Gennady A. Bocharov, Ilya S. Telatnikov, Valery A. Chereshnev, Javier Martinez, and Andreas Meyerhans

Mathematical modelling of the within-host HIV quasispecies dynamics in response to antiviral treatment

Abstract: The aim of this work is the construction, calibration, and comparative analysis of mathematical models of the evolution of the human immunodeficiency virus (HIV) in the course of infection when the models are based on deterministic principles of the quasispecies theory (Eigen–Schuster) and on stochastic approaches of genetic algorithms (Holland). The models take into account the replication of viral genomes and selection of descendants according to their fitness, point mutations, multi-infection of target cells and recombination of genomes at the stage of formation of proviral DNA. The processes of diversification of the virus population under the action of the antiviral drug azidothymidine (AZT) that blocks reverse transcription of the virus are simulated. A four-letter alphabet is used in the stochastic model for description of nucleotide sequences. The parameters of the model are estimated using original data on the degree of adaptation of the HIV mutants that are partly or completely resistant to this drug. The influence of parameters of infection on the characteristics of viral mutants population diversity is studied.

Keywords: Mathematical model, virus infection, quasispecies, genetic algorithm, human immunodeficiency virus.

MSC 2010: 49J15, 49N90, 95B05, 92C50

The central problem of mathematical modelling in immunology is the description of the immune system and outcome predictions of interactions between pathogens and the human body. The variety of cources and outcomes of infection depends on the dynamics of interaction of infecting pathogens and the immune system [29]. The fundamental basis of mathematical modelling in immunology was laid by G. I. Marchuk [19]. A significant experience in construction and application of mathematical models of the dynamics of immune system reactions in response to the reproduction of viruses or bacteria is accumulated nowadays [19, 21]. An infection, caused by the human immunodeficiency virus of first type 1 (HIV) is the object of intensive interdisciplinary research based on the methods of mathematical modelling aimed to study the mechanisms of pathogenesis and opportunities of a complete cure [6, 9].

One of the problems related to understanding the dynamics of the HIV infection is associated with a rapid variability of viruses and the selection of better adapted mutants in the context of selective pressure from the immune system and used antiviral drugs. A significant genetic diversification of the virus population appears in the course of infection, which leads to the need to use the concept of 'quasispecies' [4], i.e., an ensemble of viral RNA sequences differing from each other in several bases in populations of closely related, but not

Gennady A. Bocharov: Institute of Numerical Mathematics of the RAS, Moscow 119333, Russia. E-mail: bocharov@inm.ras.ru
Ilya S. Telatnikov: Southern Scientific Center of Russian Academy of Sciences, Rostov-on-Done 344006, Russia. E-mail: iluxt@list.ru
Valery A. Chereshnev: Institute of Immunology and Physiology, Ural Branch of the RAS, Ekaterinburg 620041, Russia. Ural Federal University, Ekaterinburg 620002, Russia. E-mail: chereshnev@duma.gov.ru
Javier Martinez, Andreas Meyerhans: Infection Biology Group, Department of Experimental and Health Sciences, University Pompeu Fabra, Barcelona 08003, Spain. E-mail: javier.martinez@upf.edu
Andreas Meyerhans: Institució Catalana de Recerca i Estudis Avancats (ICREA), Barcelona, Spain.
E-mail: andreas.meyerhans@upf.edu

identical genomes of HIV. The mathematical formalization of the theory of quasispecies was carried out by Eigen and Schuster [12–14].

Since the variability of HIV is a key factor in the pathogenesis of the disease, the construction of biologically meaningful models of the HIV infection necessarily requires the development of efficient methods for description of the heterogeneity and prediction of viral populations properties, namely, the antiviral resistance and escape from immune response. By now there are two approaches to modelling HIV variability, these are the deterministic approach based on application of systems of ordinary differential equations (ODE) [21] or partial differential equations (PDEs) [25] and the stochastic one based on genetic algorithms [5, 27, 28]. Each of them is characterized by its range of applicability and implementation details. Within the framework of the theory of quasispecies using the ODE model, one can efficiently study the effect of mutations on the genetic evolution of virus populations [18, 21]. To study the process of viral replication during HIV infection, which has a number of features including the multi-infection of cells and recombination of genomes, the use of approaches based on genetic algorithms appears to be quite appropriate.

Previously, we have constructed a stochastic model [5] of HIV variability on the base of representation of viral genomes as vectors (sequences or chains) of nucleotides of length $L = 100$ encoded using a binary alphabet. The aim of this paper is the construction, calibration, and comparative analysis of the deterministic and stochastic models of HIV evolution during the infection taking into account point mutations, recombination, replication of viral genomes, and selection of descendants according to their fitness. We simulate processes of diversification of the virus population under the action of the antiviral azidothymidine (AZT) drug which blocks the reverse transcription of viruses. To do that, we use original data for fitness values of the HIV mutants partly or completely resistant to the drug. It should be noted that we propose such a level of detail for description of the HIV mutants diversification based on real data that is absent in current attempts to modelling the HIV evolution [1, 6, 21, 27, 28].

Section 2 describes the specific aspects of HIV replication. In Section 3, we formulate the ODE model of dynamics of the HIV mutants resistant to antiviral drugs taking into account mutation, recombination, and the possibility of double infection of cells. Section 4 presents the construction of stochastic model based on genetic algorithms and taking into account mutations, recombinations under the conditions of multi-infection of target cells. In conclusion we discuss two approaches to modelling the HIV evolution and their further development.

1 HIV replication characteristics

HIV belongs to the family of retroviruses [9]. The genome of the virus consists of two single-stranded RNA molecules each of which contains about 10^4 nucleotides specifying the 4-letter alphabet A, C, G, U. The evolutionary process is caused by the action of random factors at certain stages of viral replication, first of all, at reverse transcription of viral RNA into proviral DNA which is integrated into the genome of infected target cells (macrophages, CD4 T lymphocytes). This process can be blocked using inhibitors of reverse transcription. These also include the drug AZT whose action is considered in this paper. Key characteristics of the process of the HIV replication were systematized in [6], in this case the rate of virus production by infected cells in the body is about 10^9 particles per day, with ~ 0.25 point mutations per genome in a single replication cycle. In the case of multi-infection, i.e., when the number of proviral DNA integrated into the cellular genome is greater than one, an intense recombination process takes place with the rate of $3 - 30$ times per genome in a replication cycle.

The high genetic variability of HIV and the selection of the most adapted mutants determine the ability of a population to escape from the immune response and develop a resistance to the antiviral therapy. The theoretical basis for the study of mutation-selection processes may be formed by the concept of quasispecies proposed by Eigen and Schuster [12–14] for the description of genome evolution defined by replicating polynucleotide sequences. The model of quasispecies dynamics considers the set (population) $\{\mathbf{S}_k\}$ of sequences $\mathbf{S}_1, \ldots, \mathbf{S}_n$ represented as vectors $\mathbf{S}_k = \{S_{k1}, \ldots, S_{kN}\}$, $k = 1, \ldots, n$, with the components

taking a specified number v of distinct values determined by the coding alphabet. For example, we may consider the alphabet with two elements $(v = 2,\ S_{kj} \in \{-1, 1\},\ j = 1, \ldots, N)$, or the alphabet with four letters $(v = 4,\ S_{kj} \in \{A,C,G,T\})$. It is supposed that the length N of the sequence and the size n of the population are large, i.e., $n, N \gg 1$. A variation of the population is realized as the result of interaction of the replication, mutation, and selection on the basis of fitness values. The character of evolution essentially depends on the size n of the population. If n is large, then the evolution dynamics of the population can be described by an ODE system. Canonical equations of the theory of quasispecies have the following form in the mathematical virology [21]:

$$\frac{d\mathbf{v}}{dt} = \mathbf{W}\mathbf{v} - d(\mathbf{v})\mathbf{v}, \qquad \mathbf{W} = \begin{pmatrix} a_1 Q_{11} & a_2 Q_{21} & \cdots & a_n Q_{n1} \\ a_1 Q_{12} & a_2 Q_{22} & \cdots & a_n Q_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 Q_{1n} & a_2 Q_{2n} & \cdots & a_n Q_{nn} \end{pmatrix}, \qquad d(\mathbf{v}) = \frac{\sum\limits_{i=1}^{n} a_i v_i}{\sum\limits_{i=1}^{n} v_i}. \tag{1.1}$$

The components of the vector $\mathbf{v}$ represent the number of genomes composing the population $\mathbf{v} = \{v_1, v_1, \ldots, v_n\}$; a_i is the rate of replication of the ith quasispecies, $i = 1, \ldots, n$; $\mathbf{Q} = (Q_{ij})$ is the $n \times n$ mutation matrix, i.e., the matrix of probabilities that a mutant of ith type becomes a mutant of jth type as the result of the mutation process. The second term in the right-hand side of equations (1.1) is introduced to limit the total size of the population. The solution to system 1.1 satisfies the following assertions: (1) the total number of quasispecies is constant, $\sum_{i=1}^{n} v_i = \text{const}$; (2) the equilibrium distribution is determined by the eigenvector corresponding to the maximal eigenvalue of the matrix $\mathbf{W}$.

