

Piotr Kruszyński

**Analiza zmienności w obrębie genu *env* szczepów
ludzkiego wirusa niedoboru odporności typu 1
występujących na terenie Polski**

Rozprawa na stopień doktora nauk medycznych
w dyscyplinie biologia medyczna

Promotor: dr hab. n. med. Tomasz J. Wąsik, prof. nadzw. SUM

Promotor pomocniczy: dr Joanna Smoleń-Dzirba

Śląski Uniwersytet Medyczny

Wydział Farmaceutyczny z Oddziałem Medycyny Laboratoryjnej

Katedra i Zakład Mikrobiologii i Wirusologii

Sosnowiec 2016

STRESZCZENIE

Wstęp: Genom HIV-1 cechuje się wysoką zmiennością sekwencji, szczególnie w obrębie genu *env*. Stwierdzono, że sekwencja odcinka genu *env* kodującego pętlę V3 determinuje powinowactwo HIV-1 względem koreceptora CCR5 lub CXCR4. Określenie powinowactwa jest konieczne przed wprowadzeniem do leczenia antagonisty CCR5 – marawiroku. Z kolei oszacowanie zróżnicowania genetycznego jest pomocne w monitorowaniu zakażeń i przepływu wariantów wirusa. Celem prezentowanej pracy jest określenie powinowactwa wirusa do koreceptorów z użyciem kilku metod genotypowych oraz analizy porównawcze sekwencji genu *env* pochodzących od polskich pacjentów seropozytywnych względem HIV-1.

Material i metody: Retrospektywnym analizom zostało poddanych 101 sekwencji genu *env* pochodzących od pacjentów zdiagnozowanych w latach 2008-2010, wśród których u 19 stwierdzono zakażenie wczesne HIV-1, a u 82 – dawne. Sekwencje pochodziły z prowirusowego DNA izolowanego z krwi obwodowej. Spośród badanych pacjentów, 63 było zakażonych drogą kontaktów homoseksualnych, 25 drogą kontaktów heteroseksualnych, u 6 infekcja nastąpiła poprzez stosowanie środków odurzających w iniekcjach, zaś dla pozostałych 7 osób droga zakażenia była nieznana. Określenie powinowactwa do koreceptorów odbywało się poprzez zastosowanie reguły 11/25, obliczanie wypadkowego ładunku pętli V3, oraz z użyciem narzędzi bioinformatycznych geno2pheno i PSSM. Analizy zróżnicowania sekwencji genu *env* prowadzono za pomocą programu NCBI Genotyping Tool, Hypermut, MEGA, VESPA i SNAP.

Wyniki: Po eliminacji 5 sekwencji z hipermutacjami, powinowactwo do koreceptora CXCR4 wykazano, w następstwie użycia reguły 11/25, ładunku V3, narzędzia geno2pheno i PSSM odpowiednio w przypadku 7 (7.29%), 10 (10.42%), 45 (46.88%) i 5 (5.21%) sekwencji. Nie stwierdzono istnienia statystycznie istotnych zależności pomiędzy częstością wykazywania powinowactwa do CXCR4 a etapem infekcji i drogą przeniesienia zakażenia. Stwierdzono także niedostateczną zgodność rezultatów użycia poszczególnych metod szacowania powinowactwa. Dominującym wariantem genetycznym HIV-1 był podtyp B (96.87%), pozostałe 3.13% sekwencji reprezentowało podpodtyp A1. Średnia odległość nukleotydowa była wyższa dla sekwencji pochodzących od pacjentów z dawnym zakażeniem niż dla sekwencji uzyskanych od pacjentów z zakażeniem wczesnym. Analizy filogenetyczne wykazały grupowanie się sekwencji pochodzących od pacjentów zakażonych tą samą drogą i zdiagnozowanych w tym samym mieście. Analizy wzorców kodonów wykazały dla dwóch pozycji znacząco różne częstości występowania tripletów nukleotydowych odpowiadających

aminokwasom charakterystycznym pomiędzy grupami sekwencji wirusów o niejednakowym powinowactwie do koreceptorów.

Dyskusja: Identyfikacja podtypu B dla większości analizowanych sekwencji była zgodna z wynikami innych europejskich badań. Stwierdzona w prezentowanej pracy średnia odległość nukleotydowa była zbliżona do zaobserwowanej w innych badaniach nad podtypem B. Występowanie na drzewie filogenetycznym pięciu skupisk grupujących sekwencje pochodzące od osób o jednakowej drodze zakażenia i miejscu diagnozy infekcji może świadczyć o istnieniu powiązań epidemiologicznych pomiędzy tymi pacjentami. Stwierdzono istnienie zarówno podobieństw, jak i rozbieżności pomiędzy częstością wykazywania powinowactwa do koreceptora CXCR4 w prezentowanej pracy a rezultatami badań innych autorów. Zaobserwowane niejednakowe wyniki mogą być następstwem zarówno zastosowania różnych wartości granicznych dla określania powinowactwa do CXCR4, jak i różnych metod rozróżniania zakażeń wczesnych od dawnych.

