Unveiling HIV mutational networks associated to pharmacological selective pressure: a temporal Bayesian approach

Pablo Hernandez-Leal[1], Alma Rios-Flores[1], Santiago Ávila-Rios[2], Gustavo Reyes-Terán[2], Jesus A. González[1], Felipe Orihuela-Espina[1], Eduardo F. Morales[1], and L. Enrique Sucar[1]

[1] National Institute of Astrophysics, Optics and Electronics (INAOE)
Tonantzintla, Puebla, Mexico
{pablohl,alma.rios,jagonzalez,f.orihuela-espina,emorales,esucar}@inaoep.mx
[2] Infectious Diseases Research Center at the National Institute of Respiratory Diseases (CIENI-INER)
Mexico City, Mexico
{santiago.avila}@cieni.org.mx

Abstract. Much of the HIV (Human Immunodeficiency Virus) success is due to its evolving capabilities. Understanding viral evolution and its relation to pharmacology is of utmost importance in fighting diseases caused by the HIV. Although the mutations conferring drug resistance are mostly known, the dynamics of the appearance chain of those mutations remains poorly understood. Here we apply a Temporal Nodes Bayesian Network (TNBN) to data extracted from the HIV Stanford database to explore the probabilistic relationships between mutations and antiretrovirals. We aim to unveil existing mutation networks and establish their probabilistic formation sequence. The model predictive accuracy is evaluated in terms of relative Brier score, relative time error and total number of intervals. Robustness of the model is hinted by consistency between two model instances. The learned models capture known relationships, qualitatively providing some measure of validity. Finally, two previously unseen mutational networks are retrieved and their probabilistic temporal sequentiation uncovered. We demonstrate the usefulness of TNBN for studying drug-mutation and mutation-mutation networks and expect to impact the combat against HIV infection by facilitating better treatment planning.

1 Introduction

Viral evolution is an important aspect of the epidemiology of viral diseases such as influenza, hepatitis and human immunodeficiency virus (HIV). This evolution greatly impacts the development of successful vaccines and antiviral drugs, as mutations bestowing drug resistance or immune escape often develop early after the virus is placed under selective pressure. In HIV, this is particularly relevant as the virus ranks among the fastest evolving organisms [7]. Its remarkable viral

replication capability is coupled with a high mutation rate and a high probability of recombination in the viral genome during its replication cycle. These features allow HIV to boast a wide genetic variability even considering only the viral population within a given host. The elevated variation capability of HIV gives the virus a remarkable ability to adapt to multiple selective pressures, including the immune response and antiretroviral therapy. This intra-host genetic variation raises several questions about viral evolution, for example: How much of this diversity is shaped by the selection of the immune response and how much by the antiretroviral therapy? What is the relationship between genetic diversity and clinical outcome? And finally, is it feasible to sheer the evolution of HIV in order to reduce drug resistance? Motivated by this last question, it would be desirable to develop proactive therapies that predict the advent of mutations, ergo reducing the risk of drug resistance, rather than waiting for the virus to develop resistance to reactively change the antiretroviral regimen. If we could predict the most likely evolution of the virus in any host, then it would be plausible to select an appropriate antiretroviral regimen that prevents the appearance of mutations, effectively increasing HIV control.

In this work, a Temporal Node Bayesian Networks (TNBN) model was developed to assess the occurrence of probabilistic associations among protease mutations and protease inhibitor drugs. Results of the learning of the model are presented. Our goal was to explain mutational networks in HIV evolution in the viral protease. Our probabilistic graphical model was able to predict antiretroviral drug-associated mutational pathways in the protease gene, revealing the co-occurrence of mutations and its temporal relationships. The technical challenge was to develop a model expressive enough to capture the biological complexity, yet simple enough to allow for a quick interpretation of results. The use of TNBNs for exposing mutational networks is, as far as the authors are aware of, an additional novelty to this work.

The rest of the paper is organized as follows. Section 2 highlights some important notions regarding HIV and how it develops drug resistance. Section 3 justifies the use of TNBN over other existing graphical probabilistic approaches. Section 4 describes the TNBN model. Section 5 presents the experiments and results obtained. Finally, section 6 summarizes the findings and indicates future lines of research.