The classic model of quasispecies does not take into account the action of recombination processes occurring during HIV infection. Below we propose one of possible variants of generalizations of model (1.1) to describe the influence of genome recombination in the case when chromosomes of an infected cell contain two proviral genomes. Developing our previous approach [5] and considering the fragment of HIV DNA determining the AZT-resistance, we use a four-letter alphabet corresponding to the set of nucleotide, i.e., A is for deoxyadenosine, G is for deoxyguanosine, T is for thymidine, C is for deoxycytidine. Methionine and threonine are at the 41st and 215th positions of the aminoacid sequence of the wild-type virus. Therefore, the initial virus population positions 121–123 contain a symbol combination ATG, and the positions 643–645 contain the combination of symbols ACC. We consider the following subpopulation of mutants completely or partly resistant to AZT; $v_i(t)$ is the size of the ith quasispecies in the virus population in the body, where $i = 1$ corresponds to the wild type (WT, when the positions 121–123 of the nucleotide chain contain the combination ATG, and the positions 643–645 contain the combination ACC), $i = 2$ corresponds to the mutant M41L (the same positions contain TTG and ACC, respectively), $i = 3$ is for T215N (ATG and AAC), $i = 4$ is for T215S (ATG and TCC), $i = 5$ is for T215Y (ATG and TAC), $i = 6$ is for M41L+T215N (TTG and AAC), $i = 7$ is for M41L+T215S (TTG and TCC), $i = 8$ is for M41L+T215Y (TTG and TAC).

The rate of mutation was estimated with the use of experimental data for the rate of mutations of specific nucleotides [10]. According to these data, the rate of mutation A → C is 0.543; A → G is 2.277; A → T is 0.319; C → A is 1.858; C → G is 0.228; C → T is 4.601; G → A is 7.361; G → C is 0.262; G → T is 0.655; T → A is 0.810; T → C is 2.809; T → G is 0.580. The selection of mutants in the new population after the replication cycle depends on the value of the fitness function calculated experimentally [20] for various doses of AZT, {0, 0.03, 0.3, 2, 5, 10 μM}, presented in Table 1.

## 2 Deterministic model of HIV quasispecies dynamics

We generalize the quasispecies model by enabling the recombination process in the case of single and double infection of target cells, which corresponds to clinical data [26]. According to this approach, we implemented the following simple one-point recombination operator: the point of transition from one chain to another is chosen randomly for two chains. Both chains are divided into two parts and exchange blocks of nucleotides.

Table 1. Fitness values for AZT-resistant mutants.

Concentration of AZT μM	Wild type WT	Mutation M41L	Mutation T215N	Mutation T215S	Mutation T215Y	Mutation M41L+T215N	Mutation M41L+T215S	Mutation M41L+T215Y
0	1	0.604	0.203	0.253	0.704	0.291	0.493	0.782
0.03	0.560	0.509	0.0958	0.100	0.606	0.124	0.393	0.617
0.3	0.162	0.166	0.012	0.016	0.293	0.038	0.116	0.385
2	0.043	0.024	0.0009	0.0038	0.116	0.020	0.009	0.187
5	0.0052	0.012	0	0	0.027	0.0011	0.0029	0.098
10	0	0.0011	0	0	0.011	0	0	0.020

Table 2. Transitions between mutants as the result of point mutations.

	WT	M41L	T215N	T215S	T215Y	M41L+T215N	M41L+T215S	M41L+T215Y
WT		A → T	C → A	A → T	2	2	2	3
M41L	T → A	0	2	2	3	C → A	A → T	2
T215N	A → C	2	0	2	A → T	A → T	3	2
T215S	T → A	2	2	0	C → A	3	A → T	2
T215Y	2	3	T → A	A → T	0	2	2	A → T
M41L+T215N	2	A → C	T → A	3	2	0	2	A → T
M41L+T215S	2	T → A	3	T → A	2	2	0	C → A
M41L+T215Y	3	2	2	2	T → A	T → A	A → C	0
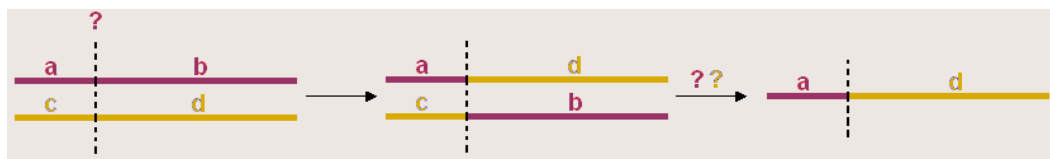
Figure 1. Single recombination scheme implemented in the ODE model of quasispecies dynamics.

Thus, the process of reverse transcription in the model involves a crosswise transfer of nucleotides from one chain of RNA to another (see Fig. 1).

We analyze the parameters characterizing the transitions between the considered variant of mutants. The mutations relevant for consideration in the model and leading to a change in sizes of the corresponding subpopulation of mutants completely or partly resistant to AZT are presented in Table 2. The numbers denote situations when the number of necessary mutations differs from 1.

Following [5], we estimate the rates of transition Q_{ij} of mutants of one type to another in the following way. If we assume that mutations of nucleotides are equiprobable, then the probability of replacement of one base at a particular position to another is equal to $p = 2.5 \cdot 10^{-5}$. Since we use the four-letter code (A, C, G, T), then each base can be replaced by one of three others. The probability that none of coding bases is changed at any position (41st or 215th) is equal to $Q_{ii} = (1 - p)^6$. In this case the probability of mutation in one of the six considered bases is equal to $Q_{ij} = p (1 - p)^5$. For the chosen value of p we have $Q_{ii} = 0.9999$, $Q_{ij} \approx 2.5 \cdot 10^{-5}, i \neq j$). The probability of transition from one type of mutant to another is assumed to be zero if this requires not less than two mutations. As was shown in [9], the mutations between bases presented above are not equiprobable. The probabilities of transition between nucleotide bases quantified experimentally and calculated on the base of roulette principle (see Fig. 4) have the following characteristics: the transition A→T occurs with the probability $p_1 = 0.1$; A→C with $p_2 = 0.17$; T→A with $p_3 = 0.19$; C→A with $p_4 = 0.28$. Taking into account these estimates, the coefficients Q_{ij} characterizing in the model the rates of transition from one type of mutant to another are additionally scaled, i.e., multiplied by the corresponding probabilities p_k, $k = 1, \ldots, 4$, of specific transition of one base to another.

In the case of double infection of target cells one has to take into account recombination caused transitions between the considered types of mutants. Variants of formation of various pairs of virus genomes in

Table 3. Transitions between mutants due to recombination in double infection.

type	1	2	3	4	5	6	7	8
1	1	1,2	1,3	1,4	1,5; 3,4	1,6; 3,2	1,7; 2,4	1,8; 2,5; 7,3
2	1,2	2	2,3; 6,1	2,4; 7,1	2,5; 8,1; 6,4	2,6	2,7	2,8; 6,7
3	1,3	2,3; 6,1	3	3,4; 1,5	3,5	3,6	3,7; 4,6; 1,8	3,8; 5,6
4	1,4	2,4; 7,1	3,4; 1,5	4	4,5	4,6; 3,7; 5,2	4,7	4,8; 5,7
5	1,5; 3,4	2,5; 8,1; 6,4	3,5	4,5	5	5,6; 3,8	5,7; 4,8	5,8
6	1,6; 3,2	2,6	3,6	4,6; 3,7; 5,2	5,6; 3,8	6	6,7; 2,8	6,8
7	1,7; 2,4	2,7	3,7; 4,6; 1,8	4,7	5,7; 4,8	6,7; 2,8	7	7,8
8	1,8; 2,5; 7,3	2,8; 6,7	3,8; 5,6	4,8; 5,7	5,8	6,8	7,8	8

double infection are presented in Table 3, where the ordinal number of the mutant is given instead of its description.

