Genetic characterisation of *Langerin* gene in human immunodeficiency virus-1-infected women from Bahia, Brazil

Giselle Calasans de Souza Costa[1,2], Jaqueline Goes Jesus[1,2],
Filipe Ferreira de Almeida Rego[1], Edson Souza Santos[2], Bernardo Galvão-Castro[1,2],
Marilda de Souza Gonçalves[1,3], Luiz Carlos Júnior Alcantara[1]/+

[1]Laboratório de Hematologia, Genética e Biologia Computacional, Centro de Pesquisas Gonçalo Moniz-Fiocruz, Salvador, BA, Brasil
[2]Escola Bahiana de Medicina e Saúde Pública, Fundação Bahiana para o Desenvolvimento das Ciências, Salvador, BA, Brasil
[3]Faculdade de Farmácia, Universidade Federal da Bahia, Salvador, BA, Brasil

Studies on human genetic variations are a useful source of knowledge about human immunodeficiency virus (HIV)-1 infection. The Langerin protein, found at the surface of Langerhans cells, has an important protective role in HIV-1 infection. Differences in Langerin function due to host genetic factors could influence susceptibility to HIV-1 infection. To verify the frequency of mutations in the Langerin *gene, 118 samples from HIV-1-infected women and 99 samples from HIV-1-uninfected individuals were selected for sequencing of the promoter and carbohydrate recognition domain (CRD)-encoding regions of the* Langerin *gene.* Langerin *promoter analysis revealed two single nucleotide polymorphisms (SNPs) and one mutation in both studied groups, which created new binding sites for certain transcription factors, such as NFAT5, HOXB9.01 and STAT6.01, according to MatInspector software analysis. Three SNPs were observed in the CRD-encoding region in HIV-1-infected and uninfected individuals: p.K313I, c.941C>T and c.983C>T. This study shows that mutations in the* Langerin *gene are present in the analysed populations at different genotypic and allelic frequencies. Further studies should be conducted to verify the role of these mutations in HIV-1 susceptibility.*

Key words: HIV-1 - *Langerin* - polymorphism

Dendritic cells (DCs) play an important role in early human immunodeficiency virus (HIV)-1 infection due to their ability to mostly capture the virus in peripheral tissues, such as the genital mucosa; migrate to the secondary lymphoid organs and present processed HIV-1 antigens to CD4 T lymphocytes, their major cell target (Ahmed et al. 2013). Different DC subsets interact with HIV-1 via selected C-type lectin receptors (CLRs), independently of CD4 and chemokine co-receptors (CCR5 and CXCR4), which is the classical pathway used by HIV-1 to enter CD4 T lymphocytes and macrophages. These CLRs are used to distinguish DC subsets in mucosal tissues (Tada et al. 2006, Hirbod et al. 2009, van den Berg & Geijtenbeek 2013).

In the genital mucosa, two different DC subsets are present: Langerhans cells (LCs), which specifically express the C-type lectin Langerin (CD207) and DC-SIGN[+]-DCs, characterised by expression of the CLR DC-SIGN. These subsets have a specific anatomical localisation: whereas LCs reside in the stratified mucosal epithelia of the vagina, the ectocervix, the epidermis of the foreskin, the glans penis and the skin, DC-SIGN[+]-DCs are located in the subepithelial layer. Thus, HIV-1 encounters different DC subsets, depending on the site of infection and the integrity of the epithelial layer. Due to their location, LCs are potentially the first DC subset to encounter HIV-1 during sexual transmission within intact genital epithelial tissues (de Witte et al. 2007a, b, Hirbod et al. 2009, 2011).

Langerin is a cell-surface C-type lectin located exclusively on LCs and is constitutively associated with Birbeck granules, which are endosomal recycling organelles consisting of superimposed and zippered membranes. Antigen capture is a function of Langerin and the induction of Birbeck granule formation in fibroblasts is a consequence of this antigen processing activity (Ward et al. 2006, de Witte et al. 2007a, Mesman & Geijtenbeek 2012). The Langerin protein comprises an extracellular neck, transmembrane and intracellular portions and a carbohydrate recognition domain (CRD), which preferentially binds glycoconjugates containing mannose, N-acetylglucosamine and fucose (Valladeau et al. 2002, Stambach & Taylor 2003, McGreal et al. 2004, Tada et al. 2006).