2 HIV and its defense against antiretroviral therapy

HIV is the causing agent of the disease known as Acquired Immunodeficiency Syndrome (AIDS), a condition in which progressive failure of the immune system allows opportunistic life-threatening infections to occur. HIV is a virus with relatively recent introduction to human populations [17] representing a huge global burden to human health (UNAIDS HIV Global Report 2010). Its structure is schematically depicted in Figure 1.

The replication cycle of HIV is characterized by a reverse-transcription step of the viral RNA genome to a double-stranded DNA molecule, which is then

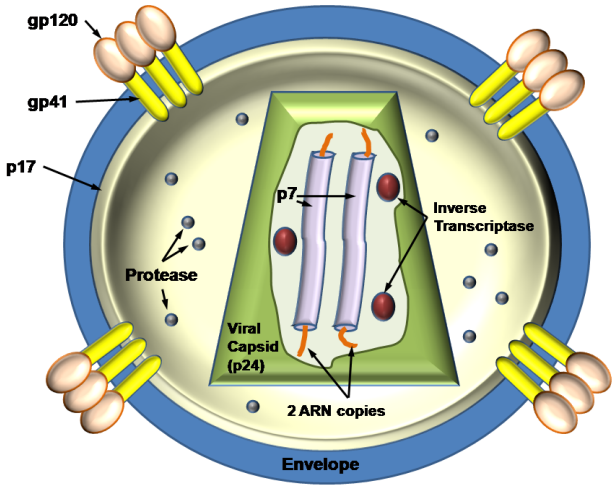

Fig. 1. Schematic representation of the HIV structure. The protease encyme helping in maturation is illustrated. This enzyme is the target of the protease inhibitors antiretrovirals which are studied in this paper. Some important proteins, e.g. gp120, gp41, p17 and p7 have also been illustrated.

inserted into the host cell genome. To combat HIV infection several antiretroviral (ARV) drugs belonging to different drug classes that affect specific steps in the viral replication cycle have been developed. Antiretroviral therapy (ART) generally consists of well-defined combinations of three or four ARV drugs. Due to its remarkable variation capabilities, HIV can rapidly adapt to the selective pressure imposed by ART through the development of drug resistance mutations, that are fixed in the viral population within the host in known mutational pathways. The development of drug resistant viruses compromises HIV control, with a consequent further deterioration of the patient's immune system. Many of these ARV drug resistance mutations reduce HIV susceptibility to ARV drugs by themselves, while others need to accumulate in order to cause resistance. Moreover, the appearance of some drug resistance mutations implies high costs for viral replication capacity. These costs in viral replication capacity are often compensated by the appearance of additional mutations known as compensating mutations. Due to their polymorphic nature the frequency of compensating mutations can vary between viruses circulating in different geographic areas, making it relevant to study HIV mutational networks in the context of different infected populations.

Abundant literature exists describing computational models aimed to better understand HIV evolution and immunopathogenesis. A good portion of this data is devoted to predict phenotypic HIV resistance to antiretroviral drugs using different approaches such as decision trees [2] or neural networks [6]. Other works try to identify relevant associations between clinical variables and HIV

disease [18]. Surprisingly, among this wealth of literature, works aimed towards the identification of temporal relationships among mutations and drugs in HIV is almost lacking.

In [3] association rules between clinical variables and the failure of the treatment are extracted, they used 15 clinical variables from 8000 patients from data collected since 1981. The results obtained are temporal rules that have as antecedent the increasing of a subset of clinical variables and as consequent the failure of the treatment, given by side effects of the drugs or by the elevated viral count (unsuccessful therapy). None of clinical variables considered are VIH mutations.

3 Bayesian Networks

Information in clinical databases is more often than not imprecise, incomplete, and with errors (noisy), and Bayesian Networks (BNs) [16] are particularly well suited to deal with uncertainty. BNs study probabilistic dependencies and relationships among domain entities. BNs models admit visual representation as a graph consisting of nodes and edges facilitating their analysis and interpretation. Nodes represent random variables and edges represent probabilistic dependencies. This graphical representation is easily understood by humans. An additional advantage is the availability of several methods to learn BN from data, e.g. [14].