In order to describe the recombination probability for a pair of chains (i, j) for the chain length of 1800 nucleotides, we introduce the notation z_{kl}^{ij}, where the superscript indicates which chains participate in recombination, the subscript indicates the chains generated in this process. These probabilities can be estimated in the following way. We denote the recombination probability by z and assume that the breaks of chain are equiprobable at any point. To calculate the recombination probability for double infection, we have to multiply z (in the model considered here the recombination probability for a particular pair of chains is $z = 10^{-8}$) by the probability of double infection q (in the model we have $q = 0.2$). In this case the probability of breaking at a given position equals zql^{-1}, where $l + 1$ is the length of the chain. Below we assume that the length of the chain is $l + 1 = 1800$. The probability that the crossover occurs at one of the first 120 positions is $120zql^{-1}$, the probability of breaking at one of $l - 645$ last positions is $(l - 645)zql^{-1}$ (for the chain length of 1800 nucleotides we have $1154zql^{-1}$), the probability of breaking between the 123rd and 643rd positions is $519zql^{-1}$. If the length of a chain differs from 1800, only the probability of breaking after the 645th position is changed. We note that the recombination of chains of one type results in formation of two chains of the same type with the probability z. If we take into account the feasibility of double infection by virions of identical type, then the recombination may produce different types with the following probabilities: for the pair (1,1) this is $z_{11}^{11} = z$; for (1,2) and (2,1) – $z_{12}^{12} = zq$; for (1,3) and (3,1) – $z_{13}^{13} = zq$; for (1,4) and (4,1) – $z_{14}^{14} = zq$; for (1,5) and (5,1) – $z_{15}^{15} = 1789zql^{-1}$ and $z_{34}^{15} = zql^{-1}$; for (1,6) and (6,1) – $z_{16}^{16} = 1276zql^{-1}$ and $z_{23}^{16} = 523zql^{-1}$; for (1,7) and (7,1) – $z_{17}^{17} = 1277zql^{-1}$ and $z_{24}^{17} = 522zql^{-1}$; for (1,8) and (8,1) – $z_{18}^{18} = 1276zql^{-1}$, $z_{25}^{18} = 522zql^{-1}$, and $z_{37}^{18} = zql^{-1}$; for (2,2) – $z_{22}^{22} = z$; for (2,3) and (3,2) – $z_{23}^{23} = 1279zql^{-1}$ and $z_{61}^{23} = 521zql^{-1}$; for (2,4) and (4,2) – $z_{24}^{24} = 1279zql^{-1}$ and $z_{71}^{24} = 521zql^{-1}$; for (2,5) and (5,2) – $z_{25}^{25} = 1276zql^{-1}$, $z_{18}^{25} = 522zql^{-1}$, and $z_{46}^{25} = zql^{-1}$; for (2,6) and (6,2) – $z_{26}^{26} = zq$; for (2,7) and (7,2) – $z_{27}^{27} = zq$; for (2,8) and (8,2) – $z_{28}^{28} = 1279zql^{-1}$ and $z_{67}^{28} = 521zql^{-1}$; for (3,3) – $z_{33}^{33} = z$; for (3,4) and (4,3) – $z_{34}^{34} = 1789zql^{-1}$ and $z_{15}^{34} = zql^{-1}$; for (3,5) and (5,3) – $z_{35}^{35} = zq$; for (3,6) and (6,3) – $z_{36}^{36} = zq$; for (3,7) and (7,3) – $z_{37}^{37} = 1276zql^{-1}$, $z_{46}^{37} = 522zql^{-1}$ and $z_{18}^{37} = zql^{-1}$; for (3,8) and (8,3) – $z_{38}^{38} = 1277zql^{-1}$ and $z_{56}^{38} = 522zql^{-1}$; for (4,4) – $z_{44}^{44} = z$; for (4,5) and (5,4) – $z_{45}^{45} = zq$; for (4,6) and (6,4) – $z_{46}^{46} = 1276zql^{-1}$, $z_{37}^{46} = 522zql^{-1}$ and $z_{25}^{46} = zql^{-1}$; for (4,7) and (7,4) – $z_{47}^{47} = zq$; for (4,8) and (8,4) – $z_{48}^{48} = 1276zql^{-1}$ and $z_{57}^{48} = 523zql^{-1}$; for (5,5) – $z_{55}^{55} = z$; for (5,6) and (6,5) – $z_{56}^{56} = 1277zql^{-1}$ and $z_{38}^{56} = 522zql^{-1}$; for (5,7) and (7,5) – $z_{57}^{57} = 1276zql^{-1}$ and $z_{48}^{57} = 523zql^{-1}$; for (5,8) and (8,5) – $z_{58}^{58} = zq$; for (6,6) – $z_{66}^{66} = z$; for (6,7) and (7,6) – $z_{67}^{67} = 1789zql^{-1}$ and $z_{28}^{67} = zql^{-1}$; for (6,8) and (8,6) – $z_{68}^{68} = zq$; for (7,7) – $z_{77}^{77} = z$; for (7,8) and (8,7) – $z_{78}^{78} = zq$; for (8,8) – $z_{88}^{88} = z$.

Finally, we formulate a system of equations of the model of population dynamics of mutants resistant to AZT on the basis of the quasispecies model where along with replication, loss, point mutations, and selection of mutants we also take into account the recombination of genomes. It is assumed that the cells infected with different mutants produce each of them in equal amounts. For simplicity of parameterization of the influence of recombination on the production-elimination of mutants, we assume that the corresponding rate linearly depends on their quantity. The generalized version of the system describing the population dynamics

of quasispecies takes the following form:

$$\dot{v}_1 = a_1 Q_{11} v_1 + a_2 p_1 Q_{21} v_2 + a_3 p_4 Q_{31} v_3 + a_4 p_1 Q_{41} v_4 - v_1 d(\mathbf{v}) + R_1(\mathbf{v})$$
$$\dot{v}_2 = a_1 p_3 Q_{12} v_1 + a_2 Q_{22} v_2 + a_6 p_4 Q_{62} v_6 + a_7 p_1 Q_{72} v_7 - v_2 d(\mathbf{v}) + R_2(\mathbf{v})$$
$$\dot{v}_3 = a_1 p_2 Q_{13} v_1 + a_3 Q_{33} v_3 + a_5 p_1 Q_{53} v_5 + a_6 p_1 Q_{63} v_6 - v_3 d(\mathbf{v}) + R_3(\mathbf{v})$$
$$\dot{v}_4 = a_1 p_3 Q_{14} v_1 + a_4 Q_{44} v_4 + a_5 p_4 Q_{54} v_5 + a_7 p_1 Q_{74} v_7 - v_4 d(\mathbf{v}) + R_4(\mathbf{v})$$
$$\dot{v}_5 = a_3 p_3 Q_{35} v_3 + a_4 p_2 Q_{45} v_4 + a_5 Q_{55} v_5 + a_8 p_1 Q_{85} v_8 - v_5 d(\mathbf{v}) + R_5(\mathbf{v}) \tag{2.1}$$
$$\dot{v}_6 = a_2 p_2 Q_{26} v_2 + a_3 p_3 Q_{36} v_3 + a_6 Q_{66} v_6 + a_8 p_1 Q_{86} v_8 - v_6 d(\mathbf{v}) + R_6(\mathbf{v})$$
$$\dot{v}_7 = a_2 p_3 Q_{27} v_2 + a_4 p_3 Q_{47} v_4 + a_7 Q_{77} v_7 + a_8 p_4 Q_{87} v_8 - v_7 d(\mathbf{v}) + R_7(\mathbf{v})$$
$$\dot{v}_8 = a_5 p_3 Q_{58} v_5 + a_6 p_3 Q_{68} v_6 + a_7 p_2 Q_{78} v_7 + a_8 Q_{88} v_8 - v_8 d(\mathbf{v}) + R_8(\mathbf{v}).$$