Regarding HIV-1, Langerin has the ability to bind the mannan-rich region of gp120 through the CRD, capturing and internalising this virus into Birbeck granules. These granules degrade HIV-1 and prevent LC infection and viral dissemination at low virus concentrations, providing a different pathway from classical antigen processing. However, HIV-1 transmission to T cells mediated by LCs can occur at high viral loads (VLs), suggesting that the

inhibition of Langerin by inhibitors or saturation enables HIV-1 binding to CD4 and CCR5 and subsequent LC infection and transmission to T cells (Ward et al. 2006, de Witte et al. 2007a, b, Schwartz 2007, Feinberg et al. 2010, van der Vlist & Geijtenbeek 2010).

As Langerin on LCs functions as a protective mechanism in intact mucosa by enabling HIV-1 clearance, during sexual transmission of the virus, differences in Langerin function due to genetic, cellular or infectious factors might influence susceptibility to HIV-1 infection (de Witte et al. 2007b). In this context, polymorphisms in the *Langerin* gene that result in different protein carbohydrate-binding capacities and/or that influence protein expression could result in different HIV-1 susceptibilities and transmission rates by LCs. Certain polymorphisms that result in amino acid changes in the CRD of human Langerin have been identified (Verdijk et al. 2005, Ward et al. 2006).

In an effort to verify the presence of *Langerin* gene polymorphisms and their frequencies in HIV-1-infected women and HIV-1-uninfected individuals from state of Bahia (BA), Brazil, the promoter and CRD-encoding regions of the human *Langerin* gene were analysed.

SUBJECTS, MATERIALS AND METHODS

Study patients - HIV-1-infected women - The promoter and CRD-encoding regions of the *Langerin* gene were studied in convenience samples of 116 and 118 HIV-1-infected women, respectively, who were recruited during a gynaecological follow-up (for technical reasons, analyses of 2 samples could not be performed in the study of the promoter region of the *Langerin* gene). These patients were followed at the Reference Centre for Sexual Transmitted Disease of Feira de Santana, BA, a facility that provides free patient management. In 2007, blood samples were collected from all of the patients who agreed to participate in the study and were then processed at the Advanced Laboratory of Public Health, Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, BA. These patients represent a low-income population, with a high risk of infection with sexually transmitted diseases. Clinical and epidemiological data were obtained from medical records. Informed consent was obtained from the patients.

HIV-1-uninfected group - Ninety-nine DNA samples from HIV-1-uninfected individuals were randomly selected to obtain the frequency of *Langerin* gene polymorphisms in the HIV-1-uninfected group. These samples were obtained from volunteers who attended the Faculty of Pharmacy of the Federal University of Bahia, Salvador, BA, for free laboratory analyses and represent a low-income population that presents a high risk of infection with several sexually transmitted diseases, including HIV-1.

The local ethical committee's approval was obtained and all subjects provided written informed consent.

DNA extraction and mutation detection - Genomic DNA was extracted from peripheral blood mononuclear cells using the QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA, USA). Analyses of the promoter and

CRD-encoding regions of the *Langerin* gene were performed through polymerase chain reaction (PCR) amplification followed by sequencing. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) and sequenced in an ABI Prism 3100 DNA Sequencer (Applied Biosystems, Foster City, CA, USA). The obtained sequences were analysed with the SeqScape programme (Applied Biosystems) and the nucleotide variations observed in the sequences were compared with the wild-type *Langerin* gene sequence (NM_015717.3) and confirmed using the BioEdit (Hall 1999) and GeneDoc (Nicholas et al. 1997) programmes.