BNs have proven to be successful in various domains such as medicine [15] and bioinformatics [20]. However, classical BNs are not well equipped to deal with temporal information. Dynamic Bayesian Networks (DBNs) evolved to tackle this shortcoming [5]. DBNs can be seen as multiple slices of a *static* BN over time, with temporal relations captured as links between adjacent slices. In a DBN, a base model is cloned for each time stage. These copies are linked via the so-called transition network. In this transition network is common that only links between consecutive stages are allowed. Whenever variable changes occur infrequently, the explicit representation of DBNs becomes unnecessarily overexpressive. The alternative are TNBNs [1].

4 Temporal Nodes Bayesian Networks

In a TNBN, each node, known as *temporal node* (TN), represents a random variable that may be in a given state i.e. value interval, throughout the different temporal intervals associated to it. An arc between two temporal nodes describes a temporal probabilistic relation. In TNBNs, each variable (node) represents an event or state change. So, only one (or a few) instance(s) of each variable is required, assuming there is one (or a few) change(s) of a variable state in the temporal range of interest. No copies of the model are needed, thus compacting the representation without losing expressiveness.

The TNBN [1, 9] is composed by a set of TNs connected by arcs representing a probabilistic relationship between TNs. A TN, v_i, is a random variable characterized by a set of states $\mathbf{S}$. Each state is defined by an ordered pair

$S = (\lambda, \tau)$, where λ is the particular value taken by v_i during its associated interval $\tau = [a, b]$, corresponding to the time interval in which the state change, i.e. change in value, occurs. In addition, each TN contains an extra default state $s =$ ('no change', $\emptyset$) with no associated interval. Time is discretized in a finite number of intervals, allowing a different number and duration of intervals for each node (multiple granularity). Each interval defined for a child node represents the possible delays between the occurrence of one of its parent events and the corresponding child event. If a node lacks defined intervals for all its states then it is referred to as *instantaneous node*. There is at most one state change for each variable (TN) in the temporal range of interest.

Formally, let $\mathbf{V}$ be a set of temporal and instantaneous nodes and $\mathbf{E}$ a set of arcs between nodes, a TNBN is defined as:

Definition 1. *A TNBN is a pair $B = (G, \Theta)$ where G is a directed acyclic graph, $G = (\mathbf{V}, \mathbf{E})$ and, Θ is a set of parameters quantifying the network. Θ contains the values $\Theta_{v_i} = P(v_i|Pa(v_i))$ for each $v_i \in \mathbf{V}$; where $Pa(v_i)$ represents the set of parents of v_i in G.*

The learning algorithm for TNBN used in this work has been presented in [11]. Briefly, the learning algorithm is described:

1. First, it performs an initial discretization of the temporal variables, for example using an Equal-Width discretization. With this process it obtains an initial approximation of the intervals for all the Temporal Nodes.
2. Then it performs a standard BN structural learning, the algorithm uses the K2 learning algorithm [4], to obtain an initial structure. This structure will be used in the third step, the interval learning algorithm.
3. The interval learning algorithm refines the intervals for each TN by means of clustering. For this, it uses the information of the configurations of the parent nodes. To obtain some intervals a clustering algorithm for the temporal data is used. The approach uses a Gaussian mixture model. Each cluster corresponds, in principle, to a temporal interval. The intervals are defined in terms of the μ and the σ of the clusters. The algorithm obtains different sets of intervals that are merged and combined, these process will generate different interval sets that will be evaluated in terms of the predictive accuracy (Relative Brier Score). The best set of intervals (that may not be those obtained in the first step) for each TN is selected based on predictive accuracy. When a TN has as parents other Temporal Nodes (an example of this situation is illustrated in Figure 4), the configurations of the parent nodes are not initially known. So, in order to solve this problem, the intervals are selected sequentially in a top-down fashion according to the TNBN structure.

The algorithm then iterates between the structure learning and the interval learning. However, for the experiments presented in this work, we present the results of the first iteration.

