
HIV decision support: from molecule to man

P.M.A. Sloom, Peter V. Coveney, G. Ertaylan, V. Müller, C.A. Boucher and M. Bubak

Phil. Trans. R. Soc. A 2009 **367**, 2691-2703
doi: 10.1098/rsta.2009.0043

References

This article cites 19 articles, 3 of which can be accessed free

<http://rsta.royalsocietypublishing.org/content/367/1898/2691.full.html#ref-list-1>



This article is free to access

Subject collections

Articles on similar topics can be found in the following collections

[medical computing](#) (13 articles)

[computational biology](#) (59 articles)

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *Phil. Trans. R. Soc. A* go to:
<http://rsta.royalsocietypublishing.org/subscriptions>

HIV decision support: from molecule to man

BY P. M. A. SLOOT^{1,*}, PETER V. COVENEY², G. ERTAYLAN¹,
V. MÜLLER³, C. A. BOUCHER⁴ AND M. BUBAK^{1,5}

¹*Computational Science, University of Amsterdam, 1081 HV Amsterdam, The Netherlands*

²*Centre for Computational Science (CCS), University College London, London WC1H 0AJ, UK*

³*Institute of Biology, Eötvös Loránd University, 1117 Budapest, Hungary*

⁴*Department of Virology, Erasmus Medical Center Rotterdam, 3000 CA Rotterdam, The Netherlands*

⁵*Institute of Computer Science AGH, 30-056 Krakow, Poland*

Human immunodeficiency virus (HIV) is recognized to be one of the most destructive pandemics in recorded history. Effective highly active antiretroviral therapy and the availability of genetic screening of patient virus data have led to sustained viral suppression and higher life expectancy in patients who have been infected with HIV. The sheer complexity of the disease stems from the multiscale and highly dynamic nature of the system under study. The complete cascade from genome, proteome, metabolome and physiome to health forms a multidimensional system that crosses many orders of magnitude in temporal and spatial scales. Understanding, quantifying and handling this complexity is one of the biggest challenges of our time, which requires a highly multidisciplinary approach. In order to supply researchers with an interactive framework and to provide the medical professional with appropriate tools and information for making a balanced and reliable clinical decision, we have developed ‘ViroLab’, a collaborative decision-support system (<http://www.virolab.org/>). ViroLab contains computational models that cover various spatial and temporal scales from atomic-level interactions in nanoseconds up to sociological interactions on the epidemiological level, spanning years of disease progression. ViroLab allows for *personalized* drug ranking. It is on trial in six hospitals and various virology and epidemiology laboratories across Europe.

Keywords: human immunodeficiency virus; multilevel modelling; drug resistance; systems biology; decision support

1. Introduction

Since it was recognized in 1981, acquired immunodeficiency syndrome (AIDS) has claimed more than 25 million lives. The human immunodeficiency virus (HIV) epidemic has become a pandemic and today at least 33 million people are living with HIV (Haub 2007; UNAIDS 2008). In a report by Mathers & Loncar (2006) from the World Health Organization, projections of global mortality up to

* Author for correspondence (p.m.a.sloot@uva.nl).

One contribution of 16 to a Theme Issue ‘Crossing boundaries: computational science, e-Science and global e-Infrastructure II. Selected papers from the UK e-Science All Hands Meeting 2008’.

2030 identified HIV/AIDS as one of the three leading causes of death. Each year, approximately a billion dollars is invested globally in HIV/AIDS research to change this scenario. Although almost every step of the virus life cycle has been unravelled in the past 25 years, and 24 distinct drugs targeted against HIV have been approved, all efforts to achieve an overall eradication of the virus have turned out to be ineffective up to now. Life expectancy under highly active antiretroviral therapy (HAART) treatment has however been extended to 21.5 years (Fang *et al.* 2007). Infection by HIV nowadays is being transformed from ‘a death sentence’ to a controllable disease that needs early diagnosis, lifelong patient-tailored treatment, continuous follow-up and strict drug adherence.

Several studies have demonstrated that better outcomes are correlated with patients treated by a clinician with greater expertise (Kitahata *et al.* 2000; Delgado *et al.* 2003). This improvement has been described in terms of mortality, hospitalization rate, cost of care, adherence to medications, etc. The definitions of expertise in these studies have varied, but most relied on the number of patients actively managed (Department of Health and Human Services 2008).

Even for the expert clinician, usually a straightforward relational approach is not possible. Given the wealth of data (sequencing and genotypic resistance testing are becoming routine practice) and confounding factors, treatment decisions are highly patient dependent and there is an urgent need for genotypic resistance interpretation tools that are up to date and transparent (provide references for their data sources). The application of artificial intelligence and computational techniques to biomedicine has resulted in the development of computer-based decision-support systems (DSSs).

DSSs are designed to enable medical professionals to easily keep up to date with the latest clinical expertise and to incorporate this directly in the decision-making process.

The Achilles heel of the DSSs is the sole dependence on the gathering of information from the published literature by experts, thus overlooking the importance of deductions from other research fields such as molecular dynamics (MD) simulations, immune system modelling or Bayesian network learning. Our vision is to provide researchers and medical doctors with a virtual laboratory (VL) for infectious diseases, which we call *ViroLab*, to enable easy access to distributed simulations as well as facilitating, processing and analysis of virological, immunological, clinical, experimental and epidemiological data.