Promoter region - The promoter region of the *Langerin* gene was analysed in two separate amplification reactions to improve the amplification of this region, which was approximately 900 bp in length. The initial 331-bp long fragment was amplified with the following constructed primers: 5'AGAGGTCCAGCTCCTTGACATGAT3'(F) and 5'AGTCCTTGCACCAAGCAGCTCATA3'(R). The last 429-bp long fragment was amplified with the following constructed primers: 5'AATGGATGGGTGGATGGATGACGA3'(F) and 5'CAAAGGGCAGGTTTGCACACAGAA3'(R).

CRD-encoding region - The Langerin CRD-encoding region, which is encoded by three exons of the *Langerin* gene, was also analysed in two separate parts to eliminate the amplification of introns. Two exons were amplified in the same 556-bp long fragment, using the following primers: 5'AGGAAGAAGCTTGAGAGGCCAGTT3'(F) and 5'ACCCACCACTTTCAAGTCCCTACA3'(R). The third exon was amplified using the primer sequences 5'TTGGGTGCAGACATTTGCTATGCC3'(F) and 5'ATTTCCAGCCAAGACAGACGGACT3'(R) and this amplification resulted in a 445-bp long fragment.

Transcription factor binding site analyses - To investigate whether the detected mutations in the *Langerin* promoter region could modify the possible transcription factor binding sites, mutated and wild-type sequences were analysed using the MatInspector tool implemented in the Genomatix programme (Cartharius et al. 2005).

Physicochemical and potential protein domain analyses - To investigate possible influences of the mutations on protein structure, a physicochemical analysis of the missense mutations observed in the sequences was performed using Network Protein Sequence Analysis (npsa-pbil.ibcp.fr/) (Combet et al. 2000). Sulfation and O-glycosylation sites and other posttranslational modification sites were analysed using the Prosite tool implemented in the GeneDoc programme. The search for potential protein domains was performed using the Pfam database (Finn et al. 2010). Finally, the SWISS-MODEL online tool (swissmodel.expasy.org/) (Arnold et al. 2006) was used as a fully automated protein structure homology-modelling server to infer the possible influence of the amino acid changes on protein secondary structure.

Statistical methods - Allelic frequencies were estimated by direct allele counting. Conformity with Hardy-Weinberg equilibrium and the existence of linkage disequilibrium were tested using Genepop v.3.4 (Ray-

mon & Rousset 1995). The mean, median and other statistical parameters were estimated by BioEstat v.5.0 (Ayres et al. 2005).

RESULTS

Sample characterisation - The sample in this study consisted of 118 women infected with HIV-1 and 99 HIV-1-uninfected individuals (49 males and 50 females). The ages of the HIV-1-infected women ranged from 20-73 years (median age = 33 ± 17.6 years). Clinical factors, such as the duration of antiretroviral treatment, the appearance of HIV-1 constitutional symptoms, VL and CD4 T cell counts, among the individuals infected with HIV-1 showed high heterogeneity in the group. The median HIV-1 VL was 386 ± 17.93 copies/mL and the median CD4 T cell count was 382.5 ± 2.03 cells/μL. In total, 88 (75.9%) patients were receiving antiretroviral therapy, whereas 23 (24.1%) were therapy-naïve. The main route of HIV-1 infection among women was sexual transmission (66.2%). One patient reported infection via a blood transfusion (1.5%) and one was infected during intravenous drug use (2.2%). Seven patients did not have data available in their medical records. The median age of the HIV-1-uninfected individuals was 29 ± 15.2 years. All of the blood samples from the HIV-1-uninfected group had negative serology for HIV-1/2 and human T-lymphotropic virus-1/2.

Molecular analysis of Langerin gene promoter region - The studies on the *Langerin* gene promoter revealed three mutations, all characterised by a substitution of a thymine nucleotide with a cytosine nucleotide, at the -577, -517 and -160 positions of the *Langerin* gene promoter region (-577T>C, -517T>C and -160T>C). Except for the frequency -577T>C mutation in the HIV-1-uninfected group (0.5%), the mutations were considered as polymorphic sites in the studied groups.

