Eur J Med Res (2007) 12: 453-462

GENOTYPIC CORECEPTOR ANALYSIS

S. Sierra[1], R. Kaiser[1], A. Thielen[2] and T. Lengauer[2]

[1]Institute of Virology, University of Cologne, Germany, [2]Max-Planck-Institute for Informatics, Saarbrücken, Germany

Abstract
HIV infects target cells by binding of its envelope gp120 protein to CD4 and a coreceptor on the cell surface. *In vivo*, the different HIV-strains use either CCR5 or CXCR4 as coreceptor. CCR5-using strains are named R5 viruses, while CXCR4-using strains are named X4. X4 viruses usually occur in the later stages. Coreceptor usage is a marker for disease progression. Additionally interest on coreceptors continually raises as a consequence of the development of a new class of antiretroviral drugs, namely the coreceptor antagonists or blockers. These specific drugs block the CCR5 or the CXCR4 coreceptors. So far, the CXCR4 blockers are not allowed to be used in the clinical practice due to their severe side effects. On the other hand, CCR5 blockers are currently in clinical practice, although they can only be administered after a baseline determination of the coreceptor usage of the predominant viral strain. Most of the coreceptor analyses in clinical cohorts have been performed with commercially available phenotypic assays. As for resistance testing of NRTIs, NNRTIs and PIs, efforts have also been made to predict the coreceptor usage from the genotype of the viruses. Different rules have been published based on the amino acid sequence of the Env-V3 region of HIV-gp120, which is known to be the major determinant of coreceptor usage. Among these, the most widely used is the 11/25 rule. Recently, bioinformatics driven prediction systems have been developed. Three of the interpretation systems are freely available via internet: WetCat, WebPSSM, geno2pheno$_{[coreceptor]}$. All three systems focus on the Env-V3 region and take the amino acid sequence only into account. They learn from phenotypic and corresponding genotypic data. So far, two cohorts have been analyzed with such a genotypic approach and provided frequencies of R5 virus strains that are within the range of those reported with phenotypic assays. For one of the systems, geno2pheno$_{[coreceptor]}$, additional clinical data (e.g. CD4$^+$T-cell counts) or structural information can be used to improve the prediction. Such genotypic systems provide the possibility for rapid screening of patients who may be administered with CCR5 blockers like the recently licensed Maraviroc.

INTRODUCTION

HIV infection begins with the attachment of the virions to the cell surface mediated by an interaction between the extracellular domain of the viral envelope protein (Env or gp120), the CD4 receptor and a cellular chemokine receptor. *In vivo* only the chemokine receptors CCR5 and CXCR4 play a role for HIV-1 and HIV-2 infection [1, 2]. The coreceptor usage (also named tropism) depends on the HIV strain and allows for the classification in i) R5 for those viral strains using CCR5, ii) X4 for those using CXCR4 and iii) R5X4 for the strains able to use both coreceptors.

The coreceptor usage is associated with disease progression in untreated patients [3, 4] and response to antiretroviral therapy (HAART) [4-10]. In addition, the recent developments of antiretrovirals which specifically block the CCR5 coreceptor (coreceptor antagonists or blockers) require the use of prediction tools for coreceptor usage. Most studies have shown that *in vivo* R5 tropism is the prevalent phenotype in the early stages of the HIV infection, irrespective of the transmission route and the predominant viral tropism present in the donor [11-14]. CXCR4-using strains evolve later in 10-20% of the therapy naïve patients and 30-60% of the severely immunodeficient HAART-experienced patients [3, 4, 15-24]. As the reasons for this non-systematic switch are still unknown and CCR5 blockers only work if the patient displays a prevalent R5 strain, a baseline determination of the predominant tropism of the circulating virus prior to treatment with these drugs is advisable. Therefore, for clinical purposes, the viral strains are classified as X4-users (classical X4 and R5X4) or R5-users, only the latter being candidates for a coreceptor blocker therapy.

IMPORTANCE OF ENV-V3 FOR THE PREDICTION OF CORECEPTOR USAGE AND RESISTANCE TO CCR5 BLOCKERS

The Env-CD4 interaction triggers a subsequent conformational change in the viral Env, which rearranges its core and exposes the coreceptor binding site, a discontinuous epitope comprising the third hypervariable loop of the viral envelope gp120 protein (Env-V3), the β-19 strand and the bridging sheet [25-29]. The charge and structure of the Env-V3 is the major determinant for the specificity for the CCR5 or CXCR4 coreceptor while the bridging sheet provides the main determinants the interaction with both coreceptors [27]. In fact, some studies showed that replacement of the Env-V3 loop from an R5 into an X4 virus can lead to a coreceptor usage switch [29, 30].

Env-V3, the bridging sheet and both coreceptors are

charged due to the presence of basic amino acids (K or R), acidic amino acids (D or E) and post-transcriptional modifications (mainly N- or O-glycosilations or tyrosine sulfation). Therefore, electrostatic interactions are greatly implicated in the efficacy and specificity of coreceptor binding [31-35]. Indeed, mutations affecting the Env-V3 charge or the overall net charge of this region correlate with coreceptor selectivity [36-49]. R5 isolates usually show lower net Env-V3 charge than X4 variants [41-48, 50-53], which fits the observation that CCR5 has a higher positive net charge than CXCR4.

On the other hand, some mutations not affecting the charge of the Env-V3 region were reported to be important for coreceptor selectivity [36-40, 42, 54-57] indicating that structural constrictions are also relevant for the gp120-coreceptor interaction.

Several studies showed that resistance to the CCR5 inhibitors AD101 (or SCH-35081), SCH-D (Vicriviroc) or Maraviroc (MVC) involves point mutations in the Env-V3 region [58-61], similar to the appearance of antiretroviral resistance against HIV RT or PR, where specific amino acid substitutions in the target genes are required. In the case of MVC resistance, the alterations within Env-V3 permit the virus to recognise the altered structure of the drug-occupied CCR5 coreceptor [61, 62].

Coreceptor Usage Prediction Based on the Bioinformatical Analysis of the Viral Env-V3 Sequence

The coreceptor which an HIV-1 strain can use for infection of a cell depends on the viral envelope-protein gp120. The ability of this protein to bind to a specific coreceptor is the result of conformational properties and binding-energies between gp120 and the chemokine receptors. The orientation and charge of the involved residues is defined by the viral genotype. Thus, in the end, all the information needed for determination of coreceptor usage of a virus strain is encoded in its genome. Hence, genotypic methods should in principle be able to determine the phenotype from the genotype.