In system (2.1) we have $d(\mathbf{v}) = \sum_{i=1}^{8} a_i v_i / \sum_{i=1}^{8} v_i$, the values of the coefficients a_i are taken from experimental data on the fitness of mutants from Table 1, the functions $R_i(\mathbf{v})$, $i = 1, \ldots, 8$, have the form

$$R_1(\mathbf{v}) = a_1 z_{11}^{11} v_1 + a_2 z_{12}^{12} v_2 + a_3 z_{13}^{13} v_3 + a_4 z_{14}^{14} v_4 + a_5 z_{15}^{15} v_5 + a_6 z_{16}^{16} v_6 + a_7 z_{17}^{17} v_7 + a_8 z_{18}^{18} v_8$$
$$+ \frac{1}{2} z_{15}^{34}(a_3 v_3 + a_4 v_4) + \frac{1}{2} z_{18}^{25}(a_2 v_2 + a_5 v_5) + \frac{1}{2} z_{17}^{24}(a_2 v_2 + a_4 v_4) + \frac{1}{2} z_{18}^{37}(a_3 v_3 + a_7 v_7)$$
$$+ \frac{1}{2} z_{18}^{23}(a_2 v_2 + a_3 v_3) - a_1 v_1\left(z_{34}^{15} + z_{23}^{16} + z_{24}^{17} + z_{25}^{18} + z_{37}^{18}\right)$$
$$R_2(\mathbf{v}) = a_1 z_{12}^{12} v_1 + a_2 z_{22}^{22} v_2 + a_3 z_{23}^{23} v_3 + a_4 z_{24}^{24} v_4 + a_5 z_{25}^{25} v_5 + a_6 z_{26}^{26} v_6 + a_7 z_{27}^{27} v_7 + a_8 z_{28}^{28} v_8$$
$$+ \frac{1}{2} z_{23}^{16}(a_1 v_1 + a_6 v_6) + \frac{1}{2} z_{24}^{17}(a_1 v_1 + a_7 v_7) + \frac{1}{2} z_{25}^{18}(a_1 v_1 + a_8 v_8) + \frac{1}{2} z_{25}^{46}(a_4 v_4 + a_6 v_6)$$
$$+ \frac{1}{2} z_{28}^{67}(a_6 v_6 + a_7 v_7) - a_2 v_2\left(z_{16}^{23} + z_{17}^{24} + z_{18}^{25} + z_{46}^{25} + z_{67}^{28}\right)$$
$$R_3(\mathbf{v}) = a_1 z_{13}^{13} v_1 + a_2 z_{23}^{23} v_2 + a_3 z_{33}^{33} v_3 + a_4 z_{34}^{34} v_4 + a_5 z_{35}^{35} v_5 + a_6 z_{36}^{36} v_6 + + a_7 z_{37}^{37} v_7 + a_8 z_{38}^{38} v_8$$
$$+ \frac{1}{2} z_{23}^{16}(a_1 v_1 + a_6 v_6) + \frac{1}{2} z_{34}^{15}(a_1 v_1 + a_5 v_5) + \frac{1}{2} z_{37}^{46}(a_4 v_4 + a_6 v_6) + \frac{1}{2} z_{37}^{18}(a_1 v_1 + a_8 v_8)$$
$$+ \frac{1}{2} z_{38}^{56}(a_5 v_5 + a_6 v_6) - a_3 v_3\left(z_{16}^{23} + z_{15}^{34} + z_{46}^{37} + z_{18}^{37} + z_{56}^{38}\right)$$
$$R_4(\mathbf{v}) = a_1 z_{14}^{14} v_1 + a_2 z_{24}^{24} v_2 + a_3 z_{34}^{34} v_3 + a_4 z_{44}^{44} v_4 + a_5 z_{45}^{45} v_5 + a_6 z_{46}^{46} v_6 + a_7 z_{47}^{47} v_7 + a_8 z_{48}^{48} v_8$$
$$+ \frac{1}{2} z_{24}^{17}(a_1 v_1 + a_7 v_7) + \frac{1}{2} z_{34}^{15}(a_1 v_1 + a_5 v_5) + \frac{1}{2} z_{46}^{25}(a_2 v_2 + a_5 v_5) + \frac{1}{2} z_{46}^{37}(a_3 v_3 + a_7 v_7)$$
$$+ \frac{1}{2} z_{48}^{57}(a_5 v_5 + a_7 v_7) - a_4 v_4\left(z_{17}^{24} + z_{15}^{34} + z_{37}^{46} + z_{25}^{46} + z_{57}^{48}\right)$$
$$R_5(\mathbf{v}) = a_1 z_{15}^{15} v_1 + a_2 z_{25}^{25} v_2 + a_3 z_{35}^{35} v_3 + a_4 z_{45}^{45} v_4 + a_5 z_{55}^{55} v_5 + a_6 z_{56}^{56} v_6 + a_7 z_{57}^{57} v_7 + a_8 z_{58}^{58} v_8$$
$$+ \frac{1}{2} z_{15}^{34}(a_3 v_3 + a_4 v_4) + \frac{1}{2} z_{25}^{18}(a_1 v_1 + a_8 v_8) + \frac{1}{2} z_{25}^{46}(a_4 v_4 + a_6 v_6) + \frac{1}{2} z_{57}^{48}(a_4 v_4 + a_8 v_8)$$
$$+ \frac{1}{2} z_{56}^{38}(a_3 v_3 + a_8 v_8) - a_5 v_5\left(z_{34}^{15} + z_{18}^{25} + z_{46}^{25} + z_{38}^{56} + z_{48}^{57}\right)$$
$$R_6(\mathbf{v}) = a_1 z_{16}^{16} v_1 + a_2 z_{26}^{26} v_2 + a_3 z_{36}^{36} v_3 + a_4 z_{46}^{46} v_4 + a_5 z_{56}^{56} v_5 + a_6 z_{66}^{66} v_6 + a_7 z_{67}^{67} v_7 + a_8 z_{68}^{68} v_8$$
$$+ \frac{1}{2} z_{16}^{23}(a_2 v_2 + a_3 v_3) + \frac{1}{2} z_{46}^{25}(a_2 v_2 + a_5 v_5) + \frac{1}{2} z_{46}^{37}(a_3 v_3 + a_7 v_7) + \frac{1}{2} z_{56}^{38}(a_3 v_3 + a_8 v_8)$$
$$+ \frac{1}{2} z_{67}^{28}(a_2 v_2 + a_8 v_8) - a_6 v_6\left(z_{23}^{16} + z_{37}^{46} + z_{52}^{46} + z_{38}^{56} + z_{28}^{67}\right)$$
$$R_7(\mathbf{v}) = a_1 z_{17}^{17} v_1 + a_2 z_{27}^{27} v_2 + a_3 z_{37}^{37} v_3 + a_4 z_{47}^{47} v_4 + a_5 z_{57}^{57} v_5 + a_6 z_{67}^{67} v_6 + a_7 z_{77}^{77} v_7 + a_8 z_{78}^{78} v_8$$
$$+ \frac{1}{2} z_{37}^{18}(a_1 v_1 + a_8 v_8) + \frac{1}{2} z_{17}^{24}(a_2 v_2 + a_4 v_4) + \frac{1}{2} z_{37}^{46}(a_4 v_4 + a_6 v_6) + \frac{1}{2} z_{57}^{48}(a_4 v_4 + a_8 v_8)$$
$$+ \frac{1}{2} z_{67}^{28}(a_2 v_2 + a_8 v_8) - a_7 v_7\left(z_{24}^{17} + z_{46}^{37} + z_{18}^{37} + z_{48}^{57} + z_{28}^{67}\right)$$

$$R_8(\mathbf{v}) = a_1 z_{18}^{18} v_1 + a_2 z_{28}^{28} v_2 + a_3 z_{38}^{38} v_3 + a_4 z_{48}^{48} v_4 + a_5 z_{58}^{58} v_5 + a_6 z_{68}^{68} v_6 + a_7 z_{78}^{78} v_7 + a_8 z_{88}^{88} v_8$$
$$+ \frac{1}{2} z_{18}^{37} (a_3 v_3 + a_7 v_7) + \frac{1}{2} z_{18}^{25} (a_2 v_2 + a_5 v_5) + \frac{1}{2} z_{28}^{67} (a_6 v_6 + a_7 v_7) + \frac{1}{2} z_{38}^{56} (a_5 v_5 + a_6 v_6)$$
$$+ \frac{1}{2} z_{48}^{57} (a_5 v_5 + a_7 v_7) - a_8 v_8 \left(z_{25}^{18} + z_{37}^{18} + z_{67}^{28} + z_{56}^{38} + z_{57}^{48} \right).$$

System (2.1) was solved numerically with the use of ode23s algorithm of MATLAB package. Figure 2 presents the solution to the system for the case when the total probability of recombination in reverse transcription is 0.5, the probability of double infection is ~ 0.2, and the total size of the population of virions is 10^4 particles.

