

Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr
 210 215 220

Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu
 225 230 235 240

Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn
 245 250 255

Leu Thr Asn Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Thr Val
 260 265 270

Asn Ile Thr Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg
 275 280 285

Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp
 290 295 300

Ile Arg Gln Ala His Cys Asn Leu Ser Arg Asp Gly Trp Asn Lys Thr
 305 310 315 320

Leu Gln Gly Val Lys Lys Lys Leu Ala Glu His Phe Pro Asn Lys Thr
 325 330 335

Ile Asn Phe Thr Ser Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His
 340 345 350

Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Gly Leu
 355 360 365

Phe Asn Gly Thr Tyr Met Pro Asn Gly Thr Asn Ser Asn Ser Ser Ser
 370 375 380

Asn Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln
 385 390 395 400

Glu Val Gly Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr
 405 410 415

Cys Arg Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser
 420 425 430

Asn Asn Gly Val Pro Asn Asp Thr Glu Thr Phe Arg Pro Gly Gly Gly
 435 440 445

Asp Met Arg Asn Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val
 450 455 460

Glu Val Lys Pro Leu Gly Val Ala Pro Thr Glu Ala Lys Arg Arg Val
 465 470 475 480

Val Glu Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly
 485 490 495

Ile Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu
 500 505 510

Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser
 515 520 525

Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr
 530 535 540

Val Trp Gly Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg
 545 550 555 560

Tyr Leu Gln Asp Gln Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly Lys
 565 570 575

Leu Ile Cys Thr Thr Ala Val Pro Trp Asn Thr Ser Trp Ser Asn Lys
 580 585 590

Ser Gln Thr Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp Lys
 595 600 605

Glu Ile Gly Asn Tyr Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu Ser
 610 615 620

Gln Asn Gln Gln Glu Lys
 625 630

<210> 10

<211> 6

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Trình tự đột biến vị trí phân cắt furin

<400> 10

Arg Arg Arg Arg Arg Arg
 1 5

<210> 11

<211> 31

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> trình tự tín hiệu

<400> 11

Met	Arg	Val	Arg	Gly	Ile	Leu	Arg	Asn	Trp	Gln	Gln	Trp	Trp	Ile	Trp
1				5					10					15	

Gly	Ile	Leu	Gly	Phe	Trp	Met	Leu	Met	Ile	Cys	Asn	Val	Val	Gly
			20					25					30	

<210> 12

<211> 8

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1

<400> 12

Asn	Pro	Asp	Trp	Leu	Pro	Asp	Met
1				5			

<210> 13

<211> 8

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1

<400> 13

Gly Ser Gly Ser Gly Ser Gly Ser

1

5

<210> 14

<211> 8

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1

<400> 14

Asp Asp Val His Pro Asp Trp Asp

1

5

<210> 15

<211> 8

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1

<400> 15

Arg Asp Thr Phe Ala Leu Met Met

1

5

<210> 16

<211> 8

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1

<400> 16

Asp Glu Glu Lys Val Met Asp Phe

1

5

<210> 17

<211> 8

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Ví dụ của 8 trình tự axit amin mà có thể thay thế vòng HR1

<400> 17

Asp Glu Asp Pro His Trp Asp Pro

1

5

<210> 18

<211> 29

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> trình tự tín hiệu

<400> 18

Met Arg Val Lys Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Arg Trp

1

5

10

15

Gly Thr Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala

20

25

<210> 19

<211> 61

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Trình tự thể

<400> 19

Ala Ala Ala Leu Pro Glu Thr Gly Gly Gly Ser Asp Tyr Lys Asp Asp
1 5 10 15

Asp Asp Lys Pro Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser Gly Gly Gly Gly Ser His His His His His His
50 55 60

<210> 20

<211> 616

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env ConC_SOSIP được làm ổn định (HIV160544)

<400> 20

Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
 1 5 10 15

Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Lys Glu
 20 25 30

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 35 40 45

Pro Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp
 50 55 60

Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp
 65 70 75 80

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 85 90 95

Leu Asn Cys Thr Asn Val Asn Val Thr Asn Thr Asn Asn Asn Asn Met
 100 105 110

Lys Glu Glu Met Lys Asn Cys Ser Phe Asn Thr Thr Thr Glu Ile Arg
 115 120 125

Asp Lys Lys Gln Lys Glu Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val
 130 135 140

Pro Leu Asn Glu Asn Ser Ser Glu Tyr Arg Leu Ile Asn Cys Asn Thr
 145 150 155 160

Ser Thr Ile Thr Gln Ile Cys Pro Lys Val Ser Phe Asp Pro Ile Pro
 165 170 175

Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn
 180 185 190

Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr Val Gln
 195 200 205

Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu Leu Asn
 210 215 220

Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr
 225 230 235 240

Asp Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Ser Val Glu Ile
 245 250 255

Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gly
 260 265 270

Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp Ile Arg
 275 280 285

Gln Ala His Cys Asn Ile Ser Glu Ala Lys Trp Asn Lys Thr Leu Gln
 290 295 300

Arg Val Lys Lys Lys Leu Lys Glu His Phe Pro Asn Lys Thr Ile Lys
 305 310 315 320

Phe Ala Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser Phe
 325 330 335

Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Lys Leu Phe Asn
 340 345 350

Ser Thr Tyr Asn Asn Thr Thr Ser Asn Ser Thr Ile Thr Leu Pro Cys
 355 360 365

Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Met
 370 375 380

Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Lys Ser Asn Ile Thr
 385 390 395 400

Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Asn Thr Glu
 405 410 415

Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu
 420 425 430

Leu Tyr Lys Tyr Lys Val Val Glu Ile Lys Pro Leu Gly Ile Ala Pro
 435 440 445

Thr Lys Cys Lys Arg Arg Val Val Glu Arg Arg Arg Arg Arg Arg Ala
 450 455 460

Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser
 465 470 475 480

Thr Met Gly Ala Ala Ser Asn Thr Leu Thr Val Gln Ala Arg Gln Leu
 485 490 495

Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala Pro Glu
 500 505 510

Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Phe Lys Gln Leu
 515 520 525

Gln Ala Arg Val Leu Ala Ile Glu Arg Tyr Leu Glu Val Gln Gln Leu
 530 535 540

Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr Ala Val
 545 550 555 560

Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Glu Asp Ile Trp Asp
 565 570 575

Asn Met Thr Trp Met Gln Trp Asp Arg Glu Ile Ser Asn Tyr Thr Asp
 580 585 590

Thr Ile Tyr Arg Leu Leu Glu Glu Ser Gln Phe Gln Gln Glu Ile Asn
 595 600 605

Glu Lys Asp Leu Leu Ala Leu Asp

610

615

<210> 21

<211> 634

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env BG505_SOSIP (HIV150673)

<400> 21

Ala Glu Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys

1

5

10

15

Asp Ala Glu Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu

20

25

30

Thr Glu Lys His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp

35

40

45

Pro Asn Pro Gln Glu Ile His Leu Glu Asn Val Thr Glu Glu Phe Asn

50

55

60

Met Trp Lys Asn Asn Met Val Glu Gln Met His Thr Asp Ile Ile Ser
 65 70 75 80

Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys
 85 90 95

Val Thr Leu Gln Cys Thr Asn Val Thr Asn Asn Ile Thr Asp Asp Met
 100 105 110

Arg Gly Glu Leu Lys Asn Cys Ser Phe Asn Met Thr Thr Glu Leu Arg
 115 120 125

Asp Lys Lys Gln Lys Val Tyr Ser Leu Phe Tyr Arg Leu Asp Val Val
 130 135 140

Gln Ile Asn Glu Asn Gln Gly Asn Arg Ser Asn Asn Ser Asn Lys Glu
 145 150 155 160

Tyr Arg Leu Ile Asn Cys Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro
 165 170 175

Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly
 180 185 190

Phe Ala Ile Leu Lys Cys Lys Asp Lys Lys Phe Asn Gly Thr Gly Pro
 195 200 205

Cys Pro Ser Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro Val
 210 215 220

Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val
 225 230 235 240

Met Ile Arg Ser Glu Asn Ile Thr Asn Asn Ala Lys Asn Ile Leu Val
 245 250 255

Gln Phe Asn Thr Pro Val Gln Ile Asn Cys Thr Arg Pro Asn Asn Asn
 260 265 270

Thr Arg Lys Ser Ile Arg Ile Gly Pro Gly Gln Ala Phe Tyr Ala Thr
 275 280 285

Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Val Ser Lys
 290 295 300

Ala Thr Trp Asn Glu Thr Leu Gly Lys Val Val Lys Gln Leu Arg Lys
 305 310 315 320

His Phe Gly Asn Asn Thr Ile Ile Arg Phe Ala Asn Ser Ser Gly Gly
 325 330 335

Asp Leu Glu Val Thr Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe
 340 345 350

Tyr Cys Asn Thr Ser Gly Leu Phe Asn Ser Thr Trp Ile Ser Asn Thr
 355 360 365

Ser Val Gln Gly Ser Asn Ser Thr Gly Ser Asn Asp Ser Ile Thr Leu
 370 375 380

Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Arg Ile Gly Gln
 385 390 395 400

Ala Met Tyr Ala Pro Pro Ile Gln Gly Val Ile Arg Cys Val Ser Asn
 405 410 415

Ile Thr Gly Leu Ile Leu Thr Arg Asp Gly Gly Ser Thr Asn Ser Thr
 420 425 430

Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg
 435 440 445

Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val
 450 455 460

Ala Pro Thr Arg Cys Lys Arg Arg Val Val Gly Arg Arg Arg Arg Arg
 465 470 475 480

Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala
 485 490 495

Gly Ser Thr Met Gly Ala Ala Ser Met Thr Leu Thr Val Gln Ala Arg
 500 505 510

Asn Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala
 515 520 525

Pro Glu Ala Gln Gln His Leu Leu Lys Leu Thr Val Trp Gly Ile Lys
 530 535 540

Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Arg Asp Gln
 545 550 555 560

Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr
 565 570 575

Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Arg Asn Leu Ser Glu Ile
 580 585 590

Trp Asp Asn Met Thr Trp Leu Gln Trp Asp Lys Glu Ile Ser Asn Tyr
 595 600 605

Thr Gln Ile Ile Tyr Gly Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu
 610 615 620

Lys Asn Glu Gln Asp Leu Leu Ala Leu Asp
 625 630

<210> 22

<211> 634

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env BG505_SOSIP được làm ổn định (HIV170863)

<400> 22

Ala	Glu	Asn	Leu	Trp	Val	Thr	Val	Tyr	Tyr	Gly	Val	Pro	Val	Trp	Lys
1				5					10					15	

Asp	Ala	Glu	Thr	Thr	Leu	Phe	Cys	Ala	Ser	Asp	Ala	Lys	Ala	Tyr	Glu
			20					25					30		

Thr	Glu	Lys	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp
		35					40					45			

Pro	Asn	Pro	Gln	Glu	Ile	His	Leu	Glu	Asn	Val	Thr	Glu	Glu	Phe	Asn
	50						55					60			

Met	Trp	Lys	Asn	Asn	Met	Val	Glu	Gln	Met	His	Thr	Asp	Ile	Ile	Ser
65						70				75				80	

Leu	Trp	Asp	Gln	Ser	Leu	Lys	Pro	Cys	Val	Lys	Leu	Thr	Pro	Leu	Cys
						85				90				95	

Val Thr Leu Gln Cys Thr Asn Val Thr Asn Asn Ile Thr Asp Asp Met
 100 105 110

Arg Gly Glu Leu Lys Asn Cys Ser Phe Asn Met Thr Thr Glu Leu Arg
 115 120 125

Asp Lys Lys Gln Lys Val Tyr Ser Leu Phe Tyr Arg Leu Asp Val Val
 130 135 140

Gln Ile Asn Glu Asn Gln Gly Asn Arg Ser Asn Asn Ser Asn Lys Glu
 145 150 155 160

Tyr Arg Leu Ile Asn Cys Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro
 165 170 175

Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly
 180 185 190

Phe Ala Ile Leu Lys Cys Lys Asp Lys Lys Phe Asn Gly Thr Gly Pro
 195 200 205

Cys Pro Ser Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro Val
 210 215 220

Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val
225 230 235 240

Met Ile Arg Ser Glu Asn Ile Thr Asn Asn Ala Lys Asn Ile Leu Val
245 250 255

Gln Phe Asn Thr Pro Val Gln Ile Asn Cys Thr Arg Pro Asn Asn Asn
260 265 270

Thr Arg Lys Ser Ile Arg Ile Gly Pro Gly Gln Ala Phe Tyr Ala Thr
275 280 285

Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Val Ser Lys
290 295 300

Ala Thr Trp Asn Glu Thr Leu Gly Lys Val Val Lys Gln Leu Arg Lys
305 310 315 320

His Phe Gly Asn Asn Thr Ile Ile Arg Phe Ala Asn Ser Ser Gly Gly
325 330 335

Asp Leu Glu Val Thr Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe
 340 345 350

Tyr Cys Asn Thr Ser Gly Leu Phe Asn Ser Thr Trp Ile Ser Asn Thr
 355 360 365

Ser Val Gln Gly Ser Asn Ser Thr Gly Ser Asn Asp Ser Ile Thr Leu
 370 375 380

Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Arg Ile Gly Gln
 385 390 395 400

Ala Met Tyr Ala Pro Pro Ile Gln Gly Val Ile Arg Cys Val Ser Asn
 405 410 415

Ile Thr Gly Leu Ile Leu Thr Arg Asp Gly Gly Ser Thr Asn Ser Thr
 420 425 430

Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg
 435 440 445

Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val
 450 455 460

Ala Pro Thr Arg Cys Lys Arg Arg Val Val Gly Arg Arg Arg Arg Arg
 465 470 475 480

Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala
 485 490 495

Gly Ser Thr Met Gly Ala Ala Ser Asn Thr Leu Thr Val Gln Ala Arg
 500 505 510

Asn Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Pro Arg Ala
 515 520 525

Pro Glu Ala Gln Gln His Leu Leu Lys Leu Thr Val Trp Gly Ile Lys
 530 535 540

Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Arg Val Gln
 545 550 555 560

Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr
 565 570 575

Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Arg Asn Leu Ser Glu Ile
 580 585 590