Certainly, current biophysical models for prediction of protein conformation and binding-energies are not accurate enough for predicting tropism directly from sequence. Therefore, genotypic methods predict coreceptor usage by correlating experimentally validated genotype-phenotype pairs. This idea has already been applied successfully in the realm of HIV drug resistance testing. There, sophisticated interpretation algorithms are widely used in support of treatment with antiretrovirals by predicting response to drugs as well as outcome of combination therapies [63].

Such an approach has several benefits. While phenotypic assays are based on cell culture experiments and are therefore relatively expensive and have relatively slow turnaround, genotypic approaches are much easier to perform and consequently cheaper, faster and easier to standardize. In addition, surrogate markers such as the patient's $CD4^+$ T-cell counts can be incorporated into these methods and improve their quality. Phenotypic assays cannot make use of this information and simulate the state of the patient's immune surveillance.

However, a problem of genotypic approaches is that the interpretation of sequence data is challenging. Statistical learning methods applied for this task are rarely descriptive and it is often hard to understand why a method classifies a sample to a certain phenotype. Moreover, although genotypic methods are continually improving and yield very good results on clonal data, their predictions are still not as accurate as phenotypic assays, especially on clinical samples.

Another benefit of genotypic methods is that they can simulate and analyze the possible evolution of the virus. Similar to the commonly used concept of the "genetic barrier" in drug resistance, this can be used to compute the number of mutations a given virus sequence has to accumulate for switching its coreceptor.

Like phenotypic approaches, genotypic methods have the limitation that they are not able to detect if a minor viral population of X4-users is "hidden" in a prevalent R5-user population.

Coreceptor Usage Prediction Tools

Several methods have been developed for prediction of coreceptor usage. Since Env-V3 is known to be the major determinant of coreceptor tropism and due to the lack of sufficient data for other regions, all current methods focus on it.

A simple but popular approach is the classical *11/25 rule* [44, 45, 64, 65]. It predicts a virus to be X4-user if basic amino acids (arginine or lysine) are present at positions 11 or 25 of Env-V3 and R5-user if no basic amino acids are to be found within these positions. This rule is quite accurate for R5-users but misclassifies many X4-users [66].

The lack of high sensitivity to detect X4-usage has led to the development of more complex methods aiming at improving the true positive rate. These methods combine the amino acid composition with the overall net charge of Env-V3 [41, 47, 66].

Further advancements could be achieved by using more sophisticated statistical learning methods. Among these, different decision tree classifiers [67, 68] and artificial neural networks [69] were the first ones reported. Both approaches can handle larger parameters sets than simple motif methods and usually employ the amino acid composition and the charge of the Env-V3 loop as explanatory variables. Alternatively suggested methods include support vector machines (SVMs) [56, 68], position specific scoring matrices (PSSMs) [66], and mixtures of localized rules [56].

Recent progress has been made by incorporating the structure of the V3 loop into prediction engines. Sander et al. developed a distance-based descriptor of the spatial arrangement of physiochemical properties that was taken as input to a SVM [57]. In comparison with a purely sequence-based SVM, they could show that the inclusion of structural features improved coreceptor usage significantly.

Although a number of prediction methods have been developed so far, only a few of them are available as web-services: *WetCat*, *WebPSSM*, and *geno2pheno[coreceptor]*. To use these tools, nucleotide or amino acid sequences of the Env-V3 region in FASTA for-

mat are sent to the server as input for predictions.

WetCat is a web-service developed and maintained at the University of California, San Diego. It implements the charge rule, three different decision trees, and support vector machines [68]. The sequences have to be manually translated into amino acids, excised so that exclusively the Env-V3 region is included, and aligned to a consensus sequence described on the website, which is very time-consuming. Especially the manual alignment to the reference is very error prone. As a result of a prediction run, a page is displayed showing the predicted phenotype for each sequence.

A benefit of WetCat is that it allows batch predictions in a single run. However, the very restricted format is not user friendly and additional tools for sequence preparation would be helpful. Furthermore, a limitation of this tool is that the different prediction models are still trained on the original dataset containing only 271 sequences. An update of the models with new data would probably improve this tool.

WebPSSM

The WebPSSM server predicts coreceptor usage from Env-V3 sequences given in amino acidic FASTA format with position specific scoring matrices developed in [66]. Alignment of the samples with the consensus sequence is automatically done by the server which uses the Needleman-Wunsch algorithm and an amino acid distance matrix to align them before scoring.

The user can choose among three different prediction matrices: two subtype B and one subtype C matrix. One of the subtype B models is trained on sequences of known coreceptor phenotype, whereas the other one is generated from sequences with known syncytium-inducing phenotype on the MT2 cell line. For subtype C, only a matrix trained on sequences of known syncytium-inducing phenotype is provided.

WebPSSM not only predicts whether a virus can use the CXCR4 coreceptor or not, but also gives additional information of the predictions. It displays a quantitative value, the prediction score, as well as percentile scores describing the certainty and trustiness of the calculated predictions. Furthermore, the amino acids at positions 11 and 25 are shown in order to facilitate a comparison with the 11/25 rule. Finally, the number of positively charged amino acids in the Env-V3 loop and its net charge are displayed together with a measure reflecting the probability of being an Env-V3 sequence.

An advantage of the server is that it accepts as many as 100 Env-V3 sequences and aligns them autonomously. In addition, the possibility of downloading results in tab-delimited format allows for easy post-processing.

A deficit of the system is that only subtype B and C matrices are offered. In principle, these matrices can also be used to predict samples from other subtypes but such predictions should be treated with extreme scepticism. Furthermore, it is not clear on which dataset the matrices have been trained on and if they have been updated since their initial generation. Because statistical learning methods improve with larger amounts of data, this information would be quite useful.

Geno2pheno$_{[coreceptor]}$

Geno2pheno$_{[coreceptor]}$ predictions are based on support vector machines [70]. Predictions can be performed from fasta-formatted nucleotide or amino acid sequences containing the Env-V3 region. The sequences can be pasted into a text field or uploaded from a file.

If required, the server allows configuration to different users requirements by varying the settings for significance levels.

The version 2.0 of geno2pheno$_{[coreceptor]}$ can make use of additional clinical parameters. If such markers are provided, the server returns two predictions: i) a prediction by the standard model trained on clonal data ii) a prediction by a model trained on data including clinical markers from 1000 therapy-naïve patients.