The influence of the recombination process on the HIV quasispecies dynamics under absence of antiretroviral therapy is shown in Fig. 2. We can see the dominance of the wild-type viruses (WT); the number of all mutants resistant to AZT is less by 5–10 orders. The recombination increases the fraction of mutants with low fitness for given set of parameters. We note that the solution of the model reflects the sizes of HIV RNA population per ml of blood. The total size of the virus population in the body is within the range from 10^9 to 10^{12}, which allows us to estimate the total number of resistant mutants.

Figure 3 illustrates the dynamics of the amount of AZT-resistant quasispecies in the HIV population in the presence of an antiviral drug (the concentration equals $0.3\,\mu$M) in the case of an initially homogeneous population consisting of wild-type (a) and heterogeneous initial population containing the same number of all eight genotypes (b).

For this set of parameters and the method of description the recombination accelerates the appearance of resistant mutants. Regardless of the initial composition of the virus populations, the proportions of quasispecies tend to their stationary values and the qualitative effect of recombination results in a more intense mixing of genomes and increases the portion of poorly adapted quasispecies.

3 Stochastic model of HIV quasispecies dynamics

A generalization of the deterministic model for the dynamics of quasispecies by inclusion of the double infection and recombination process with reverse transcription of proviral RNA into the description of the phenomenon leads to a significant technical complication in the structure of the model. A more natural approach to modelling the variability of HIV is the construction of stochastic models on the basis of genetic algorithms (GA) [6] taking into account the discreteness of the number of mutants, structure of RNA nucleotide chains, and randomness in implementation of mutations and recombination. Genetic algorithms, first proposed by Holland [17], are widely used in evolution modelling and optimization [15, 16, 24]. The general scheme of a genetic algorithm is presented in Fig. 4.

Constructing a model based on GA, the population of virus genomes is specified by an ensemble of vectors of length N, $\{\mathbf{S}_k\}$, $k = 1, \dots, n$, the components of those vectors take one of four values, $S_{kj} \in \{A,C,G,T\}$, and each of them is characterized by the fitness value $f(\mathbf{S}_k)$. The implementation of the probabilistic selection procedure based on fitness values uses the roulette principle [24]. For each generation we mark n sectors on the roulette (Fig. 4b)), the part of the kth sector in the total area of the circle is assumed to be equal to $q_k = f_k \left(\sum_{j=1}^{n} f_j \right)^{-1}$, where $f_k = f(\mathbf{S}_k)$. Further, the roulette is spun n times specifying each time the sector number where the arrow stops. The individual corresponding to that number is selected in the population for the next generation. Thus, individuals are selected to the population of new generation with the probabilities q_k proportional to their fitness values f_k.

According to the results of [26], the GA model implements single, double, and triple infections of the target cells. The generation of recombinant genomes is performed in accordance with the scheme presented in Fig. 5. In order to describe qualitatively the degree of genetic inhomogeneity of the viral population, we define the mean Hamming distance for the population in the following way:

$$D = \frac{2}{n(n-1)} \sum_{k=1}^{n} \sum_{j<k}^{n} \sum_{i=1}^{N} d(S_{ki}, S_{ji})$$
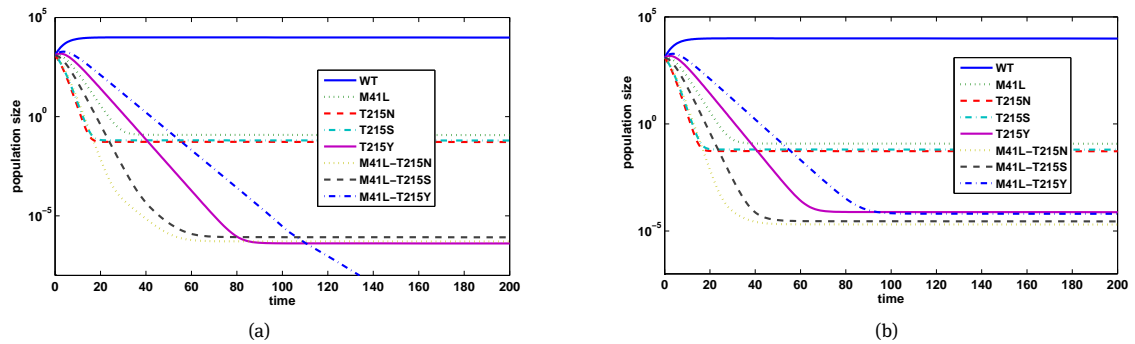
(a)　　　　　　　　　　　　　　(b)

Figure 2. The variation of the number of the HIV quasispecies for equal initial number of quasispecies without AZT in a double infection in the case of diversification; (a) under the action of point mutations, (b) subject to recombination.
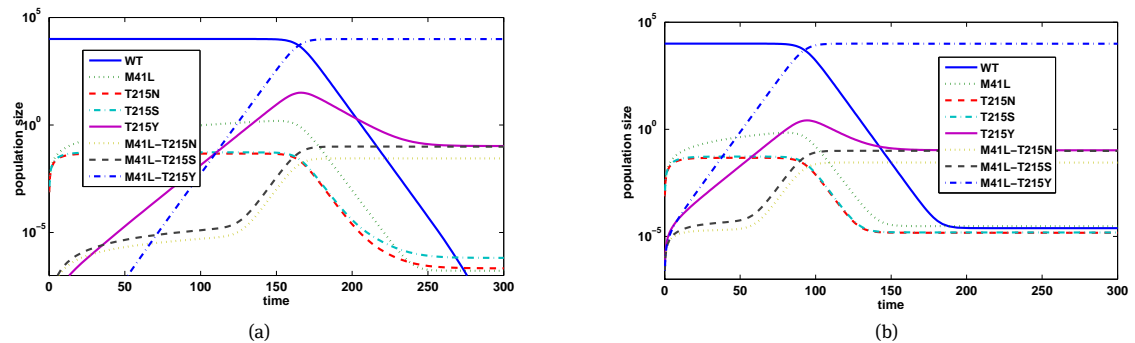

(a)　　　　　　　　　　　　　　(b)

Figure 3. The dynamics of the number of AZT-resistant mutants for the drug concentration equal to 0.3 μM subject to mutations and recombination with double infection; (a) under the action of point mutations, (b) subject to mutation and recombination.
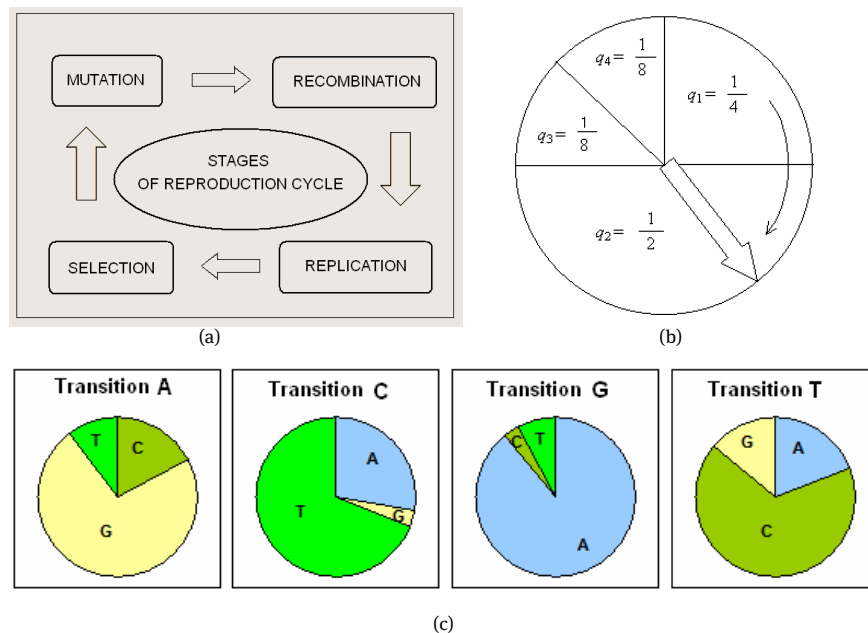