2. The ViroLab approach to HIV

ViroLab consists of complementary, multilevel computational tools as shown in figure 1; the details of the links between those will be explained in §2*a*. Its components are coordinated to form a framework capable of answering complex questions that clinicians and HIV researchers are facing on a daily basis. Hence, while contributing to fundamental research, ViroLab enhances the state of the art in DSSs by offering case-specific predictions from independent resources.

The ViroLab drug ranking system estimates the sensitivity of the patient’s virus for available drugs at the genome level. It is a rule-based scheme that is based on expert opinion, scientific literature, published data and relating genotype to phenotype experiments (Sloot *et al.* 2006).

The HIV population can compensate for the protease inhibitor (PI) resistance mutations by acquiring cleavage site (CS) mutations. To detect the possible associations between PI resistance mutations and CS mutations, covariation analysis of available HIV sequences has been performed.

At the cellular level, we are modelling the co-receptor use of HIV to understand the virus entry process and how this process evolves *in vivo* and its effects on disease progression. Our research in immune system response investigates the effects of HIV infection with HAART treatment and the onset of AIDS on medically measurable parameters such as viral load and CD4+ T-cell count.

At the epidemiological level, we aim to understand the dynamics of HIV epidemics spreading through sexual contact networks. Therefore, we employ a complex network (CN), which describes the heterogeneous population of HIV exposure groups. The CN takes as input the results of the simulations on the molecular and cellular levels and the demographic and behavioural information available.

The integrated holistic approach to decision support provided by ViroLab requires an advanced computational and data management infrastructure. For that we have developed the ViroLab VL, the backbone of ViroLab, which provides researchers and medical doctors with tools to collaborate, share data resources and services using grid technology.

VL is a set of integrated components that, used together, form a distributed and collaborative space for science (figure 2). Multiple, geographically dispersed laboratories and institutes use the VL to plan and perform *experiments* as well as to share their results (Gubała *et al.* 2007). *Experiment* in this context means a so-called *in silico* experiment; that is, a process that combines data and computations in order to obtain new knowledge on the subject in question (Bubak *et al.* 2008). Three groups of users have been identified: clinicians using DSSs for drug ranking; developers who plan complex biomedical simulations; and users who apply prepared experiments (scripts). During experiment execution, provenance data are created and stored. The advanced provenance technology of the VL allows users to trace decisions back to their origin and to share complicated workflows.

In the remainder of the paper, we give a brief overview of the various components of ViroLab and how they fit together. We also present some new results obtained with this VL for infectious diseases.

(a) *From micro to macro*

(i) *Molecular level: drug–protein binding affinity simulations*

The MD simulation takes as an input the chemical composition of the drugs and the various mutations of the viral proteins. It calculates the binding affinity of the drug–protein complexes from which we can derive the fitness of the virus and the associated viral load. This will form an input to the tools working on the cellular level as well as the transmission parameters in the sexual network.

The potential of molecular simulations to enhance our understanding of drug resistance in HIV/AIDS relies ultimately on their capability to achieve an accurate ranking of drug-binding affinities, on clinically relevant time scales.