Trp Asp Asn Met Thr Trp Leu Gln Trp Asp Lys Glu Ile Ser Asn Tyr
 595 600 605

Thr Gln Ile Ile Tyr Gly Leu Leu Glu Glu Ser Gln Phe Gln Gln Glu
 610 615 620

Ile Asn Glu Gln Asp Leu Leu Ala Leu Asp
 625 630

<210> 23

<211> 619

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env C97ZA_SOSIP kiểu đại với L535M và Q567K
 (HIV150673)

<400> 23

Asn Met Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Thr Asp Ala
 1 5 10 15

Lys Thr Thr Leu Phe Cys Ala Ser Asp Thr Lys Ala Tyr Asp Arg Glu
 20 25 30

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 35 40 45

Pro Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp
 50 55 60

Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp
 65 70 75 80

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 85 90 95

Leu His Cys Thr Asn Ala Thr Phe Lys Asn Asn Val Thr Asn Asp Met
 100 105 110

Asn Lys Glu Ile Arg Asn Cys Ser Phe Asn Thr Thr Thr Glu Ile Arg
 115 120 125

Asp Lys Lys Gln Gln Gly Tyr Ala Leu Phe Tyr Arg Pro Asp Ile Val

130		135		140											
Leu	Leu	Lys	Glu	Asn	Arg	Asn	Asn	Ser	Asn	Asn	Ser	Glu	Tyr	Ile	Leu
145				150				155							160
Ile	Asn	Cys	Asn	Ala	Ser	Thr	Ile	Thr	Gln	Ala	Cys	Pro	Lys	Val	Asn
			165					170					175		
Phe	Asp	Pro	Ile	Pro	Ile	His	Tyr	Cys	Ala	Pro	Ala	Gly	Tyr	Ala	Ile
			180					185					190		
Leu	Lys	Cys	Asn	Asn	Lys	Thr	Phe	Ser	Gly	Lys	Gly	Pro	Cys	Asn	Asn
	195						200					205			
Val	Ser	Thr	Val	Gln	Cys	Thr	His	Gly	Ile	Lys	Pro	Val	Val	Ser	Thr
	210					215					220				
Gln	Leu	Leu	Leu	Asn	Gly	Ser	Leu	Ala	Glu	Lys	Glu	Ile	Ile	Ile	Arg
225				230					235						240
Ser	Glu	Asn	Leu	Thr	Asp	Asn	Val	Lys	Thr	Ile	Ile	Val	His	Leu	Asn
			245					250					255		

Lys Ser Val Glu Ile Val Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys

260

265

270

Ser Met Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile

275

280

285

Ile Gly Asp Ile Arg Gln Ala Tyr Cys Asn Ile Ser Gly Ser Lys Trp

290

295

300

Asn Glu Thr Leu Lys Arg Val Lys Glu Lys Leu Gln Glu Asn Tyr Asn

305

310

315

320

Asn Asn Lys Thr Ile Lys Phe Ala Pro Ser Ser Gly Gly Asp Leu Glu

325

330

335

Ile Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn

340

345

350

Thr Thr Arg Leu Phe Asn Asn Asn Ala Thr Glu Asp Glu Thr Ile Thr

355

360

365

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Gly Val Gly

370		375		380											
Arg	Ala	Met	Tyr	Ala	Pro	Pro	Ile	Ala	Gly	Asn	Ile	Thr	Cys	Lys	Ser
385					390					395					400
Asn	Ile	Thr	Gly	Leu	Leu	Leu	Val	Arg	Asp	Gly	Gly	Glu	Asp	Asn	Lys
				405						410				415	
Thr	Glu	Glu	Ile	Phe	Arg	Pro	Gly	Gly	Gly	Asn	Met	Lys	Asp	Asn	Trp
			420					425					430		
Arg	Ser	Glu	Leu	Tyr	Lys	Tyr	Lys	Val	Ile	Glu	Leu	Lys	Pro	Leu	Gly
		435					440					445			
Ile	Ala	Pro	Thr	Gly	Cys	Lys	Arg	Arg	Val	Val	Glu	Arg	Arg	Arg	Arg
	450					455					460				
Arg	Arg	Ala	Val	Gly	Ile	Gly	Ala	Val	Phe	Leu	Gly	Phe	Leu	Gly	Ala
465					470					475				480	
Ala	Gly	Ser	Thr	Met	Gly	Ala	Ala	Ser	Met	Thr	Leu	Thr	Val	Gln	Ala
				485				490						495	

Arg Gln Leu Leu Ser Ser Ile Val Gln Gln Gln Ser Asn Leu Leu Arg

500

505

510

Ala Pro Glu Ala Gln Gln His Met Leu Lys Leu Thr Val Trp Gly Ile

515

520

525

Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Lys Asp

530

535

540

Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys

545

550

555

560

Thr Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Thr Asp

565

570

575

Ile Trp Asn Asn Met Thr Trp Met Glu Trp Asp Arg Glu Ile Ser Asn

580

585

590

Tyr Thr Asp Thr Ile Tyr Arg Leu Leu Glu Asp Ser Gln Thr Gln Gln

595

600

605

Glu Lys Asn Glu Lys Asp Leu Leu Ala Leu Asp

610

615

<210> 24

<211> 619

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env C97ZA_SOSIP được sửa và được làm ổn định
(HIV170690)

<400> 24

Asn	Leu	Trp	Val	Thr	Val	Tyr	Tyr	Gly	Val	Pro	Val	Trp	Lys	Asp	Ala
1				5					10					15	

Lys	Thr	Thr	Leu	Phe	Cys	Ala	Ser	Asp	Ala	Lys	Ala	Tyr	Asp	Arg	Glu
			20					25					30		

Val	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp	Pro	Asn
		35					40						45		

Pro	Gln	Glu	Ile	Val	Leu	Glu	Asn	Val	Thr	Glu	Asn	Phe	Asn	Met	Trp
		50					55					60			

Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp
 65 70 75 80

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 85 90 95

Leu His Cys Thr Asn Ala Thr Phe Lys Asn Asn Val Thr Asn Asp Met
 100 105 110

Asn Lys Glu Ile Arg Asn Cys Ser Phe Asn Thr Thr Thr Glu Ile Arg
 115 120 125

Asp Lys Lys Gln Lys Val Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val
 130 135 140

Gln Leu Lys Glu Asn Arg Asn Asn Ser Asn Asn Ser Glu Tyr Arg Leu
 145 150 155 160

Ile Asn Cys Asn Thr Ser Thr Ile Thr Gln Ile Cys Pro Lys Val Thr
 165 170 175

Phe Asp Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile
 180 185 190

Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn
 195 200 205

Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr
 210 215 220

Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Lys Glu Ile Ile Ile Arg
 225 230 235 240

Ser Glu Asn Leu Thr Asp Asn Val Lys Thr Ile Ile Val His Leu Asn
 245 250 255

Lys Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys
 260 265 270

Ser Ile Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile
 275 280 285

Ile Gly Asp Ile Arg Gln Ala Tyr Cys Asn Ile Ser Gly Ser Lys Trp
 290 295 300

Asn Glu Thr Leu Lys Arg Val Lys Glu Lys Leu Arg Glu His Phe Asn
 305 310 315 320

Asn Asn Lys Thr Ile Lys Phe Ala Pro Ser Ser Gly Gly Asp Leu Glu
 325 330 335

Ile Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn
 340 345 350

Thr Thr Arg Leu Phe Asn Asn Asn Ala Thr Glu Asn Glu Thr Ile Thr
 355 360 365

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly
 370 375 380

Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Lys Ser
 385 390 395 400

Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Glu Asp Asn Lys
 405 410 415

