Data and text mining

Effect of spatial distribution of T-Cells and HIV load on HIV progression

Frank M. Graziano¹, Samira Y. Kettoo³, Judy M. Munshower³, Jack T. Stapleton⁴ and George J. Towficž²,*

¹University of Wisconsin Hospital and Clinics, Wisconsin-Madison, USA, ²Department of Computer Science, ³Department of Mathematics, Clarke College, Dubuque, IA and ⁴Department of Internal Medicine, University of Iowa, Iowa, USA

Received on March 7, 2007; revised on December 25, 2007; accepted on January 6, 2008
Advance Access publication January 10, 2008
Associate Editor: Thomas Lengauer

ABSTRACT

Motivation: We present a spatial-temporal (ST) human immunodeficiency virus (HIV) simulation model to investigate the spatial distribution of viral load and T-cells during HIV progression. The proposed model uses the Finite Element (FE) method to divide a considered infected region into interconnected subregions each containing viral population and T-cells. HIV T-cells and viral load are traced and counted within and between subregions to estimate their effect upon neighboring regions. The objective is to estimate overall ST changes of HIV progression and to study the ST therapeutic effect upon HIV dynamics in spatial and temporal domains. We introduce sub-regional (spatial) parameters of T-cells and viral load production and elimination to estimate the spatial propagation and interaction of HIV dynamics under the influence of a 3TC D4T Reverse Transcriptase Inhibitors (RTI) drug regimen.

Results: In terms of percentage change standard deviation, we show that the average rate per 10 weeks (throughout a 10-year clinical trial) of the ST CD4+ change is 5.35% 1.3 different than that of the CD4+ rate of change in laboratory datasets, and the average rate of change of the ST CD8+ is 4.98% 1.93 different than that of the CD8+ rate of change. The half-life of the ST CD4+ count is 1.68% 3.381 different than the actual half-life of the CD4+ count obtained from laboratory datasets. The distribution of the viral load and T-cells in a considered region tends to cluster during the HIV progression once a threshold of 86–89% viral accumulation is reached.

Availability: Executable code and data libraries are available by contacting the corresponding author.

Implementation: C++ and Java have been used to simulate the suggested model. A Pentium III or higher platform is recommended.

Contact: george.towfic@clarke.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

To estimate the effect of change of human immunodeficiency virus (HIV) dynamics (drug regimens, T-cell count and viral load) on HIV progression, many top-down mathematical models, based on the solution of a set of Ordinary Differential Equations (ODE), have been proposed (Kirschner, 1999; Muller et al., 2001; Perelson, 2001; Stafford et al., 1999). One drawback of this approach is that it is restricted by the ODE’s ability to estimate temporal but not spatial changes of system variables.

Another possible approach to simulate HIV progression is the bottom-up approach based on Agent-Based (AB) models (Teweldemedhin and Marwala, 2004) and Cellular Automata (CA) models (Neumann, 1966). This approach can be used to construct and understand the components of a complex system by assigning different agents (prototypes), developed in different ways, to simulate a considered model (such as an HIV progression model). During the development process, agents cooperate to reward successful agents and impede or eliminate unsuccessful agents. Although this approach can be used to study HIV progression at the population level (Teweldemedhin and Marwala, 2004), it has not been implemented at the cellular level since it is very difficult to keep updating agents and their spatial connections when the system dynamics rapidly change during the acute phase of HIV progression (Callaghan, 2005). Incorporating a drug resistance factor into the progression process adds more randomness and hence significant complexity to the AB and CA models.

In this work, we extend our previous research (Draghici et al., 2003) in which an Independent Component Analysis (ICA) algorithm had been used for two purposes: (1) to generate identical patterns of HIV dynamics and (2) to eliminate redundant noise in each pattern to produce reliable datasets from available laboratory data.