The genotypic and allelic frequencies of the analysed sites are shown in Table and did not differ significantly from the frequencies expected under Hardy-Weinberg equilibrium (-577T>C: p = 0.6238, χ^2 = 7.3786; -517T>C: p = 0.4598, χ^2 = 6.0821; -160T>C: p = 0.2608, χ^2 = 5.2686). Linkage disequilibrium was not observed between any of the analysed polymorphisms when HIV-1-infected women and HIV-1-uninfected individuals were compared.

According to the transcription factor binding site analyses accomplished using the MatInspector tool, the -577T>C mutation created a new binding site for the transcription factor HOXB9.01 (homeobox 9.01), from the family abdominal B-type domain, with the highest level of similarity (1.0). Regarding the single nucleotide polymorphisms (SNPs) -517T>C, the analyses revealed that it was localised to the autoimmune regulatory element 01 (AIRE 01) binding site and that the nucleotide change from thymine to cytosine reduced the similarity of the core sequence from 0.857-0.774 and created a new binding site for the transcription factor nuclear factor (NF) of activated T cells 5 (NFAT 5) with similarity of 1.0. Regarding the -160T>C polymorphism, the substitution of thymine with cytosine created two new binding

TABLE

Allelic and genotypic frequencies of -577T>C, -517T>C, -160T>C, p.K313I, c.941C>T and c.983C>T mutations in human immunodeficiency virus (HIV)-1 infected women and HIV-1 non-infected group

Mutation	Genotype n (%)			Allele n (%)	
	Wild type	Heterozygous	Mutant	Wild type	Mutant
-577T>C					
HIV-1 positive	112 (96.5)	4 (3.5)	0 (0)	228 (98.3)	4 (1.7)
HIV-1 negative	97 (99)	1 (1)	0 (0)	195 (99.5)	1 (0.5)
-517T>C					
HIV-1 positive	93 (80.2)	22 (19)	1 (0.8)	208 (89.7)	24 (10.3)
HIV-1 negative	73 (74.5)	24 (24.5)	1 (1)	170 (86.7)	26 (13.3)
-160T>C					
HIV-1 positive	22 (20)	65 (59.1)	23 (20.9)	109 (49.5)	111 (50.5)
HIV-1 negative	23 (23.2)	46 (46.5)	30 (30.3)	92 (46.5)	106 (53.5)
p.K313I					
HIV-1 positive	95 (80.5)	22 (18.7)	1 (0.8)	212 (89.8)	24 (10.2)
HIV-1 negative	77 (78.6)	21 (21.4)	0 (0)	175 (89.3)	21 (10.7)
c.941C>T					
HIV-1 positive	83 (70.3)	31 (26.3)	4 (3.4)	197 (83.5)	39 (16.5)
HIV-1 negative	70 (71.4)	24 (24.5)	4 (4.1)	164 (83.7)	32 (16.3)
c.983C>T					
HIV-1 positive	68 (57.6)	47 (39.8)	3 (2.6)	183 (77.5)	53 (22.5)
HIV-1 negative	52 (53.1)	43 (43.9)	3 (3)	147 (75)	49 (25)

sites for the following transcription factors: signal transducer and activator of transcription 6.01 (STAT6.01), with similarity of 0.94, and human and murine ETS1 factors 1.02 (ETV1.02), with similarity of 0.961. Binding sites for the transcription factors EVI-1 myeloid-transforming protein and the NF1.01 were also excluded.

Molecular analysis of Langerin CRD-encoding region - Three mutations in the *Langerin* gene region that encodes the CRD of the protein were observed: p.K313I (c.938T>A), c.941C>T (g.4728C>T) and c.983C>T (g.4770C>T). The genotypic and allelic frequencies of the analysed sites are shown in Table. Except for the p.K313I SNP, which is characterised by a thymine to adenine substitution at the -4724 gene position and results in a change from the amino acid lysine to the amino acid isoleucine (p.K313I mutation) at the -313 position of Langerin protein, the mutations did not result in amino acid changes. All of the mutations were considered as polymorphic sites in both studied groups.