Thr Glu Glu Ile Phe Arg Pro Gly Gly Gly Asn Met Lys Asp Asn Trp
 420 425 430

Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Ile Lys Pro Leu Gly
 435 440 445

Ile Ala Pro Thr Lys Cys Lys Arg Arg Asn Val Thr Arg Arg Arg Arg
 450 455 460

Arg Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala
 465 470 475 480

Ala Gly Ser Thr Met Gly Ala Ala Ser Asn Thr Leu Thr Val Gln Ala
 485 490 495

Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Pro Arg
 500 505 510

Ala Pro Glu Ala Gln Gln His Met Leu Lys Leu Thr Val Trp Gly Ile
 515 520 525

Lys Gln Leu Gln Ala Arg Val Leu Ala Ile Glu Arg Tyr Leu Glu Val
 530 535 540

Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys
 545 550 555 560

Thr Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Thr Asp
 565 570 575

Ile Trp Asn Asn Met Thr Trp Met Glu Trp Asp Arg Glu Ile Ser Asn
 580 585 590

Tyr Thr Asp Thr Ile Tyr Arg Leu Leu Glu Asp Ser Gln Phe Gln Gln
 595 600 605

Glu Ile Asn Glu Lys Asp Leu Leu Ala Asn Asp
 610 615

<210> 25

<211> 628

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Biểu thể của cấu trúc Du422 được sửa và được làm ổn định
 (HIV161818)

<400> 25

Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
 1 5 10 15

Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Lys Glu
 20 25 30

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 35 40 45

Pro Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp
 50 55 60

Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp
 65 70 75 80

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 85 90 95

Leu Asn Cys Lys Asn Val Asn Ile Ser Ala Asn Ala Asn Ala Thr Ala
 100 105 110

Thr Leu Asn Ser Ser Met Asn Gly Glu Ile Lys Asn Cys Ser Phe Asn

115	120	125
Thr Thr Thr Glu Leu Arg Asp Lys Lys Gln Lys Val Tyr Ala Leu Phe		
130	135	140
Tyr Lys Pro Asp Val Val Pro Leu Asn Gly Gly Glu His Asn Glu Thr		
145	150	155
Gly Glu Tyr Ile Leu Ile Asn Cys Asn Ser Ser Thr Cys Thr Gln Ala		
165	170	175
Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile His Tyr Cys Ala Pro		
180	185	190
Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr		
195	200	205
Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys		
210	215	220
Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu		
225	230	235
		240

Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn Asn Ile Lys Thr Ile
 245 250 255

Ile Val His Leu Asn Lys Ser Val Glu Ile Asn Cys Thr Arg Pro Asn
 260 265 270

Asn Asn Thr Arg Lys Ser Val Arg Ile Gly Pro Gly Gln Thr Phe Tyr
 275 280 285

Ala Thr Gly Glu Ile Ile Gly Asp Ile Arg Glu Ala His Cys Asn Ile
 290 295 300

Ser Arg Glu Thr Trp Asn Ser Thr Leu Ile Gln Val Lys Glu Lys Leu
 305 310 315 320

Arg Glu His Tyr Asn Lys Thr Ile Lys Phe Glu Pro Ser Ser Gly Gly
 325 330 335

Asp Leu Glu Val Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe
 340 345 350

Tyr Cys Asn Thr Thr Lys Leu Phe Asn Glu Thr Lys Leu Phe Asn Glu

355		360		365											
Ser	Glu	Tyr	Val	Asp	Asn	Lys	Thr	Ile	Ile	Leu	Pro	Cys	Arg	Ile	Lys
370					375					380					
Gln	Ile	Ile	Asn	Met	Trp	Gln	Glu	Val	Gly	Arg	Cys	Met	Tyr	Ala	Pro
385					390					395					400
Pro	Ile	Glu	Gly	Asn	Ile	Thr	Cys	Lys	Ser	Asn	Ile	Thr	Gly	Leu	Leu
				405					410					415	
Leu	Thr	Arg	Asp	Gly	Gly	Glu	Asn	Ser	Thr	Glu	Glu	Val	Phe	Arg	Pro
			420					425					430		
Gly	Gly	Gly	Asn	Met	Lys	Asp	Asn	Trp	Arg	Ser	Glu	Leu	Tyr	Lys	Tyr
	435						440					445			
Lys	Val	Val	Glu	Ile	Lys	Pro	Leu	Gly	Val	Ala	Pro	Thr	Lys	Cys	Lys
	450					455					460				
Arg	Lys	Asn	Val	Thr	Arg	Arg	Arg	Arg	Arg	Arg	Ala	Val	Gly	Leu	Gly
465					470					475				480	

Ala Val Leu Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala
 485 490 495

Ala Ser Asn Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile
 500 505 510

Val Gln Gln Gln Ser Asn Leu Pro Arg Ala Pro Glu Ala Gln Gln His
 515 520 525

Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Thr Arg Val
 530 535 540

Leu Ala Ile Glu Arg Tyr Leu Lys Val Gln Gln Leu Leu Gly Leu Trp
 545 550 555 560

Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr Ala Val Pro Trp Asn Ser
 565 570 575

Ser Trp Ser Asn Lys Ser Leu Gly Asp Ile Trp Asp Asn Met Thr Trp
 580 585 590

Met Gln Trp Asp Arg Glu Ile Ser Asn Tyr Thr Asn Thr Ile Tyr Arg

595

600

605

Leu Leu Glu Asp Ser Gln Phe Gln Gln Glu Lys Asn Glu Lys Asp Leu

610

615

620

Leu Ala Asn Asp

625

<210> 26

<211> 628

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env Du422_SOSIP được sửa và được làm ổn định
(HIV170859)

<400> 26

Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala

1

5

10

15

Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Lys Glu

20

25

30

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 35 40 45

Pro Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp
 50 55 60

Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp
 65 70 75 80

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 85 90 95

Leu Asn Cys Lys Asn Val Asn Ile Ser Ala Asn Ala Asn Ala Thr Ala
 100 105 110

Thr Leu Asn Ser Ser Met Asn Gly Glu Ile Lys Asn Cys Ser Phe Asn
 115 120 125

Thr Thr Thr Glu Leu Arg Asp Lys Lys Gln Lys Val Tyr Ala Leu Phe
 130 135 140

Tyr Lys Pro Asp Val Val Pro Leu Asn Gly Gly Glu His Asn Glu Thr
 145 150 155 160

Gly Glu Tyr Ile Leu Ile Asn Cys Asn Ser Ser Thr Ile Thr Gln Ala
 165 170 175

Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile His Tyr Cys Ala Pro
 180 185 190

Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr
 195 200 205

Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys
 210 215 220

Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu
 225 230 235 240

Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn Asn Ile Lys Thr Ile
 245 250 255

Ile Val His Leu Asn Lys Ser Val Glu Ile Asn Cys Thr Arg Pro Asn
 260 265 270

Asn Asn Thr Arg Lys Ser Val Arg Ile Gly Pro Gly Gln Thr Phe Tyr
 275 280 285

Ala Thr Gly Glu Ile Ile Gly Asp Ile Arg Glu Ala His Cys Asn Ile
 290 295 300

Ser Arg Glu Thr Trp Asn Ser Thr Leu Ile Gln Val Lys Glu Lys Leu
 305 310 315 320

Arg Glu His Tyr Asn Lys Thr Ile Lys Phe Glu Pro Ser Ser Gly Gly
 325 330 335

Asp Leu Glu Val Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe
 340 345 350