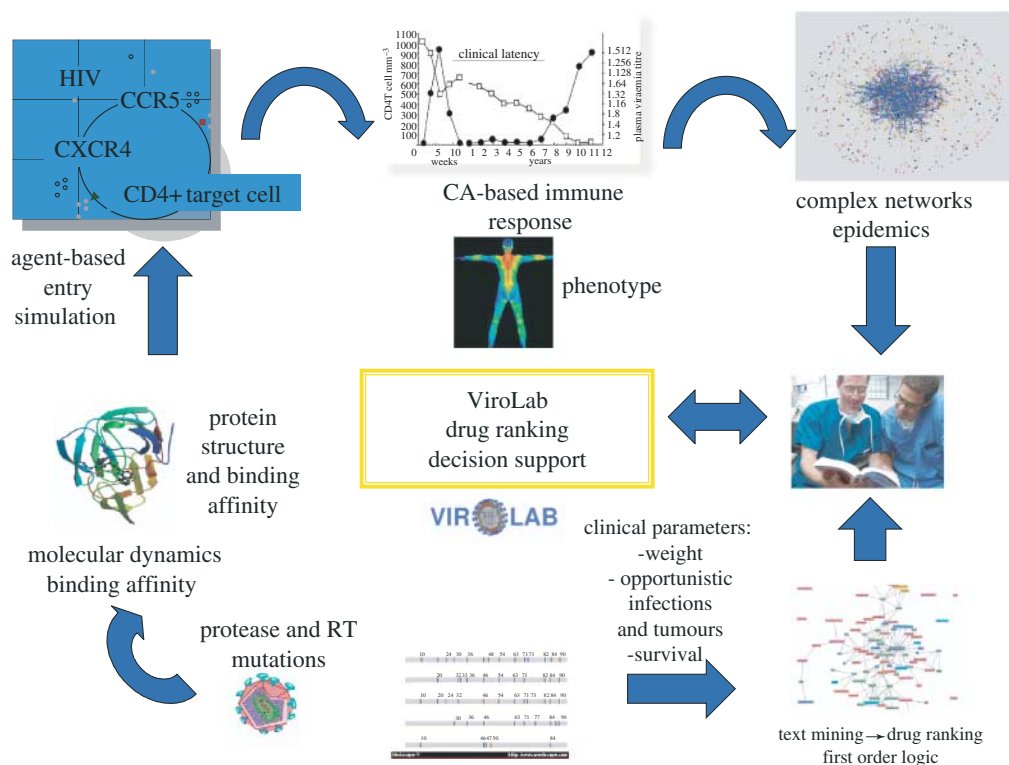


Figure 1. The ViroLab framework with its components. From the bottom going clockwise: patient-specific viral load, mutations and CD4+ information are used by the MD calculations to estimate the binding affinity of various drugs with reverse transcriptase (RT) and protease. The binding affinities as well as the mutations are used to estimate the entry of the virus into the cell. Next the binding affinity, entry process and CD4+ count (squares) and viral load (circles, viraemia) are used to estimate the immune response, and finally the immune response and the binding affinity are used in modelling the individual nodes in a complex sexual interaction network. From the bottom going anticlockwise: context-sensitive text mining distils relationships between drugs and mutations and presents these as ranking rules. The decision-support kernel integrates both types of information (simulation and text mining) into integrated decision support for drug ranking.

We have investigated the potential of the Poisson–Boltzmann surface area (MM/PBSA) methodology to accurately rank the binding affinities of the inhibitor saquinavir to the WT and resistant variants of HIV-1 protease (PR): L90M; G48V; and G48V/L90M. We have also used MM/PBSA to identify the thermodynamic determinants of binding, as a means to understand the mechanisms of emergence of the L90M and G48V mutations under chemotherapeutic pressure.

We identified conformational preferences of the inhibitor and binding site residues and characterize them structurally and energetically (Sadiq *et al.* 2007). For a detailed description of the modelling of the protease–saquinavir complexes, the MD simulation protocol and details of the MM/PBSA calculations see Stoica *et al.* (2008).

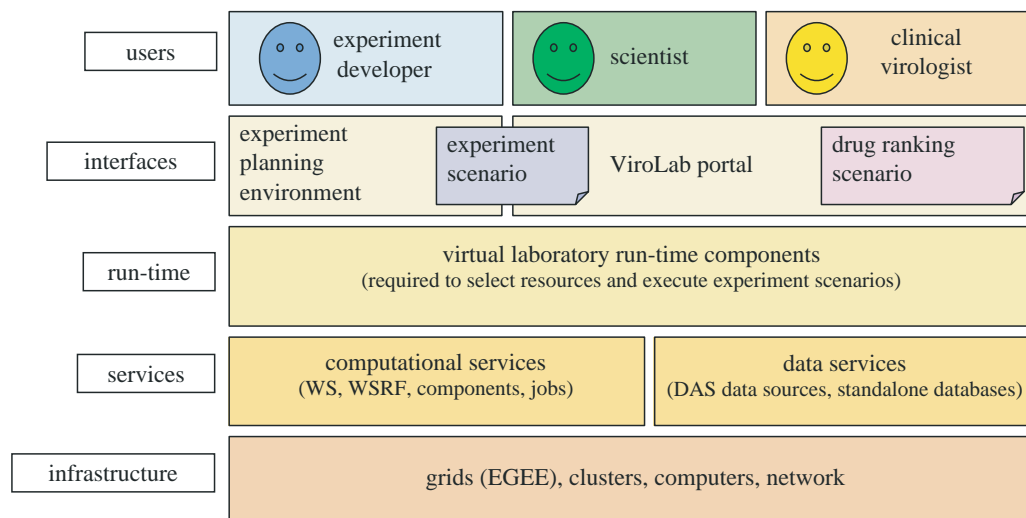


Figure 2. Architecture of the ViroLab VL.