Geno2pheno$_{[coreceptor]}$ generates an output page divided into four parts. The first two parts display general overview information and the provided clinical markers. In the third section, an alignment of the query sequence to the standard reference HXB2 is provided. The N-glycosylation motif and positions 11 and 25, all known to be significant for CXCR4-use are highlighted and coloured. The last part contains the predicted phenotype and, similar to WebPSSM, a p-value assessing the confidence of a prediction. For better understanding, the prediction field is shown with a green background in case of a predicted R5-virus, otherwise in red. If additional clinical parameters have been sent to the server, a second similar row with the results of the clinical model is displayed.

Geno2pheno$_{[coreceptor]}$ is not restricted to a specific subtype. Although most of its training sequences are from subtype B viruses, other subtypes have been included in the training process, as well.

Quality Assessment

The performance of the different web-services is hard to determine and to compare for several reasons. First of all, one cannot directly compare the performances in the respective publications because they were calculated on different datasets and different measures were used for evaluation. Second, the different authors had different demands on the quality of the data. For example, in the validation of the WetCat-models, all sequences shorter than 34 or longer than 36 residues were simply discarded. Of course, this has a high influence on the performance. However, the main reason for the difficulty in comparing the servers is the lack of data. A fair evaluation must ideally be based on an independent dataset with sequences not used for training of any method. The problem at this point is that all three different web-services have been trained on different subsets of the Los Alamos Sequence Database. Therefore, one cannot simply evaluate the predictions on all sequences in the database. Other resources with genotype-phenotype information not included in the Los Alamos database are sparse and most of them include only a small number of samples. Thus, statistically significant results cannot be inferred from these experiments.

An attempt for a fair evaluation has been made for

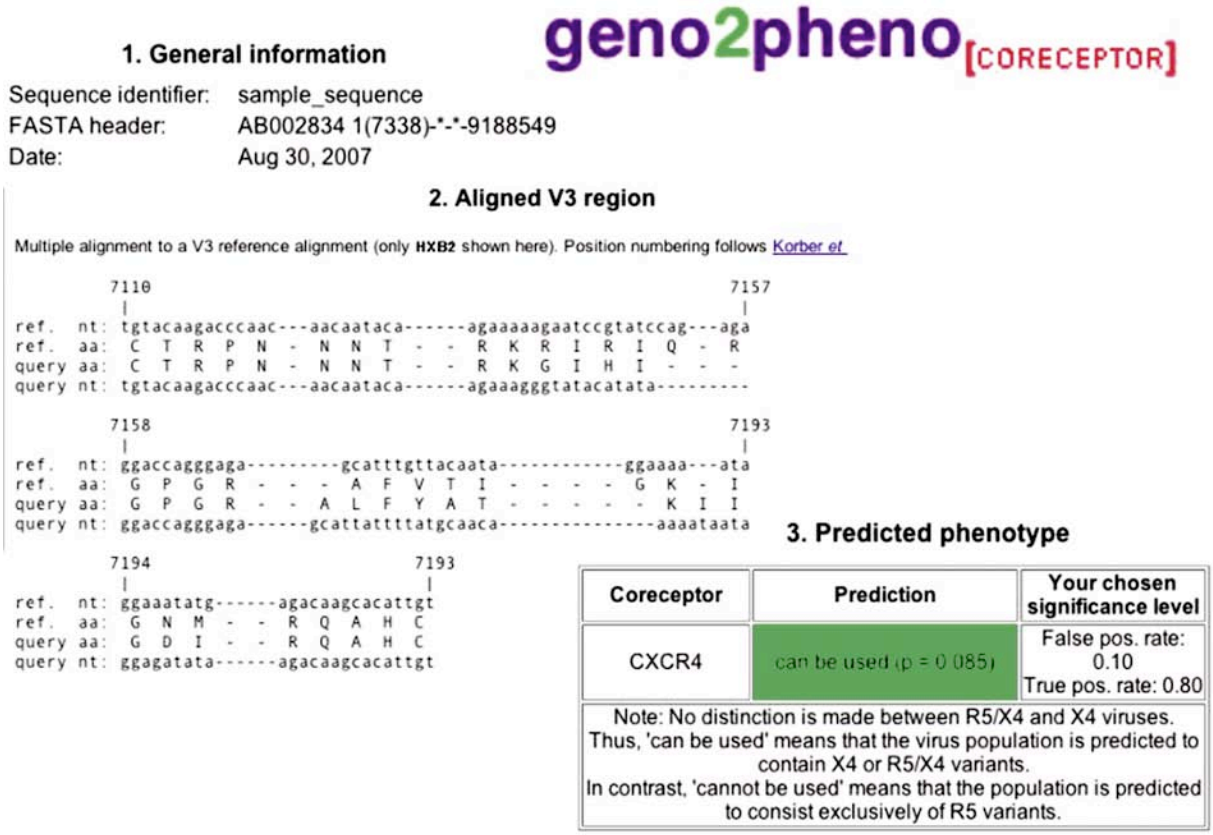

Fig. 1. A Geno2pheno$_{[coreceptor]}$ output page. The output is divided in three sections. Section 1: sample general information. Section 2: alignement of the sample sequence with the consensus HXB2. Section 3: CXCR4 usage prediction. In this example, CXCR4 usage is predicted and therefore CCR5 coreceptor blockers should not be administrated.

the development of geno2pheno$_{[coreceptor]}$. For this evaluation, all sequences with experimentally determined coreceptor-usage were downloaded from the Los Alamos Sequence Database. Different methods, including the 11/25 rule, decision trees, position specific scoring matrices, and support vector machines were implemented according to the publications in which they were proposed. In ten replicates of 10-fold cross-validation experiments, each method was evaluated on this dataset. The 11/25 rule yielded a sensitivity of 59.5% in detecting X4-using variants and a mean specificity of 92.5%. Decision trees, neural networks, mixtures of localized rules or simple modifications of the charge rule led to minor improvements (sensitivity of about 62.5%) at the corresponding specificity. Both PSSMs and SVMs, outperformed all other methods and significantly improved sensitivity by 12.4 and 16.9 percentage points, respectively. Since the SVMs showed slightly better results than PSSMs, they have been chosen as the method of choice in geno2pheno$_{[coreceptor]}$. It has to be emphasized that this evaluation only compared the different methods used in the web servers and not the different web-services themselves, because these have been trained on different datasets. The real performance of the systems is mainly based on the number and the quality of their training sets.

The results of this evaluation are based on clonal data. The reliability of all methods decreases when using clinically derived samples. These are generated with population-based "bulk" sequencing technologies and contain mixtures of co-existing viral variants.