Tyr Cys Asn Thr Thr Lys Leu Phe Asn Glu Thr Lys Leu Phe Asn Glu
 355 360 365

Ser Glu Tyr Val Asp Asn Lys Thr Ile Ile Leu Pro Cys Arg Ile Lys
 370 375 380

Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Met Tyr Ala Pro
 385 390 395 400

Pro Ile Glu Gly Asn Ile Thr Cys Lys Ser Asn Ile Thr Gly Leu Leu
 405 410 415

Leu Thr Arg Asp Gly Gly Glu Asn Ser Thr Glu Glu Val Phe Arg Pro
 420 425 430

Gly Gly Gly Asn Met Lys Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr
 435 440 445

Lys Val Val Glu Ile Lys Pro Leu Gly Val Ala Pro Thr Lys Cys Lys
 450 455 460

Arg Lys Val Val Gly Arg Arg Arg Arg Arg Arg Ala Val Gly Leu Gly
 465 470 475 480

Ala Val Leu Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala
 485 490 495

Ala Ser Asn Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile
 500 505 510

Val Gln Gln Gln Ser Asn Leu Pro Arg Ala Pro Glu Ala Gln Gln His
515 520 525

Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Thr Arg Val
530 535 540

Leu Ala Ile Glu Arg Tyr Leu Glu Val Gln Gln Leu Leu Gly Leu Trp
545 550 555 560

Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr Ala Val Pro Trp Asn Ser
565 570 575

Ser Trp Ser Asn Lys Ser Leu Gly Asp Ile Trp Asp Asn Met Thr Trp
580 585 590

Met Gln Trp Asp Arg Glu Ile Ser Asn Tyr Thr Asn Thr Ile Tyr Arg
595 600 605

Leu Leu Glu Asp Ser Gln Phe Gln Gln Glu Ile Asn Glu Lys Asp Leu
610 615 620

Leu Ala Leu Asp
625

<210> 27

<211> 599

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env_DS_sC4_SOSIP được sửa và được làm ổn định
(HIV170686)

<400> 27

Met	Gly	Asn	Leu	Trp	Val	Thr	Val	Tyr	Tyr	Gly	Val	Pro	Val	Trp	Lys
1				5				10						15	

Asp	Ala	Lys	Thr	Thr	Leu	Phe	Cys	Ala	Ser	Asp	Ala	Lys	Ala	Tyr	Glu
			20				25					30			

Lys	Glu	Val	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp
		35				40					45				

Pro	Asn	Pro	Gln	Glu	Ile	Val	Leu	Gly	Asn	Val	Thr	Glu	Asn	Phe	Asn
	50					55					60				

Met	Trp	Lys	Asn	Asp	Met	Val	Asp	Gln	Met	His	Glu	Asp	Ile	Ile	Ser
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

36291

65					70					75					80
Leu	Trp	Asp	Gln	Ser	Leu	Lys	Pro	Cys	Val	Lys	Leu	Thr	Pro	Leu	Cys
				85					90					95	
Val	Thr	Leu	Asn	Cys	Arg	Asn	Val	Arg	Asn	Val	Glu	Met	Lys	Asn	Cys
			100					105					110		
Ser	Phe	Asn	Ala	Thr	Thr	Val	Val	Arg	Asp	Arg	Lys	Gln	Lys	Val	His
		115					120					125			
Ala	Leu	Phe	Tyr	Arg	Leu	Asp	Ile	Val	Pro	Leu	Asp	Glu	Asn	Asn	Ser
	130					135					140				
Ser	Tyr	Arg	Leu	Ile	Asn	Cys	Asn	Thr	Ser	Ala	Cys	Thr	Gln	Ile	Cys
145					150					155					160
Pro	Lys	Val	Ser	Phe	Asp	Pro	Ile	Pro	Ile	His	Tyr	Cys	Ala	Pro	Ala
				165					170					175	
Gly	Tyr	Ala	Ile	Leu	Lys	Cys	Asn	Asn	Lys	Thr	Phe	Asn	Gly	Thr	Gly
			180					185					190		

Pro Cys Asn Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro
 195 200 205

Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu
 210 215 220

Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn Asn Ala Lys Thr Ile Ile
 225 230 235 240

Val His Leu Asn Glu Thr Val Asn Ile Asn Cys Thr Arg Pro Asn Asn
 245 250 255

Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala
 260 265 270

Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Leu Ser
 275 280 285

Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly Val Lys Lys Lys Leu Ala
 290 295 300

Glu His Phe Pro Asn Lys Thr Ile Lys Phe Ala Pro His Ser Gly Gly

305		310		315		320									
Asp	Leu	Glu	Ile	Thr	Thr	His	Ser	Phe	Asn	Cys	Arg	Gly	Glu	Phe	Phe
				325					330					335	
Tyr	Cys	Asn	Thr	Ser	Asn	Leu	Phe	Asn	Glu	Ser	Asn	Ile	Glu	Arg	Asn
			340					345					350		
Asp	Ser	Ile	Ile	Thr	Leu	Pro	Cys	Arg	Ile	Lys	Gln	Ile	Ile	Asn	Met
		355					360					365			
Trp	Gln	Glu	Val	Gly	Arg	Cys	Met	Tyr	Ala	Pro	Pro	Ile	Ala	Gly	Asn
	370					375					380				
Ile	Thr	Cys	Arg	Ser	Asn	Ile	Thr	Gly	Leu	Leu	Leu	Thr	Arg	Asp	Gly
385					390				395					400	
Gly	Ser	Asn	Asn	Asn	Asp	Thr	Glu	Thr	Phe	Arg	Pro	Gly	Gly	Gly	Asp
			405						410				415		
Met	Arg	Asn	Asn	Trp	Arg	Ser	Glu	Leu	Tyr	Lys	Tyr	Lys	Val	Val	Glu
		420						425				430			

Val Lys Pro Leu Gly Val Ala Pro Thr Glu Cys Lys Arg Arg Val Val

435

440

445

Glu Arg Arg Arg Arg Arg Arg Ala Val Gly Ile Gly Ala Val Phe Leu

450

455

460

Gly Ile Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Asn Thr

465

470

475

480

Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln

485

490

495

Ser Asn Leu Pro Arg Ala Pro Glu Ala Gln Gln His Met Leu Gln Leu

500

505

510

Thr Val Trp Gly Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu

515

520

525

Arg Tyr Leu Glu Val Gln Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly

530

535

540

Lys Leu Ile Cys Cys Thr Ala Val Pro Trp Asn Thr Ser Trp Ser Asn

545 550 555 560

Lys Ser Gln Thr Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp
565 570 575

Lys Glu Ile Gly Asn Tyr Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu
580 585 590

Ser Gln Phe Gln Gln Glu Ile
595

<210> 28

<211> 607

<212> PRT

<213> Trình tự Nhân tạo

 $\langle 220 \rangle$

<223> Protein Env ConC SOSIP.v3 được làm ổn định (HIV170654)