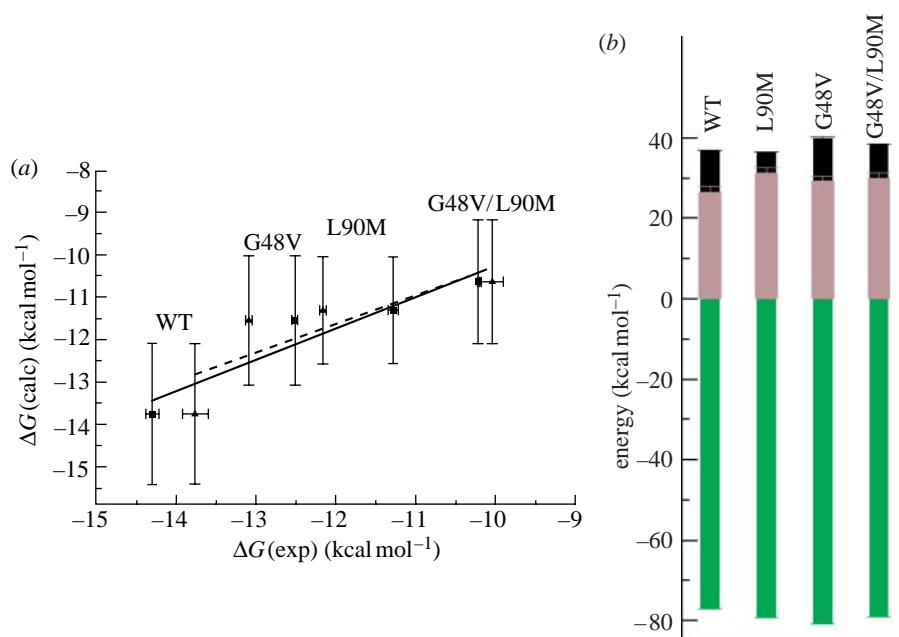


Figure 3. (a) Comparison of the free energies calculated by MM/PBSA ($\Delta G(\text{calc})$) with experimental data ($\Delta G(\text{exp})$). ϵ_1 (squares) are the data from Maschera *et al.* (1996), ϵ_2 (triangles) from Ermolieff *et al.* (1997) (correlation coefficient $\text{CC}(\epsilon_1)=0.96$; $\text{CC}(\epsilon_2)=0.81$). (b) Thermodynamic decomposition of MM/PBSA binding energies for saquinavir with the four PRs under study, WT, L90M, G48V and G48V/L90M, based on analysis of the last 4 ns of post-equilibration MD (black bars, $\Delta G_{\text{elec}-\text{tot}}$; green bars, $\Delta G_{\text{vdw}+\text{np}}$; brown bars, $-T\Delta S$).

Detailed studies indicate that we can resolve accurately the binding affinities of saquinavir with four PR mutations (see figure 3 and Stoica *et al.* 2008). Obviously, these large-scale N -body simulations require a substantial

computational infrastructure. For this, we make use of local supercomputers, the European DEISA supercomputer grid and the US TeraGrid. These large-scale computer infrastructures are accessible through the grid-based VL (Sadiq *et al.* 2008). For instance, a recent DEISA (<http://www.deisa.org>) allocation as a European Virtual Communities Project allocation (approx. 1.25M CPU hours) has allowed us to perform part of our work on EU (HPC) grid infrastructure. At Supercomputing 2008, we demonstrated that we can seamlessly link TeraGrid and DEISA resources using Web services via the application hosting environment (GT4 on TeraGrid and UNICORE 6 on DEISA; see Coveney *et al.* 2007).

The binding affinities thus resolved can be used to estimate the drug response of an individual and the individual's transmission parameter (likelihood of transmission of virus at sexual contact); they provide an indication of the effectiveness of that drug and can then be used in the subsequent immune response simulations as well as in the individual nodes of the CN where it is reflected in the transmission probability.

(ii) *Level of molecule populations: cleavage site compensatory mutation detection*

This tool takes as input a set of Gag-Pol sequences including the gene for the viral PR and one or more of its CSs, and produces a prediction for potential secondary drug resistance mutations in the latter. The cleavage of the Gag-Pol polyprotein by the PR is essential for the infectivity of HIV virions. PI therapy can give rise to primary resistance mutations in the PR, which are often associated with a decreased activity of the enzyme. Impaired function can be partially restored by compensatory mutations in the CS, probably by providing a better substrate for the mutated proteases. Associated pairs of mutations have been detected from large sets of sequences ('populations of molecules') by covariation analysis (Hoffman *et al.* 2003), and we have extended this approach to identify CS mutations that arise in response to primary resistance mutations in the PR. We analysed HIV-1 subtype B nucleotide sequences containing the protease region from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov). Our final alignment contained 30 305 sequences and spanned the entire Gag-Pol region; translation of nucleotide sequences and further analyses were performed using PERL programs and PAUP* (<http://paup.csit.fsu.edu>). Associations between PR and CS mutations were assessed by χ^2 -tests of independence. We also implemented a simple, robust 'phylogenetic jackknifing' test to exclude non-functional covariations that arose by 'common descent' only. Such apparent covariations may arise if a pair of mutations (without functional association) is present in an intensively sampled patient or patient group. Briefly, our method tests whether covariation of a mutation pair can be abolished by excluding any small subset of closely related sequences from the analysis. Failure to do so indicates that the covariation signal is robust over the whole set of sequences and is therefore likely to bear functional relevance.

We found significant association between several pairs of PR and CS mutations. Most associations were positive, which indicates that the CS mutations may indeed compensate the effect of the PR mutations. Interestingly, some mutation pairs demonstrated a negative correlation, indicating that changes in these CSs are less tolerated by the mutant than by the wild-type

protease. Remarkably, a large fraction of the correlated pairs involved a CS position that falls outside the classical definition of a CS, which suggests that positions that lie farther from the cleaved bond may also influence the efficiency of cleavage.

This method generates predictions for CS positions and CS/PR mutation interactions that might be relevant in the complex evolution of PI resistance. The functional relevance of these associations can be corroborated by calculating binding affinities for the mutant PR and CS variants by the computation methods described in §2a(i). Confirmed results are applied to refine the rule set for drug resistance prediction and used in the sexual contact network.