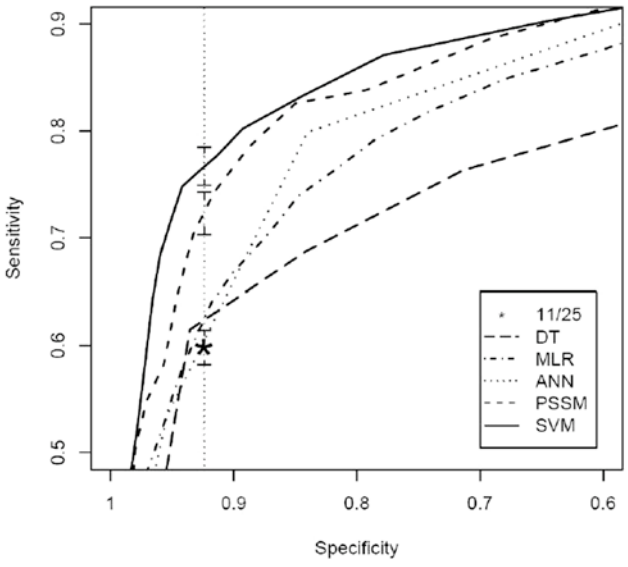

Fig. 2. Predictive performance of the 11/25 rule and five statistical learning methods (DT: decision trees, MLR: mixtures of localized rules, ANN: artificial neural networks, PSSM: Position Specific Scoring Matrices, SVM: Support Vector Machines), assessed on clonal data (from [70]).

The fact, that these variants cannot be properly discriminated from each other makes predictions more problematic. In a recent study on 952 plasma samples from antiretroviral-naïve patients, the three web-services have been evaluated [71]. The performances of all tested methods and the 11/25-rule decreased significantly. While the 11/25-rule's specificity of 93.4% was similar to results obtained on clonal data, sensitivity dropped to 30.5%. The performance of the SVM-models of WetCat were even worse as they only reached a sensitivity of 22% with a specificity of 90%. In comparison, WebPSSM and the SVM-models of geno2pheno performed much better but their performances also decreased substantially. They showed sensitivity values of about 50% at the 90% specificity-level [71].

Other Factors of Improving Prediction Efficacy

Structural Information

Current genotypic methods for tropism prediction use only sequence information. They do not consider the three-dimensional structure of the Env-V3 loop, although its conformational properties have been discussed in several studies [27, 72-74]. The deeper understanding of the interactions between the viral envelope protein and the cellular coreceptors would reduce the need for sequence data and simplify the design of new drugs. The reason for the current lack of structural predictors is that only a few NMR-structures existed until recently. Lately, the first crystal structure was published [75] and based on this, the first structural bioinformatics method for prediction of coreceptor usage was developed [57]. The work showed that the inclusion of structural information can improve predictive performance of existing methods.

Although these results are very encouraging, a lot of space for improvements still remains. One limitation of the published model is that the backbone of the loop is held fixed. Relaxing this restriction and incorporating coreceptor structures in the analysis are only two points of possible improvements. These points can be addressed as soon as more reliable crystal structures are published.

Other gp120 Regions

Both CCR5 and CXCR4 coreceptors interact with the same region of the viral Env. This region encompasses not only the Env-V3 loop but also the β-19 strand and the bridging sheet, a four-stranded antiparallel b-sheet encompassing β-2 and β-3 strands from the V1/V2 stem and the b-20 and b-21 strands within the C4 region [25-29]. Indeed, mutations in these regions have been shown to directly influence coreceptor usage [76-83].

Currently available methods do not consider other gp120-regions mainly because of lacking data. Since the Env-V3 sequence is assumed to be the major determinant of coreceptor usage and since it is more complicated to amplify and sequence larger fragments of gp120, most existing data are restricted to Env-V3. However, as new more powerful sequencing techniques are developed and the involvement of other Env-regions in coreceptor-tropism and resistance to coreceptor-antagonist becomes more obvious, more and more data containing larger regions of gp120 are being expected. Along with these, the prediction methods will probably be improved and adjusted.

Patient Clinical Factors

The detection of X4 using strains correlates with a decrease in the $CD4^+$-T-cell counts under 200 cells/ml [4, 21, 84]. It is not clear if the appearance of X4 strains is a consequence of immune system exhaustion and elimination of X4 emergence inhibitors, or *vice versa*, the appearance of X4 viruses is the origin of massive $CD4^+$-T-cell depletion and is the cause for a faster disease progression. Regardless of which direction holds, different patient clinical factors can be incorporated into the prediction methods, and support them, as shown in a recent study [70]. Plasmatic viral sequences, different clinical markers (including plasma viral load, $CD4^+$- and $CD8^+$-T-cell counts, and the percentage of $CD4^+$-T-cells at time of sampling), and the patients' genotype for the 32-base pair deletion resulting in non-functional CCR5-coreceptors from 979

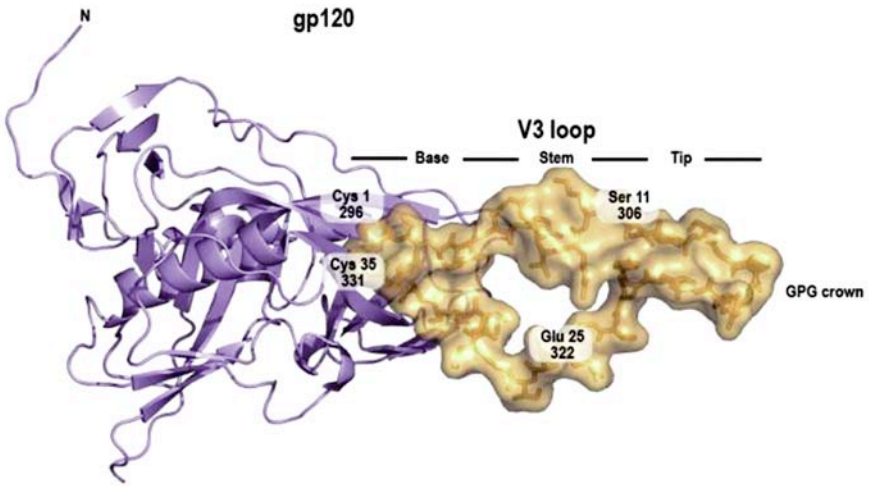

Fig. 3. Three dimensional structure of the HIV-1 Env protein . Structural informations improved the power of the predictions significantly [57].

antiretroviral-naïve patients were analysed using a support vector machine. Compared with a purely sequence-based SVM, the model including the clinical markers performed significantly better.