5 Experiments

5.1 Data and preprocessing

Data was obtained from the HIV Stanford Database (HIVDB) [19]. The isolates in the HIV Drug Resistance Database were obtained from longitudinal treatment profiles reporting the evolution of mutations in individual sequences.

In total data from 2373 patients with subtype B was retrieved. We choose to work with this subtype because it is the most common in America [10], our geographical region of interest. For each patient data retrieved contains a history consisting of a variable number of studies. Information regarding each study consists of a treatment or cocktail of drugs administered to the patient, how long the treatment lasted in weeks, and the list of more frequent mutations in the viral population within the host at the time when the treatment was suspended (changed for a different treatment). An example of the data is presented in Table 1.

Table 1. An example of the data. It presents two patients P_1 with 3 temporal studies, and P_2 with two temporal studies.

Patient	**Treatment**	**List of Mutations**	**Weeks**
P_1	LPV, FPV, RTV	L63P, L10I	15
		V77I	30
		I62V	10
P_2	NFV, RTV, SQV	L10I	25
		V77I	45

For applying the learning algorithm for TNBN the data presented in Table 1 is transformed into a table similar to the one presented in Table 2. Here, each column represents a drug or mutation, each row represents a patient case, for the drugs the values are USED or NOT USED, and for the drugs the values are: APPEAR with the number of weeks that mutation appeared the first time or Default, this is when mutation did not appear in that case. The ordering provided to the K2 algorithm is: first the antiretrovirals ordered by frequency, then the mutations ordered by frequency.

The number of studies available varies from 1 to 10 studies per patient history. Since we are interested in temporal evolution of the mutational networks, we filtered those patients having less than 2 studies, with 973 patients outliving the conditional.

Antiretrovirals are usually classified according to the enzyme that they target. We focus on protease as this is the smallest of the major enzymes in terms of number of aminoacids. There exist 9 protease inhibitors (PI), namely: Amprenavir (APV), Atazanavir (ATV), Darunavir (DRV), Lopinavir (LPV), Indinavir (IDV), Nelfinavir (NFV), Ritonavir (RTV), Tripanavir (TPV) and Saquinavir

Table 2. An example of the data used to learn the TNBN model. Each row represents a patient. Each column represents either a drug, used or not, or a mutation that appeared or not.

Drug-1	**Drug-2**	**...**	**Mutation-1**	**Mutation-2**	**...**
USED	NOT USED		(APPEAR) 30	Default	
USED	NOT USED		(APPEAR) 40	Default	
USED	USED		Default	(APPEAR) 80	

(SQV). All 9 PIs will be considered during the experiments. Figure 2 presents the histogram of the administration of the different PIs in the dataset. Data from HIVDB originates from different studies and in some cases is incomplete. In this sense, Figure 2 evidences a small portion of studies only reporting the administration of a PI, but the specific antiretroviral is missing. Also there is slightly bigger portion reporting *Unknown* as the drug used. The way we handle this cases was, if the patient case only contained *Unknown* or *None* that case was removed. However, if the case contained other drug (apart from *Unknown*), that information was used for the model.