We use the refined CD4+ and CD8+ T-cell counts and viral load acquired from the ICA algorithm to develop a Finite Element (FE) model that traces the motion of T-cells and viral load in a considered lymph domain. The objective of this spatial-temporal (ST) tracking is to provide an understanding of the internal distribution of the T-cells and the HIV viral load during the HIV life cycle (an example of the distribution is depicted in Fig. 5). A distribution profile can be used in different applications that require HIV modeling. Examples of such applications include: (1) predicting the spatial virion rate of flow to be used in drug design (including estimated drug
resistance and spatial therapeutic effect), (2) predicting viral load and T-cell interaction and (3) predicting overall HIV progression in terms of its local constituencies.

We compared the analytical results obtained from the model with laboratory datasets obtained from the University of Iowa, the University of Wisconsin and Stanford University (Rhee et al., 2003). We show that the ST tracking of HIV progression provides useful information that can be used to understand and control the overall progression of major HIV dynamics including the progression of CD4+ and CD8+ T-cell counts and viral load, and therapeutic effect.

2 PATIENTS AND METHODS

We used 1500 patient records. From the three libraries, we selected datasets that yielded 93.7% similarity, in terms of number of treatment trials, type of regimens and CD4+ on-therapy baseline.

Patients using the 3TC D4T Reverse Transcriptase Inhibitors (RTI) regimen have been selected due to the availability of more trials for this regimen. Using the resulting refined datasets, a mathematical model was developed, estimating important parameters (such as base CD4+ and CD8+ T-cell counts and viral load, during different trials) determined by the three HIV clinics and available in the three considered datasets. The proposed mathematical model projected the change of the CD4+ and CD8+ T-cell counts and viral load in terms of their rate of production, elimination and therapy duration. Validation and accuracy of the model was assessed by comparing the analytical results of the model with that of the patients’ records. Figure 1 illustrates the main steps of our simulation. We divide a proposed region of volume 1000 ml³ into 1000 FE subregions and track the HIV dynamics in each subregion. The temporal change of each subregion is calculated during the treatment duration, using the Boltzmann equation (R D Kingdon and White, 1992), by determining the change of the HIV dynamics due to the flow of T-cells and viral load from one neighboring subregion to another.

3 SYSTEM EQUATIONS

We consider the T-lymphocytes and viral load as a viscous, incompressible fluid occupying, at time \( t \), a bounded rectangular region \( \mathbb{R}^2 \) with boundary \( \Gamma \). In constructing the governing differential equations, we consider the Eulerian conservative form of the continuity equation to preserve mesh size and dimensions (Douglas et al., 2000). The rate of change of viral load and T-cells (calculated from Equation 1) will be used in the FE model (Equations 7–9) to calculate the distribution densities of the T-cells and viral load.

\[
\begin{align*}
\frac{\partial u}{\partial t} &= -u \cdot \nabla u + u \\
\frac{\partial v}{\partial t} &= -v \cdot \nabla v
\end{align*}
\]  

Equation 1 is considered under a Neumann boundary condition: \( \nabla \cdot u = 0 \) and \( \nabla \cdot v = 0 \). In Equation 1, \( u \) and \( v \) are functions of the spatial Cartesian coordinates \( x \) and \( y \) and the time coordinate \( t \). Figure 2 displays the actual and calculated CD4+ and viral load values in 10-week time periods throughout 10 years of HIV progression. In this model, \( u \) represents the viral load and \( v \) represents the CD4+ count as depicted in Figure 2. The vector \( n \) is normal to the gradient operator. The gradient \( \nabla \) is defined as:

\[
\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}.
\]

The following formulas and data values are derived from the laboratory datasets by calculating the average change of viral load, CD4+ and CD8+ T-cell counts during different trials.

\[
\begin{align*}
\psi_1 &= r_\gamma \psi_0 \\
\psi_2 &= r_\alpha \psi_1 \\
\psi_3 &= r_t \psi_2 \\
\psi_0^{(n+1)} &= \frac{\psi_1 + \psi_2 + \psi_3 \eta}{\psi_0^{(n+1)}}
\end{align*}
\]

where

\[
\begin{align*}
r_\gamma &= \gamma_0 \\
\gamma_0 &= \text{gamma} \\
r_\alpha &= \alpha_0 \\
\alpha_0 &= \text{alpha} \\
r_t &= t_0 \\
\psi_0 &= \text{initial values of } \psi
\end{align*}
\]