However, since the observations are from therapy-naïve patients, future studies have to check whether these markers can also be used for therapy-experienced patients. In addition, other clinical markers that may improve predictions (e.g. therapy history) should also be analyzed.

REFERENCES

1. Clapham PR, A McKnight. Cell surface receptors, virus entry and tropism of primate lentiviruses. J Gen Virol 2002; 83: 1809-1829.
2. Moore JP, SG Kitchen, P Pugach, JA Zack. The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. AIDS Res Hum Retroviruses 2004; 20: 111-126.
3. Tersmette M, RE de Goede, BJ Al, IN Winkel, RA Gruters, HT Cuypers, HG Huisman, F Miedema. Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. J Virol 1988; 62: 2026-2032.
4. Koot M, IP Keet, AH Vos, RE de Goede, MT Roos, RA Coutinho, F Miedema, PT Schellekens, M Tersmette. Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4$^+$ cell depletion and progression to AIDS. Ann Intern Med 1993; 118: 681-688.
5. Brumme ZL, WW Dong, B Yip, B Wynhoven, NG Hoffman, R Swanstrom, MA Jensen, JI Mullins, RS Hogg, JS Montaner, PR Harrigan. Clinical and immunological impact of HIV envelope V3 sequence variation after starting initial triple antiretroviral therapy. AIDS. 2004; 18: F1-9.
6. Davey RT, Jr., DG Chaitt, GF Reed, WW Freimuth, BR Herpin, JA Metcalf, PS Eastman, J Falloon, JA Kovacs, MA Polis, RE Walker, H Masur, J Boyle, S Coleman, SR Cox, L Wathen, CL Daenzer, HC Lane. Randomized, controlled phase I/II, trial of combination therapy with delavirdine (U-90152S) and conventional nucleosides in human immunodeficiency virus type 1-infected patients. Antimicrob Agents Chemother. 1996; 40: 1657-1664.
7. Hughes MD, VA Johnson, MS Hirsch, JW Bremer, T Elbeik, A Erice, DR Kuritzkes, WA Scott, SA Spector, N Basgoz, MA Fischl, RT D'Aquila. Monitoring plasma HIV-1 RNA levels in addition to CD4$^+$ lymphocyte count improves assessment of antiretroviral therapeutic response. ACTG 241 Protocol Virology Substudy Team. Ann Intern Med. 1997; 126: 929-938.
8. Rusconi S, MP De Pasquale, F Mainini, E Bulgheroni, S Kurtagic, A Gori, M Violin, N Zanchetta, M Moroni, C Balotta, M Galli. Viral load, viral phenotype modification, zidovudine susceptibility and reverse transcriptase mutations during the first 6 months of zidovudine monotherapy in HIV-1-infected people. Antivir Ther. 1996; 1: 211-219.
9. Katzenstein TL, C Nielsen, L Bruun, LR Mathiesen, C Pedersen, J Gerstoft. Quantification of HIV-1 RNA during antiretroviral therapy: association with viral phenotype and development of resistance. Antivir Ther. 1996; 1: 246-254.
10. Vidal C, F Garcia, JM Gatell, M Leal, B Clotet, T Pumarola, JM Miro, J Mallolas, L Ruiz, A Cruceta, C Tortajada. Predictive factors influencing peak viral load drop in response to nucleoside reverse transcriptase inhibitors in antiretroviral-naive HIV-1-infected patients. J Acquir Immune Defic Syndr Hum Retrovirol. 1998; 19: 55-60.
11. van't Wout AB, NA Kootstra, GA Mulder-Kampinga, N Albrecht-van Lent, HJ Scherpbier, J Veenstra, K Boer, RA Coutinho, F Miedema, H Schuitemaker. Macrophage-tropic variants initiate human immunodeficiency virus type 1 infection after sexual, parenteral, and vertical transmission. J Clin Invest. 1994; 94: 2060-2067.
12. Wolinsky SM, CM Wike, BT Korber, C Hutto, WP Parks, LL Rosenblum, KJ Kunstman, MR Furtado, JL Munoz. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science. 1992; 255: 1134-1137.
13. Zhu T, N Wang, A Carr, DS Nam, R Moor-Jankowski, DA Cooper, DD Ho. Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. J Virol. 1996; 70: 3098-3107.
14. Scarlatti G, V Hodara, P Rossi, L Muggiasca, A Bucceri, J Albert, EM Fenyo. Transmission of human immunodeficiency virus type 1 (HIV-1) from mother to child correlates with viral phenotype. Virology 1993; 197: 624-629.
15. de Roda Husman AM, RP van Rij, H Blaak, S Broersen, H Schuitemaker. Adaptation to promiscuous usage of chemokine receptors is not a prerequisite for human immunodeficiency virus type 1 disease progression. J Infect Dis 1999; 180: 1106-1115.
16. Li S, J Juarez, M Alali, D Dwyer, R Collman, A Cunningham, HM Naif. Persistent CCR5 utilization and enhanced macrophage tropism by primary blood human immunodeficiency virus type 1 isolates from advanced stages of disease and comparison to tissue-derived isolates. J Virol 1999; 73: 9741-9755.
17. Tersmette M, RA Gruters, F de Wolf, RE de Goede, JM Lange, PT Schellekens, J Goudsmit, HG Huisman, F Miedema. Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates. J Virol 1989; 63: 2118-2125.
18. Whitcomb JM, W Huang, S Fransen, T Wrin, EE Paxinos, J Toma, M Greenberg, P Sista, T Melby, T Matthews, R Demasi, G Heilek-Snyder, N Cammack, N Hellmann, C Petropoulos. Analysis of Baseline Enfuvirtide (T20) Susceptibility and Co-receptor Tropism in Two-phase III Study Populations. in 10th Conference on Retroviruses and Opportunistic Infetions (CROI). 2003. Boston, MA, USA Abstract 557.
19. Paxinos EE, E Fransen, W Huang, T Wrin, J Whitcomb, C Petropoulos. Measuring HIV Co-Receptor Tropism using a recombinant Virus assay. in 42nd Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). 2002. San Diego, CA, USA Abstract 2040.
20. Wilkin TJ, Z Su, DR Kuritzkes, M Hughes, C Flexner, R Gross, E Coakley, W Greaves, C Godfrey, PR Skolnik, J Timpone, B Rodriguez, RM Gulick. HIV type 1 chemokine coreceptor use among antiretroviral-experienced patients screened for a clinical trial of a CCR5 inhibitor: AIDS Clinical Trial Group A5211. Clin Infect Dis 2007; 44: 591-595.
21. Brumme ZL, J Goodrich, HB Mayer, CJ Brumme, BM Henrick, B Wynhoven, JJ Asselin, PK Cheung, RS Hogg, JS Montaner, PR Harrigan. Molecular and clinical epidemiology of CXCR4-using HIV-1 in a large population of antiretroviral-naive individuals. J Infect Dis 2005; 192: 466-474.
22. Moyle GJ, A Wildfire, S Mandalia, H Mayer, J Goodrich,