<400> 28

Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala
1 5 10 15

Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Lys Glu

	20		25		30														
Val	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp	Pro	Asn				
	35					40					45								
Pro	Gln	Glu	Met	Val	Leu	Glu	Asn	Val	Thr	Glu	Asn	Phe	Asn	Met	Trp				
	50					55					60								
Lys	Asn	Asp	Met	Val	Asp	Gln	Met	His	Glu	Asp	Ile	Ile	Ser	Leu	Trp				
65					70					75					80				
Asp	Gln	Ser	Leu	Lys	Pro	Cys	Val	Lys	Leu	Thr	Pro	Leu	Cys	Val	Thr				
				85					90					95					
Leu	Asn	Cys	Thr	Asn	Val	Asn	Val	Thr	Glu	Met	Lys	Asn	Cys	Ser	Phe				
			100					105					110						
Asn	Thr	Thr	Thr	Glu	Ile	Arg	Asp	Lys	Lys	Gln	Lys	Glu	Tyr	Ala	Leu				
	115						120					125							
Phe	Tyr	Arg	Leu	Asp	Ile	Val	Pro	Leu	Asn	Glu	Asn	Ser	Ser	Glu	Tyr				
	130						135					140							

Arg Leu Ile Asn Cys Asn Thr Ser Thr Ile Thr Gln Ile Cys Pro Lys
 145 150 155 160

Val Ser Phe Asp Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr
 165 170 175

Ala Ile Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys
 180 185 190

Asn Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro Val Val
 195 200 205

Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile
 210 215 220

Ile Arg Ser Glu Asn Leu Thr Asn Asn Val Lys Thr Ile Ile Val His
 225 230 235 240

Leu Asn Glu Ser Val Glu Ile Val Cys Thr Arg Pro Asn Asn Asn Thr
 245 250 255

Arg Lys Ser Ile Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly

36291

```

260
265
270
Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser Glu Ala
275
280
285
Lys Trp Asn Lys Thr Leu Gln Arg Val Lys Lys Lys Leu Lys Glu His
290
295
300
Phe Pro Asn Lys Thr Ile Lys Phe Gln Pro Ser Ser Gly Gly Asp Leu
305
310
315
320
Glu Ile Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys
325
330
335
Asn Thr Ser Lys Leu Phe Asn Ser Thr Tyr Asn Asn Thr Thr Ser Asn
340
345
350
Ser Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp
355
360
365
Gln Glu Val Gly Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile
370
375
380

```

Thr Cys Lys Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly
 385 390 395 400

Asn Asn Asn Asn Asn Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met
 405 410 415

Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Ile
 420 425 430

Lys Pro Leu Gly Ile Ala Pro Thr Lys Cys Lys Arg Arg Val Val Glu
 435 440 445

Arg Arg Arg Arg Arg Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly
 450 455 460

Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Asn Thr Leu
 465 470 475 480

Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser
 485 490 495

Asn Leu Leu Arg Ala Pro Glu Ala Gln Gln His Met Leu Gln Leu Thr

500	505	510
Val Trp Gly Phe Lys Gln Leu Gln Ala Arg Val Leu Ala Ile Glu Arg		
515	520	525
Tyr Leu Glu Val Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys		
530	535	540
Leu Ile Cys Cys Thr Ala Val Pro Trp Asn Ser Ser Trp Ser Asn Lys		
545	550	555
Ser Gln Glu Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp Arg		
565	570	575
Glu Ile Ser Asn Tyr Thr Asp Thr Ile Tyr Arg Leu Leu Glu Glu Ser		
580	585	590
Gln Phe Gln Gln Glu Ile Asn Glu Lys Asp Leu Leu Ala Leu Asp		
595	600	605

<210> 29

<211> 634

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env BG505_SOSIP.v2 được làm ổn định (HIV171814)

<400> 29

Ala	Glu	Asn	Leu	Trp	Val	Thr	Val	Tyr	Tyr	Gly	Val	Pro	Val	Trp	Lys
1				5					10					15	

Asp	Ala	Glu	Thr	Thr	Leu	Phe	Cys	Ala	Ser	Asp	Ala	Lys	Ala	Tyr	Glu
			20					25					30		

Thr	Glu	Lys	His	Asn	Val	Trp	Ala	Thr	His	Ala	Cys	Val	Pro	Thr	Asp
		35					40					45			

Pro	Asn	Pro	Gln	Glu	Ile	His	Leu	Glu	Asn	Val	Thr	Glu	Glu	Phe	Asn
		50					55					60			

Met	Trp	Lys	Asn	Asn	Met	Val	Glu	Gln	Met	His	Thr	Asp	Ile	Ile	Ser
65						70				75				80	

Leu	Trp	Asp	Gln	Ser	Leu	Lys	Pro	Cys	Val	Lys	Leu	Thr	Pro	Leu	Cys
					85				90					95	

Val Thr Leu Gln Cys Thr Asn Val Thr Asn Asn Ile Thr Asp Asp Met
 100 105 110

Arg Gly Glu Leu Lys Asn Cys Ser Phe Asn Met Thr Thr Glu Leu Arg
 115 120 125

Asp Lys Lys Gln Lys Val Tyr Ser Leu Phe Tyr Arg Leu Asp Val Val
 130 135 140

Gln Ile Asn Glu Asn Gln Gly Asn Arg Ser Asn Asn Ser Asn Lys Glu
 145 150 155 160

Tyr Arg Leu Ile Asn Cys Asn Thr Ser Ala Cys Thr Gln Ala Cys Pro
 165 170 175

Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly
 180 185 190

Phe Ala Ile Leu Lys Cys Lys Asp Lys Lys Phe Asn Gly Thr Gly Pro
 195 200 205

Cys Pro Ser Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro Val

210		215		220
Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val				
225		230		235
				240
Met Ile Arg Ser Glu Asn Ile Thr Asn Asn Ala Lys Asn Ile Leu Val				
	245		250	255
Gln Phe Asn Thr Pro Val Gln Ile Asn Cys Thr Arg Pro Asn Asn Asn				
	260		265	270
Thr Arg Lys Ser Ile Arg Ile Gly Pro Gly Gln Ala Phe Tyr Ala Thr				
	275		280	285
Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Val Ser Lys				
	290		295	300
Ala Thr Trp Asn Glu Thr Leu Gly Lys Val Val Lys Gln Leu Arg Lys				
305		310		315
				320
His Phe Gly Asn Asn Thr Ile Ile Arg Phe Ala Asn Ser Ser Gly Gly				
	325		330	335

Asp Leu Glu Val Thr Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe
 340 345 350

Tyr Cys Asn Thr Ser Gly Leu Phe Asn Ser Thr Trp Ile Ser Asn Thr
 355 360 365

Ser Val Gln Gly Ser Asn Ser Thr Gly Ser Asn Asp Ser Ile Thr Leu
 370 375 380

Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Arg Ile Gly Gln
 385 390 395 400