(a)　　　　　　　　　　　　　　(b)

(c)

Figure 4. Genetic algorithm; (a) scheme of genetic algorithm; (b) roulette principle [24] for calculation of descendant selection probability based on adaptation values. Example for an ensemble of 4 quasispecies; (c) roulette principle for calculation of base mutation probability according to the data from Table 1.
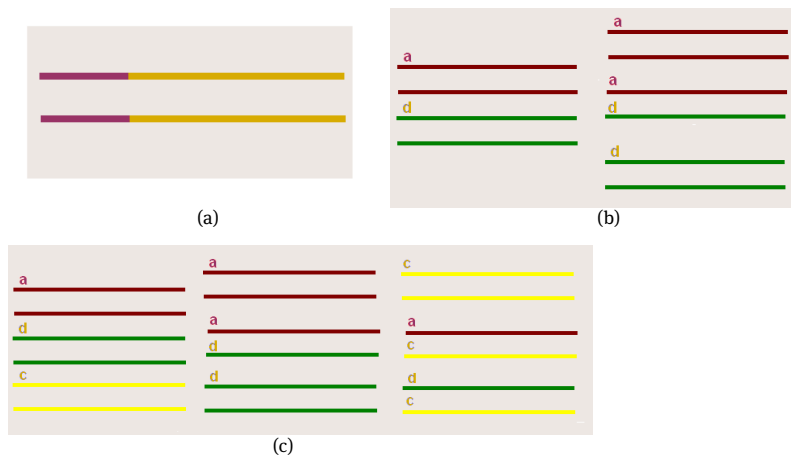
Figure 5. The results of recombination in multiply infected target cells; (a) single; (b) double; (c) triple.

where N is the length of genome, n is the size of population, and

$$d(S_{kl}, S_{km}) = \begin{cases} 0, & S_{kl} = S_{km} \\ 1, & S_{kl} \neq S_{km} . \end{cases}$$

The general structure of computation process can be represented as the following pseudocode implementing the corresponding operators of changing the size and structure of genome population.

1   Initialize: population ($t = 0$)
     do $t = 0, 1, t_{max}$
         $P(t) =$ population(t)
2        $P^{[2]}(t) =$ introduce mutations ($P(t)$)
3        $P^{[3]}(t) =$ apply recombination ($P^{[2]}(t)$)
4        $P^{[4]}(t) =$ perform replication ($P^{[3]}(t)$)
5        $P^{[5]}(t) =$ realize multi-infection ($P^{[4]}(t)$)
6        $P^{[6]}(t) =$ perform fitness based selections ($P^{[5]}(t)$)
7   Estimate the Hamming distance: D
8        $P^{[7]}(t) =$ update the population ($P^{[6]}(t)$)
9   population (t) $= P^{[7]}(t)$
     enddo

The initialization of the population (Stage 1) can be performed in different ways, e.g., by random generation of a virus RNA sequence that is copied after that up to the size of the population, by random generation of particular genomes of the population, etc. At Stage 2 (mutations) we test the probability of point mutation for each element of the 'population' array. In the case of success, we replace the nucleotide at a random position according to the roulette mechanism. At Stage 3 (recombination) we test the probability of recombination for the successive pairs of 'population' array elements. In the case of success we exchange the portions of nucleotides beginning with a random number. One of two recombinant chains are selected with equal probability. The RNA array $P^{[i]}(t)$ is updated. Further (determination of types) we check whether the elements of the RNA array belong to one or the other type of mutants. The elements of RNA types array (the number of an element in the RNA array) is assigned with the type of mutant. Implementing the processes of 'replication' and 'formation of virions', we specify the array where the virions are represented by the pairs of subsequent chains selected from the RNA pool. In the case of single infection the virion is composed from two identical chains, in the case of double and triple infection it is composed from all pairwise collections of two or three selected chains. In the case of multiple infection the index in the array of 'number of virions' is calculated to ensure the equality of the total amount to the coefficient of reproduction. The number of mutants of each

type is counted. In order to implement the 'selection' on the base of RNA types array and the array 'number of virions', we perform selection into the new population according to the fitness function for different dosing of antiviral preparation. In this case the virion containing chains with different mutations possesses the following properties: it belongs to two types of the corresponding mutants and has the fitness equal to the mean fitness value over all mutants.

A parallel version of the genetic algorithm is implemented taking into account the natural parallelism of the problem of realization of an ensemble of evolution processes, i.e., several populations are formed, each population is processed separately and is developed independently of others. The numerical implementation of the model is written in Visual Fortran and Open MPI. The key element of the stochastic model is the generator of random numbers. In this work we used an original generator of random numbers implemented in Fortran. Calls to this generator used system time as arguments, a certain linear transformation was applied to the obtained result.

Numerical calculations were performed for populations having the following parameters: the length of genome nucleotide chain was 1800; the size of genome population was 500 virions (1000 chains); the probability of point mutation was 0.2; the probability of recombination was 0.5 per genome in one replication cycle; the replication coefficient varied within a wide range in one replication cycle. The portion of single infected cells was 80%, double infected – 14%, triple infected – 6%. Although the calculation time for one replication cycle does not exceed one second on Intel CPU 2.2 GHz, the size of population and the length of genomes are critical parameters for calculations with the model. These values can be increased up to one order.

Below we present the calculation results for the scenario when the initial population contains equal number of chains (of mutants) of each type, i.e.,

1. chains of virions of the wild type (WT) contain a methionine (ATG) and threonine (ACC) at 45th and 215th positions, respectively,
2. in contrast with the wild type, the quasispecies M41L contains leucine (TTG) at 41st position,
3. the quasispecies T215N contains asparagine (AAC) at 215th position,
4. the quasispecies T215S contains serine (TCC) at 215th position,
5. the quasispecies T215Y contains tyrosine (TAC) at 215th position,
6. the quasispecies M41L+T215N differs from wild type at two positions, as well as two following ones
7. the quasispecies M41L+T215S,
8. the quasispecies M41L+T215Y.

The relative dynamics of HIV quasispecies subject to mutations and recombinations is shown in Figure 5 in the case of heterogeneous infection in the absence of therapeutic drug (AZT). The influence of the replication factor on the variation of quasispecies ratio in the HIV populations is of interest. Figures 6a–6c present the calculation results for the replication coefficients of 24, 600, and 6000 virions per cell in a replication cycle, respectively. It is seen that with increasing scale of replication the proportion of mutants is increased, while the subpopulation of wild type is reduced. We present portions of considered quasispecies averaged over 80 realizations (wild type and 7 mutants) for 10000 replication cycles (1 cycle takes about 1.5 days). The qualitative nature of the distribution of genotypes in the population is consistent with the results of the differential models taking into account the recombination (Fig. 2b), however, the differences in the sizes are significantly less. This argues for studying the correspondence between the parameters of GA and ODE models.