(iii) *Cellular level: modelling the viral entry process through HIV co-receptor tropism*

The entry simulation uses the binding affinities as well as the mutations to estimate the entry of the virus into the cell. The result is fed into the immune system response simulator. In 50 per cent of HIV-1 infected individuals, a co-receptor switch is observed. Given the transmission with an R5 tropic virus, we aim to evaluate the mutation rate by progressing the virus quasispecies to an optimum fitness by changing co-receptor tropism. The emergence rate of X4 tropic variants, the factors leading to it and the timing of the co-receptor switch are investigated to answer whether accumulation of mutations can be understood from the co-receptor switch.

Viral tropism ranges from R5 tropic to R5X4 tropic and finally to X4 tropic variants and allows for biologically natural trade-off between viral mutation rate and the overall reproduction capacity of the viral quasispecies.

A five-point mutational distance is employed in the model from pure R5 tropic to X4 tropic virus where intermediate steps stand for the R5X4 variant. Three kinds of mutation rates are defined as forward (increasing the fitness of the variant), neutral and lethal 0.3, 0.3 and 0.4, respectively (Sanjuán *et al.* 2004).

Each simulation is performed for 10 person years and the experiment has been conducted for a large range of mutation rates. We have collected the data from the population behaviour of T-cells and HIV virions for different substitution rates. For a detailed explanation of the model, see Ertaylan & Sloat (2008).

In figure 4a, we observe 10^{-5} – 10^{-4} interval as maximum mutation rate for the viral quasispecies. After this interval, the viral population experiences a fitness loss due to emergence of lethal mutations in excess numbers. The upper boundary for the mutation rate obtained from our simulation is in very good agreement with the work of Mansky & Temin (1995) and Huang & Wooley (2005).

In figure 4b, no co-receptor switch is observed under 10^{-5} . For increased substitution rates (3×10^{-5} – 10^{-4}), the viral population experiences co-receptor changes in 50–80% of the cases, which coincides with the medical data. Larger substitution rates (more than 10^{-4}) produce continuous fluctuations of each tropic variant (data not shown), indicating that the natural selection halts and all the variants coexist, although the dominant strain is changing over time.

These results clearly demonstrate that there is an upper boundary for the mutation rate for the *selection on co-receptor tropism* to work. It is tempting to speculate that there is a lower boundary for the mutation rate that is determined by the negative selection pressure of the humoral immune response. Hence, the

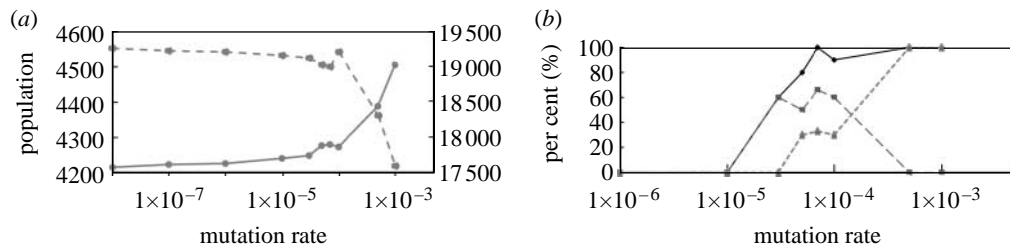


Figure 4. (a) T-cell (solid curve) and HIV (dashed curve) average population sizes versus mutation rate (per nucleotide site per replication cycle) in log scale. The y -axis scale on the left stands for the number of T-cells and that on the right indicates the HIV population size. (b) Co-receptor change rate (solid curve), co-receptor change rate to R5X4 (intermediate variants; long-dashed curve), co-receptor change rate to X4 tropic variants (short-dashed curve) versus mutation rate in log scale.

efficient and timely production of R5X4 and X4 tropic variants can be achieved only with appropriate mutation rate. This information is being fed into the immune system response simulator.

(iv) Cellular level: the immune system response

The immune simulator takes as an input the viral load, the CD4+ count and the drug efficacy and then estimates the temporal HIV dynamics. This dynamics is used as an input to the sexual network nodes to model the various stages of HIV infection of the individuals.

We simulate four phases (acute, chronic, drug treatment and onset of AIDS). We investigate the HIV infection dynamics with therapy using microscopic simulations. Non-uniform cellular automata (CA) are employed to simulate drug treatment of HIV infection, where each computational domain contains different CA rules, in contrast to normal uniform CA models.

We have defined CD4+ T-cells as the main entities. Each cell is assigned healthy/infected or dead flag depending on the interactions with its neighbours (for more details see Sloot *et al.* 2002).

Different drug therapies can be modelled by differences in a response distribution function over time. The function is chosen to model the fact that the drug therapy will not immediately influence all of the infected cells, but rather it will affect part of them at each time step. At the same time, the function accounts for the concept of drug-resistant virus strains emerging.

The simulation result for treatment naive individual dynamics is shown in figure 5. The result represents qualitatively the HIV infection dynamics. The first phase indicates the fast proliferation of the original HIV strains before the actual immune system response (acute phase). This phase ends when specific immune response occurs for these strains. The next phase, the chronic phase that takes years, is the phase where the viral load increases slowly and CD4 counts decrease slowly. When CD4+ T-cell counts drop to a certain level (normal 200–500 counts ml^{-1}), the drug therapy is started. In this phase, virus replication is blocked and CD4+ T-cell counts increase. Once resistant strains against the drugs evolve, the last phase of the disease occurs disrupting the whole immune system.