Cys Met Tyr Ala Pro Pro Ile Gln Gly Val Ile Arg Cys Val Ser Asn
 405 410 415

Ile Thr Gly Leu Ile Leu Thr Arg Asp Gly Gly Ser Thr Asn Ser Thr
 420 425 430

Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg
 435 440 445

Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val

450		455		460											
Ala	Pro	Thr	Arg	Cys	Lys	Arg	Arg	Val	Val	Gly	Arg	Arg	Arg	Arg	Arg
465					470					475					480
Arg	Ala	Val	Gly	Ile	Gly	Ala	Val	Phe	Leu	Gly	Phe	Leu	Gly	Ala	Ala
				485					490					495	
Gly	Ser	Thr	Met	Gly	Ala	Ala	Ser	Asn	Thr	Leu	Thr	Val	Gln	Ala	Arg
			500					505					510		
Asn	Leu	Leu	Ser	Gly	Ile	Val	Gln	Gln	Gln	Ser	Asn	Leu	Pro	Arg	Ala
	515						520					525			
Pro	Glu	Ala	Gln	Gln	His	Leu	Leu	Lys	Leu	Thr	Val	Trp	Gly	Ile	Lys
	530					535					540				
Gln	Leu	Gln	Ala	Arg	Val	Leu	Ala	Val	Glu	Arg	Tyr	Leu	Glu	Val	Gln
545				550					555						560
Gln	Leu	Leu	Gly	Ile	Trp	Gly	Cys	Ser	Gly	Lys	Leu	Ile	Cys	Cys	Thr
			565					570					575		

Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Arg Asn Leu Ser Glu Ile
 580 585 590

Trp Asp Asn Met Thr Trp Leu Gln Trp Asp Lys Glu Ile Ser Asn Tyr
 595 600 605

Thr Gln Ile Ile Tyr Gly Leu Leu Glu Glu Ser Gln Phe Gln Gln Glu
 610 615 620

Ile Asn Glu Val Asp Leu Leu Ala Leu Asp
 625 630

<210> 30

<211> 619

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env C97ZA_SOSIP.v2 được sửa và được làm ổn định
 (HIV171810)

<400> 30

Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Asp Ala
 1 5 10 15

Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Arg Glu
 20 25 30

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
 35 40 45

Pro Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp
 50 55 60

Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp
 65 70 75 80

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 85 90 95

Leu His Cys Thr Asn Ala Thr Phe Lys Asn Asn Val Thr Asn Asp Met
 100 105 110

Asn Lys Glu Ile Arg Asn Cys Ser Phe Asn Thr Thr Thr Glu Ile Arg
 115 120 125

Asp Lys Lys Gln Lys Val Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val
 130 135 140

Gln Leu Lys Glu Asn Arg Asn Asn Ser Asn Asn Ser Glu Tyr Arg Leu
 145 150 155 160

Ile Asn Cys Asn Thr Ser Thr Cys Thr Gln Ile Cys Pro Lys Val Thr
 165 170 175

Phe Asp Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile
 180 185 190

Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn
 195 200 205

Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr
 210 215 220

Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Lys Glu Ile Ile Ile Arg
 225 230 235 240

Ser Glu Asn Leu Thr Asp Asn Val Lys Thr Ile Ile Val His Leu Asn
 245 250 255

Lys Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys
 260 265 270

Ser Ile Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile
 275 280 285

Ile Gly Asp Ile Arg Gln Ala Tyr Cys Asn Ile Ser Gly Ser Lys Trp
 290 295 300

Asn Glu Thr Leu Lys Arg Val Lys Glu Lys Leu Arg Glu His Phe Asn
 305 310 315 320

Asn Asn Lys Thr Ile Lys Phe Ala Pro Ser Ser Gly Gly Asp Leu Glu
 325 330 335

Ile Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn
 340 345 350

Thr Thr Arg Leu Phe Asn Asn Asn Ala Thr Glu Asn Glu Thr Ile Thr
 355 360 365

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly
 370 375 380

Arg Cys Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Lys Ser
 385 390 395 400

Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Glu Asp Asn Lys
 405 410 415

Thr Glu Glu Ile Phe Arg Pro Gly Gly Gly Asn Met Lys Asp Asn Trp
 420 425 430

Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Ile Lys Pro Leu Gly
 435 440 445

Ile Ala Pro Thr Lys Cys Lys Arg Arg Asn Val Thr Arg Arg Arg Arg
 450 455 460

Arg Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala
 465 470 475 480

Ala Gly Ser Thr Met Gly Ala Ala Ser Asn Thr Leu Thr Val Gln Ala
 485 490 495

Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Pro Arg
500 505 510

Ala Pro Glu Ala Gln Gln His Met Leu Lys Leu Thr Val Trp Gly Ile
515 520 525

Lys Gln Leu Gln Ala Arg Val Leu Ala Ile Glu Arg Tyr Leu Glu Val
530 535 540

Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Cys
545 550 555 560

Thr Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Thr Asp
565 570 575

Ile Trp Asn Asn Met Thr Trp Met Glu Trp Asp Arg Glu Ile Ser Asn
580 585 590

Tyr Thr Asp Thr Ile Tyr Arg Leu Leu Glu Asp Ser Gln Phe Gln Gln
595 600 605

Glu Ile Asn Glu Val Asp Leu Leu Ala Asn Asp

610

615

<210> 31

<211> 628

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env Du422_SOSIP.v1 được sửa và được làm ổn định
(HIV171812)

<400> 31

Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala

1

5

10

15

Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Lys Glu

20

25

30

Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn

35

40

45

Pro Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp

50

55

60

Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp
 65 70 75 80

Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
 85 90 95

Leu Asn Cys Lys Asn Val Asn Ile Ser Ala Asn Ala Asn Ala Thr Ala
 100 105 110

Thr Leu Asn Ser Ser Met Asn Gly Glu Ile Lys Asn Cys Ser Phe Asn
 115 120 125

Thr Thr Thr Glu Leu Arg Asp Lys Lys Gln Lys Val Tyr Ala Leu Phe
 130 135 140

Tyr Lys Pro Asp Val Val Pro Leu Asn Gly Gly Glu His Asn Glu Thr
 145 150 155 160

Gly Glu Tyr Ile Leu Ile Asn Cys Asn Ser Ser Thr Cys Thr Gln Ala
 165 170 175

Cys Pro Lys Val Ser Phe Asp Pro Ile Pro Ile His Tyr Cys Ala Pro

	180		185		190
Ala Gly Tyr Ala Ile Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr					
	195		200		205
Gly Pro Cys Asn Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys					
	210		215		220
Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu					
	225		230		235
					240
Glu Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn Asn Ile Lys Thr Ile					
		245		250	255
Ile Val His Leu Asn Lys Ser Val Glu Ile Asn Cys Thr Arg Pro Asn					
	260		265		270
Asn Asn Thr Arg Lys Ser Val Arg Ile Gly Pro Gly Gln Thr Phe Tyr					
	275		280		285
Ala Thr Gly Glu Ile Ile Gly Asp Ile Arg Glu Ala His Cys Asn Ile					
	290		295		300

Ser Arg Glu Thr Trp Asn Ser Thr Leu Ile Gln Val Lys Glu Lys Leu
 305 310 315 320

Arg Glu His Tyr Asn Lys Thr Ile Lys Phe Glu Pro Ser Ser Gly Gly
 325 330 335

Asp Leu Glu Val Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe
 340 345 350

Tyr Cys Asn Thr Thr Lys Leu Phe Asn Glu Thr Lys Leu Phe Asn Glu
 355 360 365

Ser Glu Tyr Val Asp Asn Lys Thr Ile Ile Leu Pro Cys Arg Ile Lys
 370 375 380

Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Cys Met Tyr Ala Pro
 385 390 395 400