Wnioski: Powszechność stwierdzania powinowactwa do koreceptora CXCR4 wskazuje na konieczność rutynowego testowania powinowactwa przed wprowadzeniem do leczenia antagonisty CCR5. Rozbieżności wyników oznaczania powinowactwa do koreceptorów z użyciem różnych metod genotypowych przemawiają za koniecznością ich udoskonalania. Metody filogenetyczne pozwoliły na identyfikację łańcuchów transmisji HIV-1 mogących świadczyć o lokalnym szerzeniu się wirusa.

Słowa kluczowe: HIV-1, gen *env*, glikoproteina gp120, pętla V3, analiza sekwencji, powinowactwo do koreceptorów, zróżnicowanie genetyczne

ABSTRACT

Introduction: Genome of the HIV-1 is characterized by high sequence variability, especially within the *env* gene. The sequence of the *env* spanning the V3 loop coding region determines HIV-1's affinity for the CCR5 or CXCR4 coreceptor. Determination of tropism is necessary before initiation of therapy with CCR5's antagonist – maraviroc. Assessment of genetic diversity is useful in monitoring the infections and transmission of viral variants. The aim of the presented study is to determine viral affinity for the coreceptors with the use of several genotypic methods and comparative analyses of the *env* gene sequences derived from Polish HIV-1 seropositive patients.

Material and methods: Retrospective analyses were performed on 101 *env* gene sequences obtained from patients diagnosed between 2008 and 2010, including 19 with recent HIV-1 and 82 with long-term infection respectively. All sequences had been derived from proviral DNA isolated from peripheral blood. The route of HIV-1 infection was homosexual intercourses for 63 patients, heterosexual contacts for 25 patients, injecting drug usage for 6 patients and unknown for the remaining 7 persons. The affinity for the coreceptors was determined with the 11/25 rule, V3 net charge calculation, as well as the geno2pheno, and PSSM bioinformatic tools. Analyses of the *env* gene diversity were conducted with the use of NCBI Genotyping Tool, Hypermut, MEGA, VESPA and SNAP programs.

Results: Five sequences were excluded due to the presence of hypermutations. Affinity for the CXCR4 coreceptor was identified with the use of the 11/25 rule, V3 charge and geno2pheno and PSSM tools in 7 (7.29%), 10 (10.42%), 45 (46.88%) and 5 (5.21%) sequences, respectively. There were no statistically significant associations between the frequency of affinity for the CXCR4 coreceptor and the stage or route of HIV-1 infection. Additionally incomplete correlation between the results of particular methods of the affinity's determination was observed. Subtype B was determined to be predominant HIV-1 genetic variant (96.87%) whereas the remaining 3.13% of sequences represented the A1 subsubtype. Mean nucleotide distance was higher for the sequences derived from the patients with long-term infection than for the sequences obtained from the patients with recent infection. Phylogenetic analyses revealed grouping of sequences derived from patients with the same infection route and diagnosed in the same place. Analyses of the codons' patterns for two positions revealed frequencies significantly different between the groups of sequences of viruses with various coreceptor's affinity.

Discussion: Identification of the subtype B as a dominant HIV-1 variant was consistent with the results of other European studies. Mean nucleotide distance established in presented study was similar to that observed in other research studies on the B subtype. Identification of five transmission clusters within a phylogenetic tree with sequences obtained from persons of the same route of infection and place of diagnosis suggests the existence of epidemiological relationships between these persons. Both similarities and discrepancies were found between the frequency of determining of the affinity for the CXCR4 coreceptor observed in the presented study and those noticed in other authors' analyses, what can be a consequence of both application of different cutoffs for determining of affinity to CXCR4 and various methods used for discrimination of recent and long-term infections.

Conclusions: Prevalence of the affinity for the CXCR4 coreceptor in the analyzed population indicates the need for routine affinity testing before treatment with maraviroc. Discrepancies between the results of different methods used for determination of coreceptors' affinity support the necessity of their improvement. Phylogenetic methods allowed for identification of HIV-1 transmission chains which may indicate local viral spread.

Key words: HIV-1, *env* gene, gp120 glycoprotein, V3 loop, sequence analysis, affinity for the coreceptors, genetic diversity