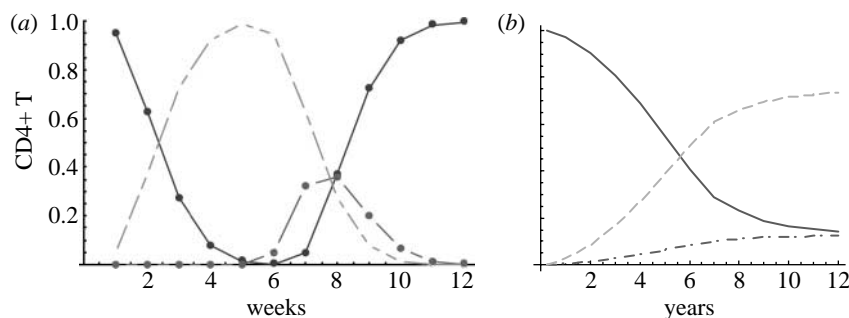


Figure 5. Three-phase dynamics: acute, chronic and AIDS with two time scales (a) weeks and (b) years. The solid, dashed and dot-dashed lines represent healthy, infected and dead cells, respectively.

The main success of the present model is the adequate modelling of the four phases of HIV infection with different time scales in a single model. Our simulation results show a qualitative correspondence to clinical data. The individual HIV dynamics similar to that shown [figure 5](#) is given as an input to the individual nodes in the sexual networks and population dynamics simulation.

(v) *Population level: complex network model for HIV population dynamics*

The CN model takes as an input the binding affinities from the MD simulator and the immune response together with epidemiological and demographic information from Centre of Disease Control-like databases. It stochastically predicts the evolution of the HIV dynamics as well as the transmission of resistance.

Despite the availability of a large number of mathematical models describing the spreading of HIV, a good understanding of the dynamics through numerical analyses is still a major challenge. The estimate of HIV/AIDS incidence is quite uncertain since many people are still unaware of their infection due to the long asymptotic incubation period. Heterogeneity of the infected population and the various routes of infection among its members exert additional layers of complexity. Therefore, it is essential to combine epidemiological processes with sociological models and network sciences. Here we propose a new way to model HIV infection spreading through the use of dynamic CNs.

The heterogeneous population of HIV exposure groups is described through a unique network degree probability distribution. The network model is constructed as a dynamical scale-free network, wherein a node represents each individual and the edges are the links between the individuals. The infection spreads along these edges.

The time evolution of the network nodes is modelled by a Markov process and incorporates HIV disease progression through a *configuration model* of T-cell CD4 count for each node.

While constructing the model, we have taken into account that HIV is primarily a sexually transmittable disease (STD). In our model, we identify three ways of transmission:

- men who have sex with men (MSM),
- heterosexual population, and
- injection drug users (IDU); STD dynamics is not implemented for this group.

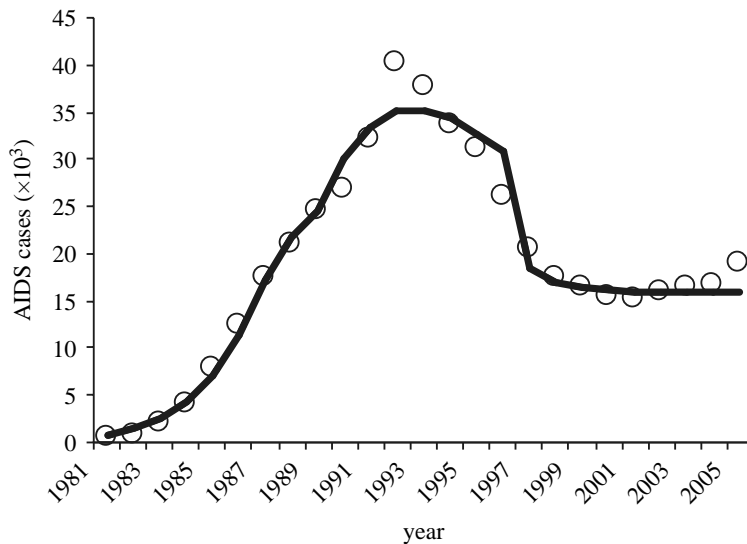


Figure 6. Simulation results and reported data for the AIDS epidemic in the USA. MSM exposure groups, power-law distribution with $\gamma=1.6$ and $k_{\max}=250$, $I_0=0.32\%$, $d=0.04$, $\lambda=0.44$. Circles are the annual officially registered number of AIDS cases; the solid curve indicates the simulation results. Comparable results were obtained for heterosexuals and IDU (data not shown).

Each risk group has its own probability of infection per partner λ as well as specific demographic factors and related contact networks.

The results are validated against historical data of AIDS cases in the USA as recorded by the Center of Disease Control. These data include at least three epochs of the epidemic evolution, pre-ARV, ARV and HAART treatments, and three kinds of HIV spreading.