Figure 7 presents the comparative dynamics of the virus population for the replication coefficient equal to 24; (a) without antiretroviral preparation and (b) for the case when the concentration of AZT equals 0.3 μM. For the replication coefficient equal to 24, only the most adapted quasispecies (M41L+T215Y) remain in the virus population. If we reduce the replication coefficient down to 6 in the case of injection of the antiviral preparation (AZT concentration equals 0.3 μM), all the genomes, except for one (M41L+T215Y or T215Y), disappear during the first few replication cycles (graph is not shown). The growth of the replication coefficient also causes a growth of the stochasticity of population dynamics as is shown in Fig. 8. The curves presented in Figs. 8a and 8b illustrate the variation of relative sizes of quasispecies in the virus population averaged over 80 implementations for the replication coefficient equal to 600 and 6000, respectively, and for the AZT
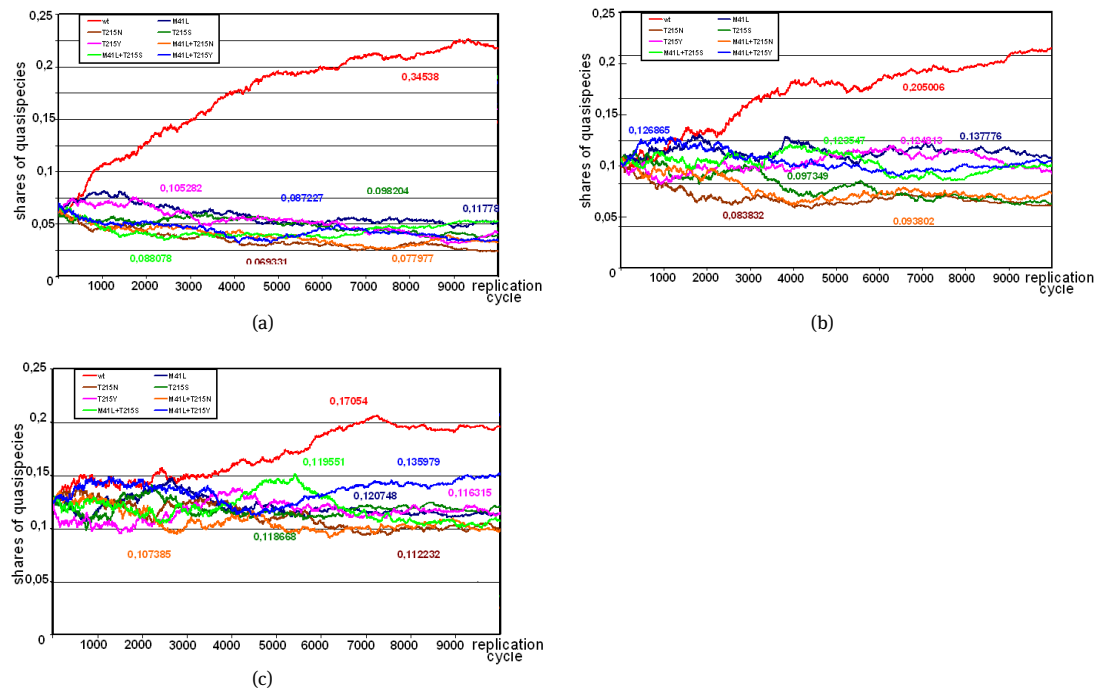
Figure 6. Averaged dynamics of quasispecies without AZT for different values of the replications coefficient; (a) 24; (b) 600; (c) 6000.
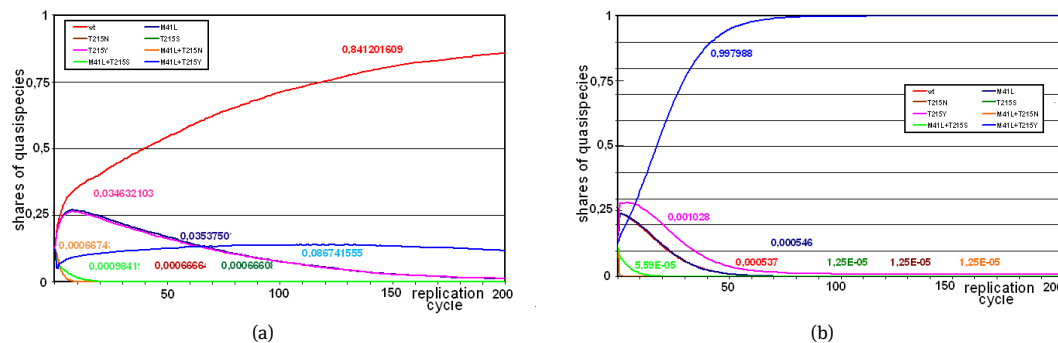

Figure 7. Dynamics of the HIV quasispecies; (a) without AZT; (b) for AZT concentration 0.3μ M. The coefficient of replication is 24 in both cases.
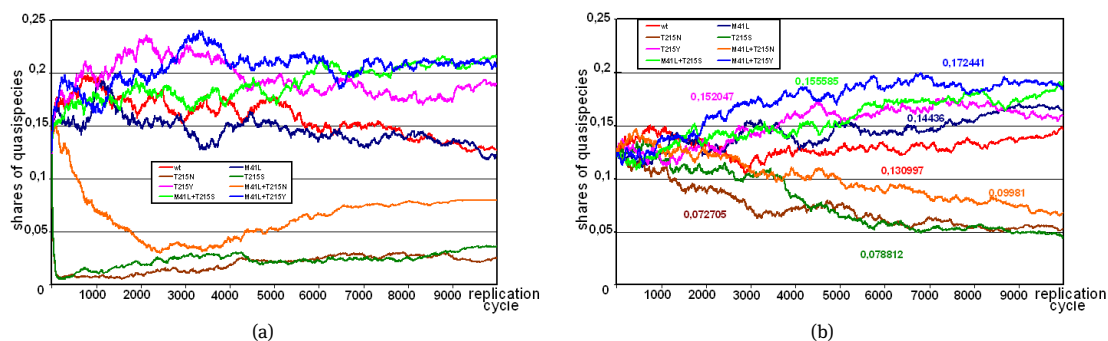

Figure 8. Dynamics of the HIV quasispecies resistant to AZT for the AZT concentration equal to 0.3μ M for the replication coefficient equal to (a) 600; (b) 6000.

concentration equal to $0.3\,\mu$M. It is seen that large population reproduction coefficients allow both better and weaker adapted mutants to survive.

Thus, the choice of the replication coefficient essentially affects the proportion of the HIV population quasispecies. For small values of the replication coefficient an initially heterogeneous population turns into a practically homogeneous one (the genotype characterized by the greatest value of the adaptation function remains). On the contrary, a large replication coefficient leads to a heterogeneous population with a less pronounced dominance of most fitness quasispecies.

4 Conclusions

In this paper we present two approaches to modelling the mutants population dynamics on the base of deterministic and stochastic description of the processes. Specifically we model the HIV quasispecies evolution under the action of selection caused by differences in fitness values of mutants partly or completely resistant to AZT.

A mathematical model of HIV diversification is constructed on the base of an ODE system of the quasispecies framework to study the genetic variability of HIV in the course of infection. Along with point mutations, the model describes recombination processes considering the possibility of double infection of cells. We note that a more biologically adequate approach to parameterization of the recombination process action on the appearance of mutants has to be based on consideration of linkage disequilibrium in the model, which leads to a bilinear dependence of birth–elimination processes of mutants as was shown in [5] for the model of two-locus two-allele mutants, or in a more general form in [2]. However, the practical implementation of this type of parameterization is restricted by significant complication of the ODE model of quasispecies dynamics and requires further studies with the use of tensor products and construction of small-parametric representations [22, 23] whose great potential in the problems of mathematical immunology was pointed out by G. I. Marchuk.

Further, a model of the virus population evolution taking into account random point mutations, recombination, replication of viral genomes, and selection of progeny according to the degree of adaptation is constructed on the base of genetic algorithms. A four-letter alphabet of nucleotide encoding is used. Single- and multi-infection of target cells are implemented in the model. The calibration of the model uses original experimental data on the fitness of mutants resistant to the antiviral AZT preparation that blocks reverse transcription of the viral RNA into proviral DNA.

For a number of characteristics, the mathematical models constructed here are more realistic from the viewpoint of the HIV infection biology than other existing models in this area. The stochastic model implements the following replication process: (1) viral genomes (with the length of 1800 bases) encoded by a 4-letter alphabet, (2) the sizes of viral populations (up to 10^4), and (3) the coefficient of virus replication (up to 10^4). Within the framework of the deterministic model, processes of recombination leading to the appearance of drug-resistant mutants are parameterized. The comparison of solutions of two types of models demonstrates the need of further refinement of the parameter estimation procedure in the deterministic model according to experimental data on replication of mutants and clinical data on drug resistance. For small HIV populations, characteristic of replication processes localized in individual lymph nodes a natural simulation approach is the use of genetic algorithms. Consideration of HIV throughout the whole organism makes it necessary to follow descriptions based on ODE systems. However, some problems of parameter calibration and description of recombination in the context of multi-infection appear here. It seems that an approach considering virus populations together with classes of target cells infected by different kinds of specific mutants will be meaningful here.