Physicochemical analysis revealed that the p.K313I polymorphism enhanced the hydropathy (Kyte & Doolittle 1982) and transmembrane profiles (Argos et al. 1982) and reduced the hydrophilicity (Hopp & Woods 1983), accessibility (Janin 1979) and antigenicity profiles (Hopp & Woods 1983) (Figure). Analysis of the post-translational modification sites showed that all modification occurred at wild-type and mutated alleles and that the same potential protein domains were found in the wild-type and mutated sequences. This mutation is localised to the end of the Langerin CRD and did not alter the predicted protein secondary structure after submission of the wild-type and mutated sequences to the SWISS-MODEL server.

The genotypic and allelic frequencies of the analysed sites did not differ significantly from those frequencies expected under Hardy-Weinberg equilibrium (p.K313I: p = 0.9023, χ^2 = 1.0486; c.941C>T: p = 0.6245, χ^2 = 2.6133; c.983C>T: p = 0.4362, χ^2 = 3.7827). A strong linkage disequilibrium was observed when HIV-1-infected women and HIV-1-uninfected individuals were compared between the c.941C>T and the c.983C>T polymorphisms (p = 0.010388, χ^2 = 13.19).

DISCUSSION

Sexual transmission of HIV-1 is the most common mode of infection worldwide. The stratified epithelium of the genital mucosa contains only one specific DC subset. The first DC subset to encounter HIV-1 during sexual transmission within intact genital epithelial tissues is LCs. These cells' extensive dendrites form a barrier against pathogens. In the case of HIV-1, the binding and degradation of the virus by Langerin and Birbeck granules, respectively, prevents HIV-1 infection and posterior transmission to CD4 T cells. HIV-1 gp120 recognition by Langerin depends on the presence of the CRD, which binds the gp120 mannose-rich region, allowing HIV-1 entry (Tada et al. 2006, de Witte et al. 2007a, b, Hirbod et al. 2009, van der Vlist & Geijtenbeek 2010).

The co-expression of CD4 and CCR5 by LCs enables HIV-1 infection at high VLs or after Langerin inactivation, as HIV-1 entry into LCs may occur by the classical

pathway of infection: via CD4/CCR5. In this context, Langerin genetic variations could interfere in the expression of the protein and/or in its sugar binding activity, affecting HIV-1 infection susceptibility and acquired immune deficiency syndrome progression (Ward et al. 2006, de Witte et al. 2007a, b, Schwartz 2007, Feinberg et al. 2010, van der Vlist & Geijtenbeek 2010). Although the role of Langerin in HIV-1 infection is of high importance, there are no studies associating mutations in the *Langerin* gene with the susceptibility to viral infection. This study is the first to analyse mutations in the promoter and CRD-encoding regions of the *Langerin* gene in HIV-1-infected patients. The samples from individuals infected with HIV-1 that were analysed in this study originated in the city of Feira de Santana. This city is located at one of the major road junctions in the Northeast Region of the country, connecting this area with the Central-West, the South and the North Regions.