As can be seen from [figure 6](#), the simulation results provide very close agreement to the officially registered annual AIDS cases. The obvious feature in [figure 6](#) is a substantial decline of AIDS cases after introduction of HAART (1996). Our model correctly reproduces this decline and the temporal separation of the peaks of the AIDS epidemic in the heterosexual and MSM populations (see [Sloot *et al.* 2008a](#)).

In conclusion, the presented parametrized CN model accurately describes the dynamics of HIV spreading taking into account all the existing kinds of HIV spreading. We find a remarkably good correspondence between the number of simulated and registered HIV cases, indicating that our approach to modelling the dynamics of HIV spreading through a sexual network is a valid approach that opens up completely new ways of reasoning about various medication scenarios.

This tool, which integrates the clinical information with social behaviour, is used by epidemiologists and by decision-makers to study scenarios of treatment and prevention measures.

(b) *The virtual laboratory*

The VL hosts biomedical information related to viruses (proteins and mutations), patients (viral load, CD4 count, etc.) and literature (on drug resistance etc.), and it enables one to plan and run experiments transparently on distributed resources in grid systems, clusters and standalone computers.

The VL integrates various tools, modules, protocols and interfaces into a single virtual space, where various types of users are able to (collaboratively) perform tasks. In the upper layer, we have the set of users performing tasks using the tools that are grouped in the interfaces: there are dedicated tools for each group. The centrepiece of the architecture, the unified run-time system, is the part that allows these various users and their tools to understand each other. What is more, this run-time layer serves as a bridge to resources dispersed in the VL: both data sources and computational services. Finally, these services run on physical equipment that is at their disposal. There are multiple solutions here, which are supported by our VL; among others, it supports also large grid computing test beds. Currently, the EGEE test bed through the LCG software is supported as well as the DEISA resources.

In this way, the VL provides an integrated environment to organize computational and ranking experiments in HIV. Details of the ViroLab VL are beyond the scope of this paper; for downloads and demonstrations, we refer the interested reader to: <http://virolab.cyfronet.pl/trac/vlvi>.

3. Conclusions

ViroLab is advancing the state of the art by promoting the theoretical understanding of infectious diseases while bridging the gaps not only within each research field but also across different fields. It is integrated in a grid-based VL (Bubak *et al.* 2008) that facilitates medical knowledge discovery and the interaction and integration of the wide variety of HIV modelling that is available. This offers the potential to provide medical doctors with a DSS to rank drugs targeted at individual patients. Its infrastructure provides virologists with an advanced environment to study trends on molecular, genome and cellular, individual and epidemiological levels.

For instance, to tackle problems such as the development of drug resistance to saquinavir, we envisage being able to simulate the probability of drug resistance with the binding affinity calculator (BAC; Sadiq *et al.* 2008) and identify compensatory mutations in the CSs. These results could then be integrated with a fitness landscape. This could then be used to interpret patient data in high resolution and provide an improved DSS.

Another potential scenario concerns the transmission of drug resistance at the population level. Viral load and infectivity can be estimated on an individual level using the BAC and the immune system response model. These parameters are then supplied to the HIV population dynamics model, eventually resulting in a network with predicted drug-resistant nodes. This could help policy makers to answer questions such as: ‘which drug should be made free and publicly available without creating a universal drug-resistant population’ or ‘which combination of HAART would give the best outcome in resource-poor areas where adherence is a major problem?’

In this ongoing research, interesting results have already been obtained, showing the viability and the extensibility of the approach taken (e.g. Deforche *et al.* 2007, 2008; Bubak *et al.* 2008; Sloom *et al.* 2008*a,b*; Stoica *et al.* 2008). The system we have described is still under development with new functionalities added frequently through extensive usability studies in a series of hospitals across Europe.

The authors would like to thank Breannán Ó. Nualláin, Alfredo Tirado Ramos and Annemie Vandamme for their valuable feedback. This research was supported by the European Union through the ViroLab project (www.virolab.org), EU project no. IST-027446.