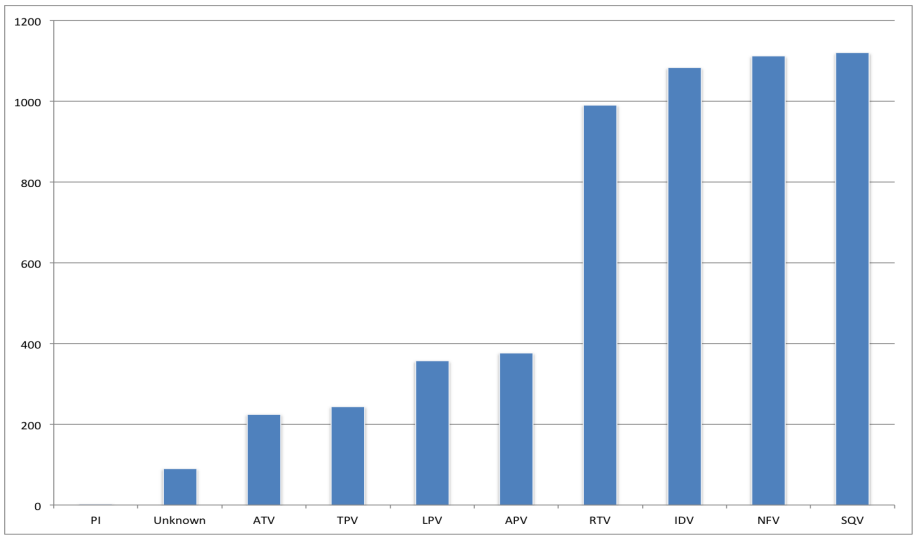

Fig. 2. Histogram of the protease inhibitors administration in the full dataset including all 2373 patients.

We still have to define a target set of mutations of interest. Figure 3 presents the histogram of mutations as they appear in the dataset. 733 different mutations appeared at least once in the data. But as evident from the histogram, most

mutations are rare presenting low frequency. Indeed, only few mutations exhibit high frequency. Mutation inclusion criteria is detailed below.

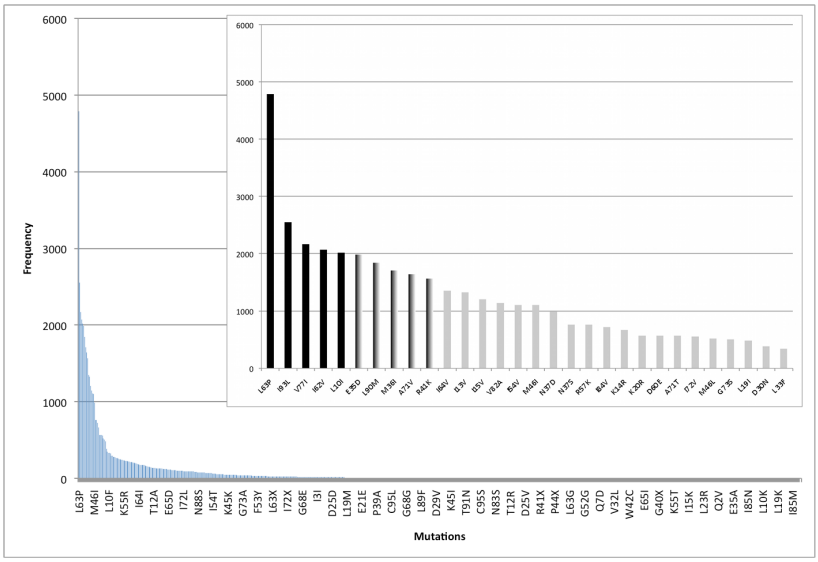

Fig. 3. A group of mutations and their frequency in the full dataset including all 2373 patients. The higher frequency end of the spectrum is zoomed. Mutations used for the first experiment are shown in black, and the completing set for the second experiment are indicated in shaded black.

5.2 Model evaluation

The learning algorithm arrives to local maxima, and thus is influenced by initial parameterization. For our experiments, the number of initial intervals for each node was allowed to vary from 2 to 4 and Equal Width Discretization [13] was used to initialize those intervals. Since it does not exist a gold standard or a reference TNBN, in evaluating the model's performance three indirect measures were used: the relative Brier score (RBS), the relative time error and the total number of intervals in the model. From the 973 total patients 80% of them were used for learning and the rest for evaluation.

The Brier Score is a measure of the predictive accuracy of the network, is defined as

$$BS = \frac{1}{n}\sum_{i=1}^{n}(1 - P_i)^2$$

where P_i is the marginal posterior probability of the correct value of each node given the evidence, this applies for all the selected nodes, n, of the TNBN. The

RBS is defined as:

$$RBS \ \ (in \ \ \%) = (1 - BS) \times 100$$

For each case of the data (a row in table 1), the RBS is obtained by instantiating a random subset of variables in the model, predicting the unseen variables, and obtaining the RBS for these predictions. The relative time error is a measure to evaluate how far the real events are from the intervals and it is defined as the difference between the real event and the middle point of the interval divided by the range of the temporal node. The range of the node is the difference between the maximum and the minimum values of the intervals in a temporal node. Finally, the number of intervals is defined as the total number of intervals learned across all variables, this is a rough estimate of the complexity of the network and a low number of intervals is a desirable property for simplicity of the model. The best model would afford a high RBS, a low time error and a low complexity (low number of intervals). The technical performance of the model reflects its predictive accuracy and complexity, but should not be confused with the biological/physiological plausibility of the model.