J Whitcomb, BG Gazzard. Epidemiology and predictive factors for chemokine receptor use in HIV-1 infection. J Infect Dis 2005; 191: 866-872.
23. Demarest J, T Bonny, C Vavro, C Labranche, KM Kitrinos, C McDanal, S Sparks, S Chavers, S Castillo, D Elrick, D McCarty, J Whitcomb, W Huang, C Petropoulos, S Piscitelli. HIV-1 co-receptor tropism in treatment naive and experienced subjects. in 44th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). 2004. Washington, DC, USA Abstract H-1136.
24. Lehmann C, M Daumer, I Boussaad, T Sing, N Beerenwinkel, T Lengauer, N Schmeisser, C Wyen, G Fatkenheuer, R Kaiser. Stable coreceptor usage of HIV in patients with ongoing treatment failure on HAART. J Clin Virol 2006; 37: 300-304.
25. Kwong PD, R Wyatt, J Robinson, RW Sweet, J Sodroski, WA Hendrickson. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 1998; 393: 648-659.
26. Rizzuto CD, R Wyatt, N Hernandez-Ramos, Y Sun, PD Kwong, WA Hendrickson, J Sodroski. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. Science 1998; 280: 1949-1953.
27. Lusso P. HIV and the chemokine system: 10 years later. Embo J. 2006; 25: 447-456.
28. Myszka DG, RW Sweet, P Hensley, M Brigham-Burke, PD Kwong, WA Hendrickson, R Wyatt, J Sodroski, ML Doyle. Energetics of the HIV gp120-CD4 binding reaction. Proc Natl Acad Sci U S A 2000; 97: 9026-9031.
29. Biscone MJ, JL Miamidian, JM Muchiri, SS Baik, FH Lee, RW Doms, JD Reeves. Functional impact of HIV coreceptor-binding site mutations. Virology 2006; 351: 226-236.
30. Hwang SS, TJ Boyle, HK Lyerly, BR Cullen. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science 1991; 253: 71-74.
31. Thordsen I, S Polzer, M Schreiber. Infection of cells expressing CXCR4 mutants lacking N-glycosylation at the N-terminal extracellular domain is enhanced for R5X4-dualtropic human immunodeficiency virus type-1. BMC Infect Dis 2002; 2: 31.
32. Ross TM, PD Bieniasz, BR Cullen. Multiple residues contribute to the inability of murine CCR-5 to function as a coreceptor for macrophage-tropic human immunodeficiency virus type 1 isolates. J Virol 1998; 72: 1918-1924.
33. Dragic T, V Litwin, GP Allaway, SR Martin, Y Huang, KA Nagashima, C Cayanan, PJ Maddon, RA Koup, JP Moore, WA Paxton. HIV-1 entry into CD4$^+$ cells is mediated by the chemokine receptor CC-CKR-5. Nature 1996; 381: 667-673.
34. Bannert N, S Craig, M Farzan, D Sogah, NV Santo, H Choe, J Sodroski. Sialylated O-glycans and sulfated tyrosines in the NH2-terminal domain of CC chemokine receptor 5 contribute to high affinity binding of chemokines. J Exp Med 2001; 194: 1661-1673.
35. Zaitseva M, K Peden, H Golding. HIV coreceptors: role of structure, posttranslational modifications, and internalization in viral-cell fusion and as targets for entry inhibitors. Biochim Biophys Acta 2003; 1614: 51-61.
36. Shioda T, JA Levy, C Cheng-Mayer. Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. Proc Natl Acad Sci U S A 1992; 89: 9434-9438.
37. Westervelt P, DB Trowbridge, LG Epstein, BM Blumberg, Y Li, BH Hahn, GM Shaw, RW Price, L Ratner. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J Virol 1992; 66: 2577-2582.
38. Kuiken CL, JJ de Jong, E Baan, W Keulen, M Tersmette, J Goudsmit. Evolution of the V3 envelope domain in proviral sequences and isolates of human immunodeficiency virus type 1 during transition of the viral biological phenotype. J Virol 1992; 66: 4622-4627.
39. Chesebro B, K Wehrly, J Nishio, S Perryman. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J Virol 1992; 66: 6547-6554.
40. Hoffman TL, RW Doms. HIV-1 envelope determinants for cell tropism and chemokine receptor use. Mol Membr Biol 1999; 16: 57-65.
41. Donaldson YK, JE Bell, EC Holmes, ES Hughes, HK Brown, P Simmonds. In vivo distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. J Virol 1994; 68: 5991-6005.
42. Xiao L, SM Owen, I Goldman, AA Lal, JJ deJong, J Goudsmit, RB Lal. CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global HIV-1 isolates: delineation of consensus motif in the V3 domain that predicts CCR-5 usage. Virology 1998; 240: 83-92.
43. Milich L, B Margolin, R Swanstrom. V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. J Virol 1993; 67: 5623-5634.
44. de Jong JJ, A de Ronde, W Keulen, M Tersmette, J Goudsmit. Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J Virol 1992; 66: 6777-6780.
45. Fouchier RA, M Groenink, NA Kootstra, M Tersmette, HG Huisman, F Miedema, H Schuitemaker. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J Virol 1992; 66: 3183-3187.
46. Distler O, PW McQueen, ML Tsang, LA Evans, L Hurren, C Byrne, R Penny, DA Cooper, SF Delaney. Primary structure of the V3 region of gp120 from sequential human immunodeficiency virus type 1 isolates obtained from patients from the time of seroconversion. J Infect Dis 1995; 172: 1384-1387.
47. Briggs DR, DL Tuttle, JW Sleasman, MM Goodenow. Envelope V3 amino acid sequence predicts HIV-1 phenotype (co-receptor usage and tropism for macrophages). AIDS 2000; 14: 2937-2939.
48. Connor RI, KE Sheridan, D Ceradini, S Choe, NR Landau. Change in coreceptor use coreceptor use correlates with disease progression in HIV-1--infected individuals. J Exp Med 1997; 185: 621-628.
49. Sierra S, T Nolden, N Beerenwinkel, B Kupfer, M Däumer, J Selbig, D Hoffmann, JK Rockstroh, M Oette, G Fäktenheuer, H Pfister, T Lengauer, R Kaiser. Identification of HIV-1 determinants of cytopathogenicity and coreceptor usage. Submitted to publication.
50. Innocenti-Francillard P, K Brengel, C Guillon, F Mallet, P Morand, R Gruters, JM Seigneurin. Blood monocytes infected in vivo by HIV-1 variants with a syncytium-inducing phenotype. AIDS Res Hum Retroviruses 1994; 10: 683-690.
51. Albert J, P Stalhandske, S Marquina, J Karis, RA Fouchier, E Norrby, F Chiodi. Biological phenotype of HIV type 2 isolates correlates with V3 genotype. AIDS Res Hum Retroviruses. 1996; 12: 821-828.
52. Isaka Y, A Sato, S Miki, S Kawauchi, H Sakaida, T Hori, T Uchiyama, A Adachi, M Hayami, T Fujiwara, O Yoshie. Small amino acid changes in the V3 loop of human immunodeficiency virus type 2 determines the coreceptor usage for CXCR4 and CCR5. Virology. 1999; 264: 237-243.
53. Shi Y, E Brandin, E Vincic, M Jansson, A Blaxhult, K