The variables \( u \), \( v \), \( \omega \) and \( \varphi \) are the rate of viral change due to therapy, the rate of CD4+ change with respect to time, the rate of CD8+ change with respect to time, and the overall rate of change in viral load and CD4+ and CD8+ count respectively. The considered initial values (measured in days)
are: the rate of viral load production, $\beta = 0.001$; the rate of natural viral load elimination, $\gamma = 7.4 \times 10^{-4}$; the rate of viral load elimination due to therapy $\kappa = 3.1 \times 10^{-2}$; the uninfected cell production rate $\varepsilon_{uv} = 0.01$; the uninfected cell elimination rate $\varepsilon_{uv} = 0.007$; the average weight of virus effect as time changes is $r = 0.001$; the average rate for a virus to locate a CD4+ cell is $\eta = 1.09$ and $\delta = 0.2$, is the rate of production of immature T-cells that will not contribute in the production of new T-cells. The CD8+ count at a new time step is calculated using the CD8+ count at the current time step, the rate of production, $r_1 = 0.001$, and the rate of elimination of CD8+ cells, $r_2 = 0.018$. $\varepsilon$ is the change in the number of CD4+ cells at a given time ($u^{(n)}_j - t^{(n-1)}_j$).

In the above systems of equations (Equations 2–5), the value of the HIV variables at an advanced time step is estimated according to the existing value of the considered variable plus the extra values that are added to the system due to the considered production rate of that variable minus the eliminated values due to natural death or therapy. We discuss the meaning of individual terms of Equations 2–5 in the appendix.

We considered the following FE trial function (Douglas et al., 2000):

$$f(r) = \begin{cases} 1 - 6r^2 + 8r^3 - 3r^4 & 0 \leq r \leq 1 \\ 0 & r > 1 \end{cases}$$

The above trial function (Belytschko et al., 1996; Chen and Raju, 2003) is used to calculate the following FE weak solution:

$$\varphi_j(x, y) = f \left( \frac{d_j(x) d_j(y)}{m_{nj}} \right)$$

The approximation value $\varphi_j(x, y)$ is considered to be weak since it is defined at each element of the finite element mesh rather than the entire $\mathbb{R}^d$ domain (i.e. the solution of the system of Equation 1 is a combination of local solutions calculated at each finite element). For any variable, $\eta$, $d_j(\eta) = |\eta - \eta_j|/\varepsilon$.

4 MODEL EVALUATION

To evaluate our suggested mathematical model, we first compared the analytical datasets obtained from the mathematical model with those recorded in the three considered HIV libraries. Figure 3 depicts the result of the comparison obtained for 500 patients with an on therapy CD4+ count of 300 and on-the-therapy baseline HIV RNA of 4.1 (measured in Log_{10} copies/ml).