Many studies have demonstrated an association between polymorphisms in promoter regions of human genes and disease manifestation. It has been suggested that SNPs in the promoter regions of human genes could influence protein expression in the cell, leading to high
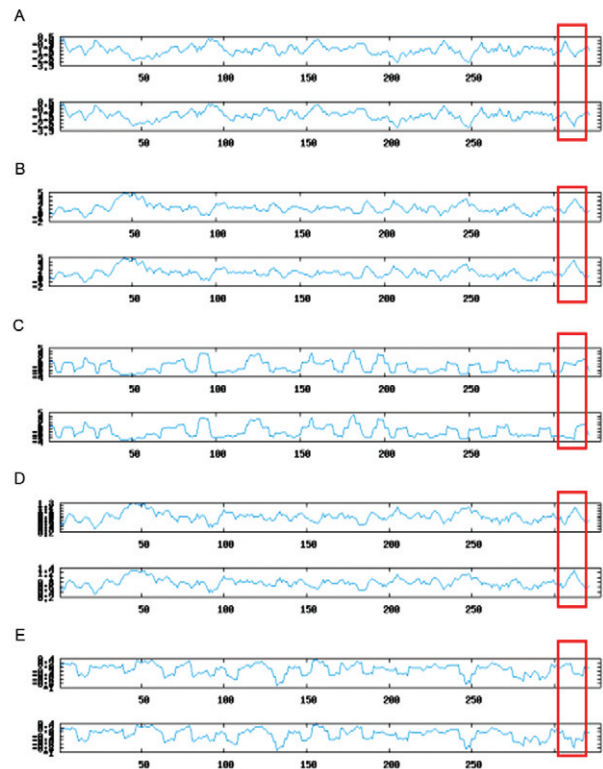
Physicochemical analysis of wild and mutated alleles of p.K313I polymorphism. A: hydrophilicity for wild allele and hydrophilicity for mutated allele, respectively; B: hydropathy for wild allele and hydropathy for mutated allele, respectively; C: accessibility for wild allele and accessibility for mutated allele, respectively; D: transmembranous helices for wild allele and transmembranous helices for mutated allele, respectively; E: antigenicity for wild allele and antigenicity for mutated allele, respectively.

or low levels of protein expression (Sabouri et al. 2004). These different protein expression phenotypes could be related to its functionality and affect its activity. Regarding HIV-1 infection, polymorphisms in the promoter region of the human *Langerin* gene could interfere with virus entry through Langerin and HIV-1 susceptibility. For this reason, the presence of polymorphisms in the *Langerin* promoter region was investigated and three single nucleotide substitutions were observed in this region: -577T>C, -517T>C and -160T>C. According to the transcription factor binding site analyses performed using the MatInspector tool, all of these SNPs created new binding sites for transcription factors.

The -577T>C mutation, for example, created a binding site for the transcription factor HOXB9.01, which encodes a nuclear protein that functions as a sequence-specific transcription factor involved in cell proliferation and differentiation (Bürglin 2005). Regarding the -517T>C polymorphism, according to the MatInspector analyses, this polymorphism is localised to the AIRE 01 binding site, which is expressed in peripheral monocytes and differentiated DC lineages (Kogowa et al. 2002). This SNP reduced the similarity of the core sequence from 0.857-0.774, although this reduction was not expressive. The SNP also created a new binding site for the transcription factor NFAT5, with a similarity of 1.0. The NFAT family of transcription factors plays a central role in inducible gene transcription during the immune response (Rinne et al. 2009). Finally, -160T>C SNP created binding sites for the transcription factors STAT6.01 and ETV1.02, with similarities of 0.94 and 0.961, respectively. STAT proteins are involved in the development and function of the immune system and play a role in maintaining immune tolerance and tumour surveillance (Vinkemeier et al. 1998). Regarding the transcription factor ETV1.02, it belongs to the ETS family, which is involved in a wide variety of functions, including the regulation of cellular differentiation, cell migration and proliferation, apoptosis and angiogenesis (Lee et al. 2005).

MatInspector is a software tool that uses a large library of matrix descriptions for transcription factor binding sites. It is possible that the creation of new binding sites does not truly affect LC functionality and that the transcription factors that bind to the new binding sites created by the mutations do not occur in LCs. There are no functional studies based on the *Langerin* gene promoter that could confirm or refute the results of these MatInspector analyses. It would be interesting to investigate the functional impact of these mutations, as the creation of new binding sites for transcription factors suggested by MatInspector analyses cannot be confirmed due to a lack of information about the transcription factors' existence in LCs.

Several of the SNPs found in the *Langerin* gene have been studied in different populations and can alter the amino acid sequence of the Langerin protein, resulting in changes in the protein's functionality. The p.A300P and p.N288D polymorphisms, for example, decrease Langerin's sugar-binding activity by decreasing the stability of the CRD and mediating a local effect on the structure around the sugar-binding site, respectively.