The treatment of the HIV infection based on antiretroviral therapy reduces the viral load, but does not solve the problem of elimination of viruses. In this case, as was shown in recent studies [11], viral genomes in latently-infected cells contain variants of HIV with mutations in Gag genes coding the regions recognized by T-cells, i.e. escaping from the cytotoxic immune response. The ability of HIV to escape from immune response

is one of the most important mechanisms of pathogenesis of this infectious diseases and requires further studies with the use of mathematical models.

The virus population is a heterogeneous (with respect to antigens) ensemble of genomes with high adaptive potential to the action of humoral and cellular neutralization factors (selection). It is based on a high frequency of mutations and recombination of virus genomes in the case of multiple infections of target cells. The classical models of antiviral immune response developed by G. I. Marchuk consider virus populations being antigenically homogeneous. The application of these models for the analysis of the HIV infection involves the development of description methodology to take into account the antigenic heterogeneity and virus populations variability. The results presented in this paper, as well as other works in this area [3], show that the ability to describe the evolution of HIV using standard models of quasispecies dynamics based on differential equations by introducing into consideration the recombination of genomes is very limited due to the complexity of parameterization of transitions between particular quasispecies forming the ensemble of considered mutants. The use of discrete approaches based on genetic algorithms modelling the evolution of quasispecies as a population of vectors of the given length significantly enhances our ability to naturally describe the mutation and recombination processes [27]. Simulation of genomic sequences based on GA makes the urgent task of adjustment of discrete-stochastic models of the HIV quasispecies to continuous-deterministic modelling of immuno-physiological reactions. Overall, the proposed approach to the description of the HIV quasispecies diversification forms a basis for further research related to the analysis of development of 'resistance' patterns to the action of a combination of drugs and 'escape' from the immune response with the use of mathematical modelling.

References

[1]   P. Arora and N. M. Dixit, Timing the emergence of resistance to anti-HIV drugs with large genetic barriers. *PLoS Comp. Biol.* **5** (2009), No. 3, 1–10.

[2]   E. Baake, Mutation and recombination with tight linkage. *J Math. Biol.* **42** (2001), No. 5, 455–488.

[3]   E. Baake and I. Herms, Single-crossover dynamics: finite versus infinite populations. *Bull. Math. Biol.* **70** (2008), No. 2, 603–624.

[4]   C. K. Biebricher and M. Eigen, What is quasispecies? *Current Topics Microbiol. Immunol.* **299** (2006), 1–31.

[5]   G. A. Bocharov, N. J. Ford, J. Edwards, T. Breinig, S. Wain-Hobson, and A. Meyerhans, A genetic algorithm approach to simulating human immunodeficiency virus evolution reveals the strong impact of multiply infected cells and recombination. *J. General Virology* **86** (2005), No. 11, 3109–3118.

[6]   G. Bocharov, V. Chereshnev, I. Gainova, S. Bazhan, B. Bachmetyev, J. Argilaguet, J. Martinez, and A. Meyerhans, Human immunodeficiency virus infection: from biological observations to mechanistic mathematical modelling. *Math. Model. Natur. Phenom.* **7** (2012), No. 5, 78–104.

[7]   M. T. Bretscher, C. L. Althaus, V. Muller, and S. Bonhoeffer, Recombination in HIV and the evolution of drug resistance: for better or for worse? *BioEssays* **26** (2004), No. 2, 180–188.

[8]   D. S. Burke, K. A. De Jong, J. J. Grefenstette, and K. L. Ramsey, Putting more genetics into genetic algorithms. *J. Evolut. Comp.* **6** (1998), No. 4, 387–410.

[9]   V. A. Chereshnev, G. A. Bocharov, S. Bazhan, B. Bachmetyev, I. Gainova, V. Likhoshvai, J. M. Argilaguet, J. P. Martinez, J. A. Rump, B. Mothe, C. Brander, A. Meyerhans, Pathogenesis and treatment of HIV infection: the cellular, the immune system and the neuroendocrine systems perspective. *Int. Rev. Immunol.* **32** (2013), 282–306.

[10] K. Deforche, R. Camacho, K. V. Laethem, B. Shapiro, Y. Moreau, A. Rambaut, A.Vandamme, and P. Lemey, Putting more genetics into genetic algorithms. *J. Comp. Biol.* **14** (2007), No. 8, 1105–1114.

[11] K. Deng, M. Pertea, A. Rongvaux, L. Wang, C. M. Durand, G. Ghiaur, J. Lai, H. L. McHugh, H. Hao, H. Zhang, J. B. Margolick, C. Gurer, A. J. Murphy, D. M. Valenzuela, G. D. Yancopoulos, S. G. Deeks, T. Strowig, P. Kumar, J. D. Siliciano, S. L. Salzberg, R. A. Flavell, L. Shan, R. F. Siliciano, Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* **517** (2015), No. 7534, 381–385.

[12] M. Eigen, Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58** (1971), No. 10, 465–523.

[13] M. Eigen and P. Schuster, The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* **64** (1977), No. 11, 541–565.

[14] M. Eigen and P. Schuster, The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle. *Naturwissenschaften.* **65** (1978), No. 11, 7–41.

[15] V. V. Emelyanov, V. M. Kureichik, and V. V. Kureichik, *Theory and Practice of Evolutionary Modelling Genetic Algorithms.* Fizmatlit, Moscow, 2003 (in Russian).

[16] L. A. Gladkov, V. V. Kureichik, and V. M. Kureichik, *Genetic Algorithms. Textbook.* Fizmatlit, Moscow, 2006 (in Russian).

[17] J. H. Holland, *Adaptation in Natural and Artificial Systems.* Univ. of Michigan Press, Ann Arbor, 1975.

[18] A. N. Ignatovich, V. E. Kulikova, N. A. Medvedeva, E. A. Chiglintsev, and G. A. Bocharov, Mathematical technologies of modelling the virus and immune response dynamics. *Electr. J. Struct. Dynam. Molec. Syst.* Kazan State Univ. Part A. **2** (2008), No. 4, 350–386.

[19] G. I. Marchuk, *Mathematical Modelling of Immune Response in Infectious Diseases.* Kluwer Acad. Publ., Dordrecht, 1997.

[20] J. P. Martinez, G. Bocharov, A. Ignatovich, J. Reiter, M. T. Dittar, S. Wain-Hobson, and A. Meyerhans, Fitness ranking of individual mutants drives patterns of epistatic interactions in HIV-1. *PLoS ONE* **6** (2011), No. 3, e18375.

[21] M. A. Nowak and R. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology.* Oxford Univ. Press, Oxford, 2000.

[22] I. V. Oseledets and E. E. Tyrtyshnikov, Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM J. Sci. Comput.* **31** (2009), No. 5, 3744–3759.

[23] I. V. Oseledets and E. E. Tyrtyshnikov, TT-cross approximation for multidimensional arrays. *Linear Algebra Appl.* **432** (2010), No. 1, 70–88.

[24] V. G. Redko, *Evolutionary Cybernetics.* Nauka, Moscow, 2001 (in Russian).

[25] I. M. Rouzine and J. M. Goffin, Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol Mol. Biol. Rev.* **65** (2001), No. 1, 151–185.

[26] A. Schultz, S. Sopper, U. Sauermann, A. Meyerhans, and R. Suspene, Stable multi-infection of splenocytes during SIV infection : the basis for continuous recombination. *Retrovirology* **9** (2012), No. 1, 31.

[27] K. Tripathi, R. Balagam, N. K. Vishnoi, and N. M. Dixit, Stochastic simulations suggest that HIV-1 survives close to its error threshold. *Comput. Biol.* **8** (2012), No. 9, e1002684.

[28] N. V. Vijay, Vasantika, R. Ajmani, A. S. Perelson, and N. M. Dixit, Recombination increases human immunodeficiency virus fitness, but not necessarily diversity. *J. General Virology* **89** (2008), 1467–1477.

[29] R. M. Zinkernagel, *Grundlagen Der Immunologie. Medizinische Mikrobiologie* (Eds. F. H. Kayser, K. A. Bienz, J. Eckert, R. M. Zinkernagel). Georg Thieme Verlag, Stuttgart–New York, 2005, pp. 47–152.