Gyllensten, L Moberg, C Brostrom, EM Fenyö, J Albert. Evolution of human immunodeficiency virus type 2 coreceptor usage, autologous neutralization, envelope sequence and glycosylation. J Gen Virol 2005; 86: 3385-3396.
54. Coetzer M, T Cilliers, LH Ping, R Swanstrom, L Morris. Genetic characteristics of the V3 region associated with CXCR4 usage in HIV-1 subtype C isolates. Virology 2006; 356: 95-105.
55. Cardozo T, T Kimura, S Philpott, B Weiser, H Burger, S Zolla-Pazner. Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. AIDS Res Hum Retroviruses 2007; 23: 415-426.
56. Sing T, O Sander, N Beerenwinkel, T Lengauer. Learning mixtures of localized rules by maximizing the area under the ROC curve. in Proceedings of the First International Workshop on ROC analysis in Artificial Intelligence. 2004. Valencia, Spain 96-98.
57. Sander O, T Sing, I Sommer, AJ Low, PK Cheung, P Harrigan, T Lengauer, FS Domingues. Structural descriptors of gp120 V3 loop for the prediction of HIV-1 coreceptor usage. PLoS Computational Biology 2007; 3: e58.
58. Kuhmann SE, P Pugach, KJ Kunstman, J Taylor, RL Stanfield, A Snyder, JM Strizki, J Riley, BM Baroudy, IA Wilson, BT Korber, SM Wolinsky, JP Moore. Genetic and phenotypic analyses of human immunodeficiency virus type 1 escape from a small-molecule CCR5 inhibitor. J Virol 2004; 78: 2790-2807.
59. Marozsan AJ, SE Kuhmann, T Morgan, C Herrera, E Rivera-Troche, S Xu, BM Baroudy, J Strizki, JP Moore. Generation and properties of a human immunodeficiency virus type 1 isolate resistant to the small molecule CCR5 inhibitor, SCH-417690 (SCH-D). Virology 2005; 338: 182-199.
60. Trkola A, SE Kuhmann, JM Strizki, E Maxwell, T Ketas, T Morgan, P Pugach, S Xu, L Wojcik, J Tagat, A Palani, S Shapiro, JW Clader, S McCombie, GR Reyes, BM Baroudy, JP Moore. HIV-1 escape from a small molecule, CCR5-specific entry inhibitor does not involve CXCR4 use. Proc Natl Acad Sci U S A 2002; 99: 395-400.
61. Mori J, M Mosley, M Lewis, P Simpson, J Toma, W Huang, J Whitcomb, G Ciaramella, M Westby. Characterization of maraviroc resistance in patients failing treatment with CCR5-tropic virus in MOTIVATE 1 and MOTIVATE 2. Antiviral Ther 2007; 12: S12.
62. Mosley M, C Smith-Burchnell, J Mori, M Lewis, M Stockdale, W Huang, J Whitcomb, C Petropoulos, M Perros, M Westby. Resistance to the CCR5 antagonist maraviroc is characterised by dose-response curves that display a reduction in maximal inhibition. in 13th Conference on Retroviruses and Opportunistic Infections (CROI). 2006. Denver, USA Abstract 598.
63. Altmann A, N Beerenwinkel, T Sing, I Savenkov, M Doumer, R Kaiser, SY Rhee, WJ Fessel, RW Shafer, T Lengauer. Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance. Antivir Ther 2007; 12: 169-178.
64. Fouchier RA, M Brouwer, SM Broersen, H Schuitemaker. Simple determination of human immunodeficiency virus type 1 syncytium-inducing V3 genotype by PCR. J Clin Microbiol 1995; 33: 906-911.
65. Korber BT, RM Farber, DH Wolpert, AS Lapedes. Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. Proc Natl Acad Sci U S A 1993; 90: 7176-7180.
66. Jensen MA, FS Li, AB van 't Wout, DC Nickle, D Shriner, HX He, S McLaughlin, R Shankarappa, JB Margolick, JI Mullins. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol 2003; 77: 13376-13388.
67. Beerenwinkel N, T Nolden, B Kupfer, J Selbig, M Daeumer, D Hoffmann, J Rockstroh, R Kaiser. Predicting HIV-1 cytopathogenicity and co-receptor usage from V3 envelope sequences by machine lerning. in 1st International AIDS Society Conference on HIV Pathogenesis and Treatment. 2001. Buenos Aires, Argentina Abstract 268.
68. Pillai S, B Good, D Richman, J Corbeil. A new perspective on V3 phenotype prediction. AIDS Res Hum Retroviruses 2003; 19: 145-149.
69. Resch W, N Hoffman, R Swanstrom. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. Virology 2001; 288: 51-62.
70. Sing T, AJ Low, N Beerenwinkel, O Sander, PK Cheung, FS Domingues, J Büch, M Däumer, R Kaiser, T Lengauer, PR Harrigan. Predicting HIV co-receptor usage based on genetic and clinical covariates. Antiviral Ther in press.
71. Low AJ, W Dong, D Chan, T Sing, R Swanstrom, M Jensen, S Pillai, B Good, PR Harrigan. Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. Aids 2007; 21: F17-F24.
72. Watabe T, H Kishino, Y Okuhara, Y Kitazoe. Fold recognition of the HIV-1 V3 loop and flexibility of its crown structure during the course of adaptation to a host. Genetics 2005; 15: 15.
73. Rosen O, M Sharon, SR Quadt-Akabayov, J Anglister. Molecular switch for alternative conformations of the HIV-1 V3 region: implications for phenotype conversion. Proc Natl Acad Sci U S A 2006; 103: 13950-13955.
74. Sharon M, O Rosen, J Anglister. NMR studies of V3 peptide complexes with antibodies suggest a mechanism for HIV-1 co-receptor selectivity. Curr Opin Drug Discov Devel 2005; 8: 601-612.
75. Huang CC, M Tang, MY Zhang, S Majeed, E Montabana, RL Stanfield, DS Dimitrov, B Korber, J Sodroski, IA Wilson, R Wyatt, PD Kwong. Structure of a V3-containing HIV-1 gp120 core. Science 2005; 310: 1025-1028.
76. Cho MW, MK Lee, MC Carney, JF Berson, RW Doms, MA Martin. Identification of determinants on a dualtropic human immunodeficiency virus type 1 envelope glycoprotein that confer usage of CXCR4. J Virol 1998; 72: 2509-2515.
77. Cormier EG, T Dragic. The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor. J Virol 2002; 76: 8953-8957.
78. Hoffman NG, F Seillier-Moiseiwitsch, J Ahn, JM Walker, R Swanstrom. Variability in the human immunodeficiency virus type 1 gp120 Env protein linked to phenotype-associated changes in the V3 loop. J Virol 2002; 76: 3852-3864.
79. Nabatov AA, G Pollakis, T Linnemann, A Kliphius, MI Chalaby, WA Paxton. Intrapatient alterations in the human immunodeficiency virus type 1 gp120 V1V2 and V3 regions differentially modulate coreceptor usage, virus inhibition by CC/CXC chemokines, soluble CD4, and the b12 and 2G12 monoclonal antibodies. J Virol 2004; 78: 524-530.
80. Pollakis G, S Kang, A Kliphuis, MI Chalaby, J Goudsmit, WA Paxton. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. J Biol Chem 2001; 276: 13433-13441.
81. Ross TM, BR Cullen. The ability of HIV type 1 to use CCR-3 as a coreceptor is controlled by envelope V1/V2 sequences acting in conjunction with a CCR-5 tropic V3 loop. Proc Natl Acad Sci U S A 1998; 95: 7682-7686.
82. Rizzuto C, J Sodroski. Fine definition of a conserved