In addition, the W264R mutation completely abolishes Langerin's sugar-binding activity, preventing Langerin from binding to microorganisms, such as HIV-1 (Verdijk et al. 2005, Ward et al. 2006). This study demonstrated the existence of three polymorphisms in the CRD-encoding region: two synonymous (c.941C>T - rs2080390 and c.983C>T - rs13421115) and one nonsynonymous (p.K313I - rs57302492).

The p.K313I polymorphism resulted in a substitution of a lysine by an isoleucine amino acid at the end of the CRD of the protein (position 313). The enhancement of the hydropathy and transmembrane profiles and the reduction in the hydrophilicity, accessibility and antigenicity profiles at this polymorphic site, noted in the physicochemical analysis, could reduce the access of HIV-1 gp120 to this CRD region and affect HIV-1 entry through Langerin. However, this mutation did not alter the predicted protein secondary structure.

In summary, the results presented in this study reveal several polymorphisms in the promoter and CRD-encoding regions of the human *Langerin* gene (-577T>C, -517T>C, -160T>C, p.K313I, c.941C>T and c.983C>T) in both analysed groups. As this is the first study to investigate *Langerin* gene mutations in individuals infected with HIV-1, more extensive human genetic variation analyses are necessary to clarify Langerin's major role in HIV-1 infection.

REFERENCES

Ahmed Z, Czubala M, Blanchet F, Piguet V 2013. HIV impairment of immune responses in dendritic cells. *Adv Exp Med Biol 763*: 201-238.

Argos P, Rao JK, Hargrave PA 1982. Structural prediction of membrane-bound proteins. *Eur J Biochem 15*: 565-575.

Arnold K, Bordoli L, Kopp J, Schwede T 2006. The SWISS-MODEL Workspace: a web-based environment for protein structure homology modelling. *Bioinformatics 22*: 195-201.

Ayres M, Ayres Jr M, Ayres DL, Santos AS 2005. BioEstat v.5.0, Sociedade Civil Mamirauá, MCT-CNPq, Belém.

Bürglin TR 2005. Homeodomain proteins. In RA Meyers, *Encyclopedia of molecular cell biology and molecular medicine*, Wiley-VCH Verlag GmbH and Co, Weinheim, p. 179-222.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics 21*: 2933-2942.

Combet C, Blanchet C, Geourjon C, Deléage G 2000. NPS@: network protein sequence analysis. *Trends Biochem Sci 25*: 147-150.

de Witte L, Nabatov A, Geijtenbeek TB 2007a. Distinct roles for DC-SIGN+-dendritic cells and Langerhans cells in HIV-1 transmission. *Trends Mol Med 14*: 12-19.

de Witte L, Nabatov A, Pion M, Fluitsma D, de Jong M, Gruijl T, Piguet V, van Kooyk, Geijtenbeek TB 2007b. Langerin is a natural barrier to HIV-1 transmission by Langerhans cells. *Nat Med 13*: 367-371.

Feinberg H, Powlesland A, Taylor M, Weis W 2010. Trimeric structure of Langerin. *J Biol Chem 285*: 13285-13293.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer

ELL, Eddy SR, Bateman A 2010. The Pfam protein families database. *Nucl Acids Res 38*: 211-222.

Hall TA 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser 41*: 95-98.

Hirbod T, Kaldensjo T, Broliden K 2011. *In situ* distribution of HIV-binding CCR5 and C-type lectin receptors in the human endocervical mucosa. *PLoS ONE 6*: e25551.

Hirbod T, Kaldensjo T, Lopalco L, Klareskog E, Andersson S, Uberti-Foppa C, Ferrari D, Manghi M, Andersson J, Loré K, Broliden K 2009. Abundant and superficial expression of C-type lectin receptors in ecotocervix of women at risk of HIV infection. *J Acquir Immune Defic Syndr 51*: 239-247.