The square boxes in Figure 3 represent the average of the calculated CD4+ cells for a 10-year period. The solid line represents the average values obtained from the three laboratory datasets for the same time period. Therapy at Figure 3...
started at the 967-th day of viral infection and ended at the 2817-th day. The figure shows that there was approximately 30 days delay for the therapy to take effect and approximately 63 days for the therapy to lose its effect in restraining the viral progression (this is shown in the figure as a surge in the CD4⁺ cells after the continuous decline of CD4⁺ cells during the first 63 days). The analytical datasets and the laboratory datasets for the 500 patient records differ by approximately 2.4 SDs. Using a confidence interval table for 500 data points, this accounts for about 92% confidence in the proposed finite element model. Given that the three libraries are completely independent in measuring patient datasets, and the results obtained from the ST model (depicted in Fig. 2), we concluded that the derived Finite Element model can be used for further analysis to estimate the effect of spatial viral propagation and drug resistance on the overall HIV progression. Using the considered datasets and in terms of mean SD, we calculated the difference of the average value of the actual laboratory CD8⁺ count to be 216.413 /C6 30.568 and the average CD8⁺ ST values as 250.586 /C6 33.649, which confirms the model’s robustness in predicting the change in the T-cells and viral load. Using the datasets in Figure 2, we calculated the difference between the average rate of change of the actual laboratory CD4⁺ count and the calculated ST CD4⁺ count to be 5.35% /C6 1.3 and the difference between the average rate of change of the actual laboratory CD8⁺ dataset and the calculated ST datasets to be 4.98% /C6 1.93. To test the convergance of the FE model with respect to different mesh sizes, we used ANOVA. Table 1 shows comparison between the analytical data obtained from four finite element mesh sizes of 300, 500, 1000 and 1500. The fact that the F value between the four groups is approximately 0.001 indicates that there is no statistically significant difference between the CD4⁺ values calculated using the four mesh sizes. Figure 4 depicts the half-life average for both actual laboratory and ST change of 14 CD4⁺ on-therapy baseline values. Figure 4 shows that a reasonable r² value of 0.913 is obtained from the polynomial y = -0.0058x² + 0.2106x + 2.8029x + 18.613. This indicates that a polynomial fit can be used with 82% confidence to predict the half-life value of CD4⁺ count given a considered CD4⁺ baseline value. This result is consistent with that obtained by Drusano and Daniel (Drusano and Stein, 1998). The average difference between the spatial and overall half-life of the CD4⁺ count is 1.68% /C6 3.381. This shows that there is no significant difference between the spatial and overall models in calculating the half-life for the CD4⁺ count.

5 RESULTS

Figure 5 shows three samples of viral and T-cell distribution obtained from the implemented FE model. The figure helps in understanding viral load progression under the 3TC D4T RTI regimen. The following observations can be inferred from Figure 5: (1) initial viral distribution is important in determining concentration of viral load and T-cells during the HIV progression. The second column shows that in 10 years, the viral load will spread rapidly: 57% faster than the third column’s initial state and 73% faster than the distribution shown in the fourth column. (2) Viral load and T-cells tend to accumulate as they expand or shrink during HIV progression. When a therapy manages to break the viral load concentration (as in the third and fourth columns), the effect of HIV progression would be significantly reduced.

Table 1. A single factor ANOVA test showing that there is no significant difference between the CD4⁺ values calculated for 4.3Log10 viral load baseline using 300, 500, 100 and 1500 finite element mesh

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>300 mesh</td>
<td>33</td>
<td>11697</td>
<td>354.4545</td>
<td>1202.818</td>
</tr>
<tr>
<td>500 mesh</td>
<td>33</td>
<td>11641</td>
<td>352.7576</td>
<td>979.8769</td>
</tr>
<tr>
<td>1000 mesh</td>
<td>33</td>
<td>11654</td>
<td>353.1515</td>
<td>1015.008</td>
</tr>
<tr>
<td>1500 mesh</td>
<td>33</td>
<td>11792</td>
<td>357.3333</td>
<td>790.9792</td>
</tr>
</tbody>
</table>

ANOVA

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>17976.15</td>
<td>3</td>
<td>5992.06</td>
<td>0.0097</td>
<td>0.0547</td>
</tr>
<tr>
<td>Within groups</td>
<td>127637.8</td>
<td>128</td>
<td>997.1705</td>
<td>F- Crit</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>145614</td>
<td>131</td>
<td>2.675387</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Mean CD4⁺ values for 500 patient records with baseline CD4⁺ count of 300 and baseline HIV RNA of 4.1 Log10 copies/ml.

Fig. 4. Baseline CD4⁺ Count versus ST CD4⁺ half-life (days).