CCR5-binding region on the human immunodeficiency virus type 1 glycoprotein 120. AIDS Res Hum Retroviruses. 2000; 16: 741-749.
83. Hartley O, PJ Klasse, QJ Sattentau, JP Moore. V3: HIV's switch-hitter. AIDS Res Hum Retroviruses 2005; 21: 171-189.
84. Kulkarni S, S Tripathy, K Agnihotri, N Jatkar, S Jadhav, W Umakanth, K Dhande, P Tondare, R Gangakhedkar, R Paranjape. Indian primary HIV-2 isolates and relationship between V3 genotype, biological phenotype and coreceptor usage. Virology 2005; 337: 68-75.

Received: August 29, , 2007 / Accepted: August 30, 2007

Address for correspondence:
Rolf Kaiser
Institut für Virologie
der Universität zu Köln
Fürst Pückler Str. 56
50935 Köln
Germany
Tel.: +49 221 478 3903

Curriculum Vitae

Saleta Sierra-Aragón

Saleta Sierra-Aragón, Ph.D. (born 1972) studied Pharmacy at the University of Granada, Spain. She graduated in June 1995 and in June 1996 completed a master of Sciences in Microbiology. She finished her Ph.D. in February 2001 at the University Autonoma of Madrid, Spain, analizing variability of RNA virus quasispecies and the application of error catastrophe as a new antiviral strategy. Currently she has post-doc position in the Institute of Virology at the University of Cologne where she is engaged in HIV coreceptors research.

Curriculum Vitae

Rolf Kaiser

Rolf Kaiser, PhD, born in 1957, studied biology and performed his PhD in 1990 in Human Genetics at the University of Bonn. Since 1990 he is involved in virus research and diagnostics. From 1990 till 1998 he was working as a group leader in the Institue of Medical Microbiology, University of Bonn and since 1999 as a project leader in the Institute of Virology, University of Cologne. He is heading the research group on HIV- and HBV-resistance and is taking care of the molecular diagnostics in the Institute of Virology. He is member of the scientific board of the European HIV Drug Resistance Workshop and as co-founder of GenaFor being responsible for the organization of the annual AREVIR-GenaFor meeting on technical, bioinformatical, virological and clinical aspects of drug resistance in HIV, HBV.

Curriculum Vitae

Thomas Lengauer

Prof. Dr. Thomas Lengauer, Ph.D. (born 1952) is Director at the Max-Planck Institute for Informatics in Saarbrucken, Germany. His background is in Math (Ph.D. Berlin, Germany 1976) and Computer Science (Ph.D, Stanford 1979). In the seventies he performed research in theoretical computer science, in the 80s on design methods for integrated circuits. He has been engaged in research in computational biology since the beginning of the 90s. His major focuses of research are protein bioinformatics, computational drug screening and design and bioinformatics for understanding and curing diseases. Previously, he held the positions of a full professor at the University of Paderborn, Germany (1984-1992) and of a Director of the Institute for Algorithms and Scientific Computing at the German National Research Center for Computer Science in Sankt Augustin, near Bonn Germany (1992-2001). Dr. Lengauer is a founding member of the International Society for Computational Biology (ISCB), a member of the steering board of the international conference series RECOMB, and he headed the steering board of the European bioinformatics conference series ECCB since its foundation in 2002 until 2005. In 2001 he co-founded the BioSolveIT GmbH, Sankt Augustin, Germany, which develops and distributes Cheminformatics software. In 2003 he received the Konrad Zuse Medal of the German Informatics Society and the Karl Heinz Beckurts Award. He is a member and a senator of the German Academy of Sciences Leopoldina.

Fax: +49 221 478 3904
Email: rolf.kaiser@uk-koeln.de

Curriculum Vitae

Alexander Thielen

Alexander Thielen, born 1979, studied Bioinformatics at the Eberhard Karls University Tübingen. He graduated in March 2006, finishing his studies with a master thesis about the prediction of linear B-cell epitopes and MHC class II binding peptides. Since March 2006, he is a PhD student at the Max-Planck Institute for Informatics in the department of Computational Biology and Applied Algorithmics.