Hopp TP, Woods KR 1983. Physico-chemical profiles: a computer program for predicting protein antigenic determinants. *Mol Immunol 20*: 483-489.

Janin J 1979. Surface and inside volumes in globular proteins. *Nature 277*: 491-492.

Kogowa K, Nagafuchi S, Katsuta H, Kudoh J, Tamiya S, Sakai Y, Shimizu N, Harada M 2002. Expression of AIRE gene in peripheral monocyte/dendritic cell lineage. *Immunol Lett 80*: 195-198.

Kyte J, Doolittle R 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol 157*: 105-132.

Lee GM, Donaldson LW, Pufall MA, Kang HS, Pot I, Graves BJ, McIntosh LP 2005. The structural and dynamic basis of ETS-1 DNA binding autoinhibition. *J Biol Chem 280*: 7088-7099.

McGreal EP, Martinez-Pomares L, Gordon S 2004. Divergent roles for C-type lectins expressed by cells of the innate immune system. *Mol Immunol 41*: 1109-1121.

Mesman AW, Geijtenbeek TB 2012. Pattern recognition receptors in HIV transmission. *Front Immunol 3*: 1-9.

Nicholas KB, Nicholas Jr HB, Deerfield DW II 1997. GeneDoc: analysis and visualization of genetic variation. *EMBNEW.NEWS 4*: 14.

Raymon M, Rousset F 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered 86*: 248-249.

Rinne A, Banach K, Blatter LA 2009. Regulation of nuclear factor of activated T cells (NFAT) in vascular endothelial cells. *J Mol Cell Cardiol 47*: 400-410.

Sabouri AH, Saito M, Lloyd AL, Vine AM, Witkover AW, Furukawa Y, Izumo S, Arimura K, Marshall SEF, Usuku K, Bangham CRM, Osame M 2004. Polymorphism in the interleukin-10 promoter affects both provirus load and the risk of human T lymphotropic virus type I-associated myelopathy/tropical spastic paraparesis. *J Infect Dis 190*: 1279-1285.

Schwartz O 2007. Langerhans cells lap up HIV-1. *Nat Med 13*: 245-246.

Stambach N, Taylor M 2003. Characterization of carbohydrate recognition by Langerin, a C-type lectin of Langerhans cells. *Glycobiology 13*: 401-410.

Tada Y, Riedl E, Lowenthal M, Liotta L, Briner D, Crouch E, Udey M 2006. Identification and characterization of endogenous Langerin ligands in murine extracellular matrix. *J Invest Dermatol 126*: 1549-1558.

Valladeau J, Clair-Moninot V, Dezutter-Dambuyant C, Pin J, Kissenpfennig A, Mattéi MG, Yahia SA, Bates EEM, Malissen B, Koch F, Fossiez F, Romani N, Lebecque S, Saeland S 2002. Identification of mouse Langerin/CD207 in Langerhans cells and some dendritic cells of lymphoid tissues. *J Immunol 168*: 782-792.

van den Berg LM, Geijtenbeek TB 2013. Antiviral immune responses by human Langerhans cells and dendritic cells in HIV-1 infection. *Adv Exp Med Biol 762*: 45-70.

van der Vlist M, Geijtenbeek T 2010. Langerin functions as an antiviral receptor on Langerhans cells. *Immunol Cell Biol 88*: 410-415.

Verdijk P, Dijkman R, Plasmeijer EI, Mulder AA, Zoutman WH, Mommaas AM, Tensen CP 2005. A lack of birbeck granules in Langerhans cells is associated with a naturally occurring point mutation in the human *Langerin* gene. *J Invest Dermatol 124*: 714-717.

Vinkemeier U, Moarefi I, Darnell JE, Kuriyan J 1998. Structure of the amino-terminal protein interaction domain of STAT-4. *Science 279*: 1048-1052.

Ward E, Stambach N, Drickamer K, Taylor M 2006. Polymorphisms in human Langerin affect stability and sugar binding activity. *J Biol Chem 281*: 15450-15456.