Pro Ile Glu Gly Asn Ile Thr Cys Lys Ser Asn Ile Thr Gly Leu Leu
 405 410 415

Leu Thr Arg Asp Gly Gly Glu Asn Ser Thr Glu Glu Val Phe Arg Pro

420

425

430

Gly Gly Gly Asn Met Lys Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr

435

440

445

Lys Val Val Glu Ile Lys Pro Leu Gly Val Ala Pro Thr Lys Cys Lys

450

455

460

Arg Lys Val Val Gly Arg Arg Arg Arg Arg Arg Ala Val Gly Leu Gly

465

470

475

480

Ala Val Leu Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala

485

490

495

Ala Ser Asn Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile

500

505

510

Val Gln Gln Gln Ser Asn Leu Pro Arg Ala Pro Glu Ala Gln Gln His

515

520

525

Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Thr Arg Val

530

535

540

Leu Ala Ile Glu Arg Tyr Leu Glu Val Gln Gln Leu Leu Gly Leu Trp
 545 550 555 560

Gly Cys Ser Gly Lys Leu Ile Cys Cys Thr Ala Val Pro Trp Asn Ser
 565 570 575

Ser Trp Ser Asn Lys Ser Leu Gly Asp Ile Trp Asp Asn Met Thr Trp
 580 585 590

Met Gln Trp Asp Arg Glu Ile Ser Asn Tyr Thr Asn Thr Ile Tyr Arg
 595 600 605

Leu Leu Glu Asp Ser Gln Phe Gln Gln Glu Ile Asn Glu Val Asp Leu
 610 615 620

Leu Ala Leu Asp
 625

<210> 32

<211> 599

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> Protein Env sC4_SOSIP.v4 được sửa và được làm ổn định

<400> 32

Met Gly Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys
 1 5 10 15

Asp Ala Lys Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu
 20 25 30

Lys Glu Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp
 35 40 45

Pro Asn Pro Gln Glu Ile Val Leu Gly Asn Val Thr Glu Asn Phe Asn
 50 55 60

Met Trp Lys Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser
 65 70 75 80

Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys
 85 90 95

Val Thr Leu Asn Cys Arg Asn Val Arg Asn Val Glu Met Lys Asn Cys

	100		105		110										
Ser	Phe	Asn	Ala	Thr	Thr	Val	Val	Arg	Asp	Arg	Lys	Gln	Lys	Val	His
	115					120					125				
Ala	Leu	Phe	Tyr	Arg	Leu	Asp	Ile	Val	Pro	Leu	Asp	Glu	Asn	Asn	Ser
	130					135					140				
Ser	Tyr	Arg	Leu	Ile	Asn	Cys	Asn	Thr	Ser	Ala	Cys	Thr	Gln	Ile	Cys
145					150					155					160
Pro	Lys	Val	Ser	Phe	Asp	Pro	Ile	Pro	Ile	His	Tyr	Cys	Ala	Pro	Ala
				165					170					175	
Gly	Tyr	Ala	Ile	Leu	Lys	Cys	Asn	Asn	Lys	Thr	Phe	Asn	Gly	Thr	Gly
			180						185				190		
Pro	Cys	Asn	Asn	Val	Ser	Thr	Val	Gln	Cys	Thr	His	Gly	Ile	Lys	Pro
	195							200				205			
Val	Val	Ser	Thr	Gln	Leu	Leu	Leu	Asn	Gly	Ser	Leu	Ala	Glu	Glu	Glu
	210						215				220				

Ile Ile Ile Arg Ser Glu Asn Leu Thr Asn Asn Ala Lys Thr Ile Ile
 225 230 235 240

Val His Leu Asn Glu Thr Val Asn Ile Val Cys Thr Arg Pro Asn Asn
 245 250 255

Asn Thr Arg Lys Ser Ile Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala
 260 265 270

Thr Gly Asp Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Leu Ser
 275 280 285

Arg Asp Gly Trp Asn Lys Thr Leu Gln Gly Val Lys Lys Lys Leu Ala
 290 295 300

Glu His Phe Pro Asn Lys Thr Ile Lys Phe Ala Pro His Ser Gly Gly
 305 310 315 320

Asp Leu Glu Ile Thr Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe
 325 330 335

Tyr Cys Asn Thr Ser Asn Leu Phe Asn Glu Ser Asn Ile Glu Arg Asn

	340		345		350														
Asp	Ser	Ile	Ile	Thr	Leu	Pro	Cys	Arg	Ile	Lys	Gln	Ile	Ile	Asn	Met				
	355						360					365							
Trp	Gln	Glu	Val	Gly	Arg	Cys	Met	Tyr	Ala	Pro	Pro	Ile	Ala	Gly	Asn				
	370					375						380							
Ile	Thr	Cys	Arg	Ser	Asn	Ile	Thr	Gly	Leu	Leu	Leu	Thr	Arg	Asp	Gly				
385					390					395					400				
Gly	Ser	Asn	Asn	Asn	Asp	Thr	Glu	Thr	Phe	Arg	Pro	Gly	Gly	Gly	Asp				
			405						410					415					
Met	Arg	Asn	Asn	Trp	Arg	Ser	Glu	Leu	Tyr	Lys	Tyr	Lys	Val	Val	Glu				
		420						425					430						
Val	Lys	Pro	Leu	Gly	Val	Ala	Pro	Thr	Glu	Cys	Lys	Arg	Arg	Asn	Val				
	435						440					445							
Thr	Arg	Arg	Arg	Arg	Arg	Arg	Ala	Val	Gly	Ile	Gly	Ala	Val	Phe	Leu				
	450					455					460								

Gly Ile Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Asn Thr
 465 470 475 480

Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln
 485 490 495

Ser Asn Leu Pro Arg Ala Pro Glu Ala Gln Gln His Met Leu Gln Leu
 500 505 510

Thr Val Trp Gly Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu
 515 520 525

Arg Tyr Leu Glu Asp Gln Gln Leu Leu Gly Leu Trp Gly Cys Ser Gly
 530 535 540

Lys Leu Ile Cys Cys Thr Ala Val Pro Trp Asn Thr Ser Trp Ser Asn
 545 550 555 560

Lys Ser Gln Thr Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp
 565 570 575

Lys Glu Ile Gly Asn Tyr Thr Gly Glu Ile Tyr Arg Leu Leu Glu Glu

580

585

590

Ser Gln Phe Gln Gln Glu Ile

595

<210> 33

<211> 29

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> trình tự tín hiệu

<400> 33

Met Arg Val Arg Gly Met Leu Arg Asn Trp Gln Gln Trp Trp Ile Trp

1

5

10

15

Ser Ser Leu Gly Phe Trp Met Leu Met Ile Tyr Ser Val

20

25

<210> 34

<211> 29

<212> PRT

<213> Trình tự Nhân tạo

<220>

<223> trình tự tín hiệu

<400> 34

Met	Arg	Val	Met	Gly	Ile	Gln	Arg	Asn	Cys	Gln	His	Leu	Phe	Arg	Trp
1				5					10					15	

Gly	Thr	Met	Ile	Leu	Gly	Met	Ile	Ile	Ile	Cys	Ser	Ala
				20					25			