5.3 Results

Two experiments have been carried out. The first experiment with a smaller model aims to assess the capability of TNBN for capturing known relations and thus providing a qualitative validation of the approach. The second, with a more complete model is aimed at uncovering the more common existing mutational networks and capturing the temporal aspect of the network formation.

In the first experiment, only mutations with more than 2000 counts were used: L63P, I93L, V77I, I62V and L10I. For tractability, in the second experiment only those mutations appearing more than 1500 times were included: L63P, I93L, V77I, I62V, L10I, E35D, L90M, M36I, A71V and R41K.

Table 3. Evaluation of the models in terms of RBS, relative time error (in percentage) and number of intervals.

Experiment	**Initial intervals**	**RBS**	**Relative Time error**	**Number of total intervals**
1	2	**89.8**	**13.0**	**17**
	3	88.3	13.6	20
	4	88.5	13.9	19
2	2	87.3	15.0	**30**
	3	**88.5**	**14.7**	31
	4	87.5	15.9	35

Table 3 summarizes the results for the two experiments. Figure 4 illustrates the TNBN of the first experiment exhibiting the best scores. The figure represents the network, the intervals and the prior probabilities obtained for each TN.

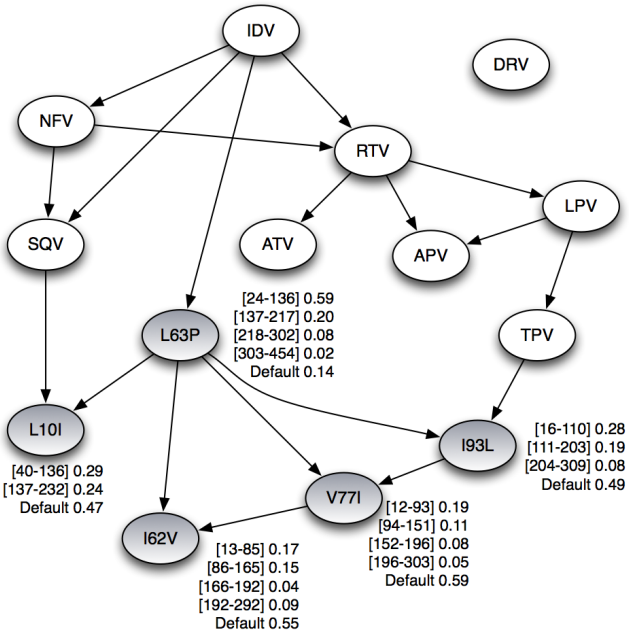

Fig. 4. A learned TNBN with 9 Protease Inhibitors and 5 mutations that appear frequently. Drugs are indicated in white bubbles and mutations in gray bubbles. Learned time intervals and their probabilities for the TNs are indicated beside the bubble as is standard in the representation of TNBNs.

Exploring the model reveals that RTV has arcs linking it with IDV, NFV, ATV, APV and LPV. This important relationship of RTV with other medicaments is explained due to the fact that the Ritonavir drug has been proved to boost the effect of other PIs, and therefore most of the times it is administered in combination with other drugs. The link between SQV and L10I was also already known to clinicians [12] and our model has also been successful in uncovering it. The observation of these known fact boosts our confidence that the model was meaningful; but can the model reveal new knowledge? The DRV node in the model is isolated because in the data, was never given as part of a first treatment for any of the patients. This is perhaps because DRV is a relatively new drug. L63P is an extremely common mutation in the viral genotype coding the protease as revealed by the histogram in Figure 3. However, "when" this mutation appears in the evolving virus remained mysterious. Our model suggests that most times this mutation tends to appear early in time, and that its probability to appear decreases over it.