References

- Bubak, M. *et al.* 2008 Virtual laboratory for development and execution of biomedical collaborative applications. In *Proc. 21st IEEE Int. Symp. on Computer-Based Medical Systems (CBMS 2008)*, Jyväskylä, Finland, 17–19 June 2008, pp. 373–378. Washington, DC: IEEE Computer Society.
- Coveney, P. V., Saksena, R. S., Zasada, S. J., McKeown, M. & Pickles, S. 2007 The application hosting environment: lightweight middleware for grid-based computational science. *Comput. Phys. Commun.* **176**, 406–418. (doi:10.1016/j.cpc.2006.11.011)
- Deforche, K. *et al.* 2007 Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. *Infect. Genet. Evol.* **7**, 382–390. (doi:10.1016/j.meegid.2006.09.004)
- Deforche, K. *et al.* 2008 Estimation of an *in vivo* fitness landscape experienced by HIV-1 under drug selective pressure useful for prediction of drug resistance evolution during treatment. *Bioinformatics (Oxf., Engl.)* **24**, 34–41. (doi:10.1093/bioinformatics/btm540)
- Delgado, J., Heath, K., Yip, B., Marion, S., Alfonso, V., Montaner, J., O'Shaughnessy, M. V. & Hogg, R. S. 2003 Highly active antiretroviral therapy: physician experience and enhanced adherence to prescription refill. *Antiviral Ther.* **8**, 471–478.
- Department of Health and Human Services 2008 *AIDSinfo*. Retrieved 8 April 2008, from Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. See <http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf>.
- Ermolieff, J., Lin, X. & Tang, J. 1997 Kinetic properties of saquinavir-resistant mutants of human immunodeficiency virus type 1 protease and their implications in drug resistance *in vivo*. *Biochemistry* **36**, 12 364–12 370. (doi:10.1021/bi971072e)
- Ertaylan, G. & Sloot, P. 2008 A complex automata model of HIV-1 co-receptor tropism. In *Proc. 3rd Int. Workshop in Targeting HIV Entry, Washington, DC*, vol. 2 (eds C. Boucher & J. Schapiro). *Reviews in antiretroviral therapy*, p. 25. Utrecht, The Netherlands: Virology Education.
- Fang, C. T. *et al.* 2007 Life expectancy of patients with newly-diagnosed HIV infection in the era of highly active antiretroviral therapy. *QJM Mon. J. Assoc. Phys.* **100**, 97–105. (doi:10.1093/qjmed/hcl141)
- Gubała, T. *et al.* 2007 ViroLab virtual laboratory. In *Cracow Grid Workshop 2007 Workshop Proc.*, Cracow, Poland, pp. 35–40.
- Haub, C. 2007 World population highlights: key findings from PRB's 2007 world population data sheet. *Population bulletin*. Washington, DC: Population Reference Bureau.
- Hoffman, N. G., Schiffer, C. A. & Swanstrom, R. 2003 Covariation of amino acid positions in HIV-1 protease. *Virology* **314**, 536–548. (doi:10.1016/S0042-6822(03)00484-7)
- Huang, K. & Wooley, D. 2005 A new cell-based assay for measuring the forward mutation rate of HIV-1. *J. Virol. Methods* **124**, 95–104. (doi:10.1016/j.jviromet.2004.11.010)
- Kitahata, M., Van Rompaey, S. & Shields, A. 2000 Physician experience in the care of HIV-infected persons is associated with earlier adoption of new antiretroviral therapy. *J. Acquir. Immune Defic. Syndr.* **24**, 106–114.
- Mansky, L. & Temin, H. 1995 Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* **69**, 5087–5094.
- Maschera, B., Darby, G., Palú, G., Wright, L., Tisdale, M., Myers, R., Blair, E. D. & Furfine, E. S. 1996 Human immunodeficiency virus. Mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex. *J. Biol. Chem.* **271**, 33 231–33 235. (doi:10.1074/jbc.271.52.33231)
- Mathers, C. & Loncar, D. 2006 Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* **3**, e442. (doi:10.1371/journal.pmed.0030442)

- Sadiq, S., Wan, S. & Coveney, P. V. 2007 Insights into a mutation-assisted lateral drug escape mechanism from the HIV-1 protease active site. *Biochemistry* **46**, 14 865–14 877. (doi:10.1021/bi700864p)
- Sadiq, S., Wright, D., Watson, S., Zesada, S., Stoica, S. & Coveney, P. V. 2008 Automated molecular simulation based binding affinity calculator for ligand-bound HIV-1 proteases. *J. Chem. Inf. Model.* **48**, 1909–1919. (doi:10.1021/ci8000937)
- Sanjuán, R., Moya, A. & Elena, S. 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl Acad. Sci. USA* **101**, 8396–8401. (doi:10.1073/pnas.0400146101)
- Sloot, P. M. A., Chen, F. & Boucher, C. 2002 Cellular automata model of drug therapy for HIV infection. In *5th Int. Conf. on Cellular Automata for Research and Industry, Geneva, Switzerland*, vol. 2493 (eds S. Bandini & B. T. Chopard), pp. 282–293. Lecture Notes in Computer Science.
- Sloot, P. M. A., Tirado-Ramos, A., Altintas, I., Bubak, M. T. & Boucher, C. A. 2006 From molecule to man: decision support in individualized e-health. *IEEE Comput.* **39**(11), 40–46.
- Sloot, P. M. A., Ivanov, S., Boukhanovsky, A., van de Vijver, D. & Boucher, C. 2008a Stochastic simulation of HIV population dynamics through complex network modeling. *Int. J. Comput. Math.* **85**, 1175–1187. (doi:10.1080/00207160701750583)
- Sloot, P. M. A., Coveney, P. V., Bubak, M. T., Vandamme, A.-M., Nualláin, B.Ó., van de Vijver, D. & Boucher, C. A. B. 2008b ViroLab: a collaborative decision support system in viral disease treatment. *Rev. Antiviral Ther.* **3**, 4–7.
- Stoica, I., Sadiq, S. K. & Coveney, P. V. 2008 Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. *J. Am. Chem. Soc.* **130**, 2639–2648. (doi:10.1021/ja0779250)
- UNAIDS 2008 *Report on the global AIDS epidemic*. Geneva, Switzerland: UNAIDS.