Downloaded from https://academic.oup.com/bioinformatics/article-abstract/24/6/855/193068 by guest on 04 May 2018
The distribution patterns depicted in Figure 5 show that the finite element model provides insight into the complicated dynamic process of HIV progression. Figure 5 also shows that the initial viral distribution determines the pace of HIV progression. The viral load and T-cells tend to spread in groups during HIV progression once a threshold of 86-89% viral accumulation is reached.

ACKNOWLEDGEMENTS

We acknowledge the valuable contribution of Dr Robert Shafer of Stanford University for providing us with password access to the Stanford HIV Database (http://hivdb.stanford.edu). We also thank Dr Sorin Draghici of Wayne State University for his valuable comments. Dr Thomas Lengauer and this article’s anonymous reviewers made valuable suggestions and comments that significantly improved the content of this article. This research was funded through a grant by the Maytag Foundation.

Conflict of Interest: none declared.

REFERENCES


APPENDIX

Equation 2 calculates the number of viral load in the next time step in terms of the values of different system parameters. The first term of the right-hand side of Equation 2 represents the number of viral load at the current time step. The second term measures the increase in viral load due to viral load production rate $\beta$. The third and fourth terms calculate the reduction of viral load due to rate of natural death ($\gamma$) and rate of therapy effect ($\delta$), respectively. The final two terms in Equation 2 calculate the overall change ($\varphi$) in the system and the new value of CD8+ count ($\omega$), respectively. Both $\varphi$ and $\omega$ will be discussed in more detail below in the discussion of Equations 4–5. It is worthwhile to mention here that the CD8+ count is added to Equation 2 due to the effect of CD8+ count on the production of new CD4+ cells [some authors simply estimate the number of CD8+ count as 2/3 of the total CD4+ count (Haraba and Dolezal, 1996)]. After multiplying, the first two terms in the ratio of Equation 3 are the current CD4+ count and the new naturally produced CD4+ cells ($\varepsilon_{2p} + v^{(n)}$) respectively. The last three terms in Equation (3) represent the number of CD4+ T-cells that will be eliminated from the system due to natural death ($\varepsilon_{2e} + v^{(n)}$), the number of CD4 -cells that will be eliminated as viral loads detect and infect them at the considered time step ($\tau_{e}^{(n)}$), and the number of immature CD4+ cells that will not contribute in the production process of new T-cells $\delta v^{(n)}$. The variable $\varepsilon$ (the change in number of CD4+ cells in consecutive times) is used to simulate the biological response of the immune system to produce more CD4+ cells when the current number of CD4+ cells reduces and vice versa. As $\varepsilon \to 1$, the newly calculated CD4+ count stays the same. However, when $\varepsilon > 1$, the production of the new CD4+ count is restricted by the value of $\varepsilon$. If the value of $\varepsilon < 1$, the production of the new CD4+ cells will decelerate according to the value of $\varepsilon$.

In Equation 4, the overall system change (changes in number of CD4+ cells and viral load) is calculated as a ratio between viral elimination and viral production. The numerator in Equation 4 estimates: (1) accumulative effect of viral elimination and reduced effectiveness of viral load on the system ($\psi_1$). This is calculated using the rate of viral elimination on the system due to natural elimination, $\gamma$ and rate of reduced effectiveness $\tau$ over time, (2) The effect of viral production on the CD4+ cells and viral load ($\psi_2$), and (3) The effect of viral elimination due to therapy ($\delta$) and reduces effectiveness ($\tau$) on the viral load. All the above three numerator terms are affected by the rate ($\eta$) at which viral load(s) locate a CD4+ cell. The denominator in Equation 4 simulates an HIV progression fact (Perelson, 2001), that every infected CD4+ cell will on average produce one other infected cell. Hence, like Equation 3, the denominator $\left(\varepsilon_{2p} + v^{(n)} + \tau_{e}^{(n)} + \delta v^{(n)}\right)$ will accelerate or decelerate the production of T-cells and viral loads according to the overall possible viral loads that can be generated by the system. Equation (5) calculates the CD8+ count at next time step using the CD8+ production rate $\tau_1$ and the CD8+ elimination rate $\tau_2$. 

860