Figure 5 illustrates the best TNBN model instantiation in terms of higher RBS for the second experiment. Most of the arcs from the smaller model were retained. Only the relation linking I62V and V77I and TPV with I93L are drop. Moreover, only two new arcs from SVQ and TPV to L63P appear among previ-

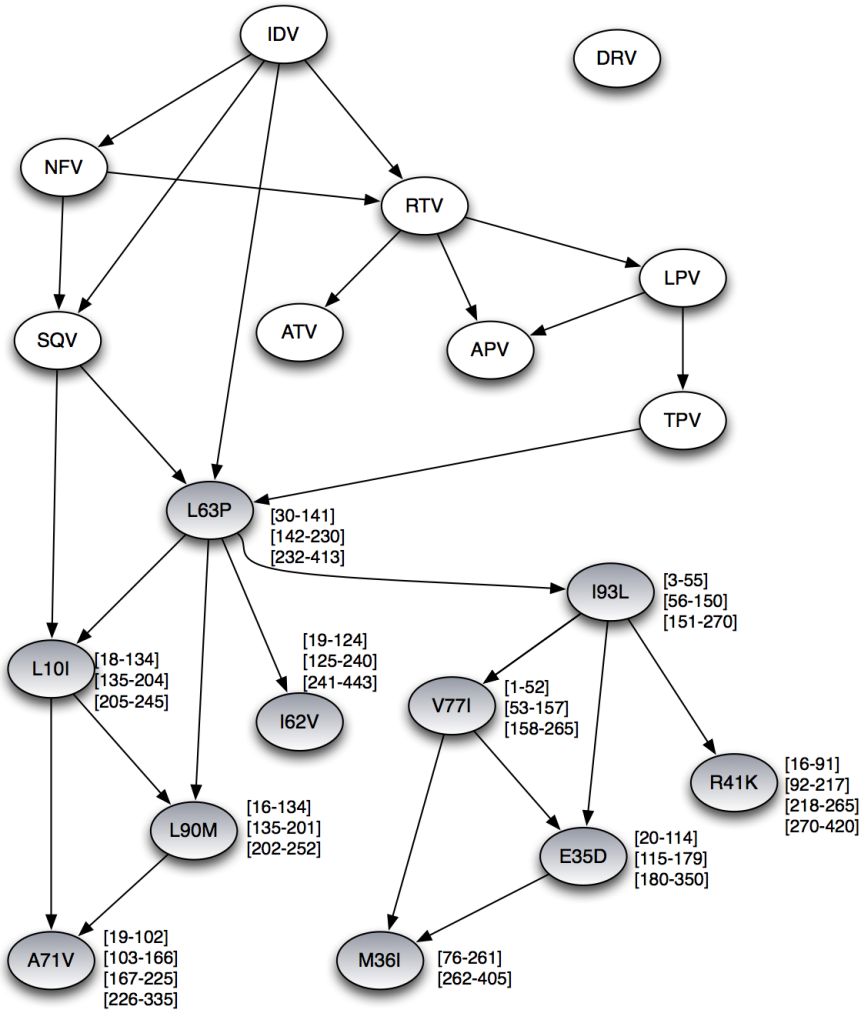

Fig. 5. A learned TNBN with 9 Protease Inhibitors and 10 mutations that appear frequently. Drugs are indicated in white bubbles and mutations in gray bubbles. Learned time intervals for the TNs are indicated beside the bubble. Probabilities and the Default state for the TNs are hidden for readability.

ously considered elements. This small variation among the two models is a good indicator of the robustness of the modeling approach. From this more complete model the prevalence of mutation L63P is evident from its relation to most drugs. It can then diverge to other mutations. There are two possible explanations for this observation; either the frequency of appearance of L63P is biasing the formation of the associations in the model -L63P almost doubles in frequency that of the following most common mutation-, or L63P is a key mutation to unleash

others. Additionally, the local neighborhoods in the graph clearly reveal two mutational networks;

* L63P, I62V, L10I, L90M and A71V
* I93L, V77L, M36I, E35D and R41K

We are not aware of these mutational networks to have previously been reported. Moreover, the TNBN further reveals the temporal sequence of mutation appearance.

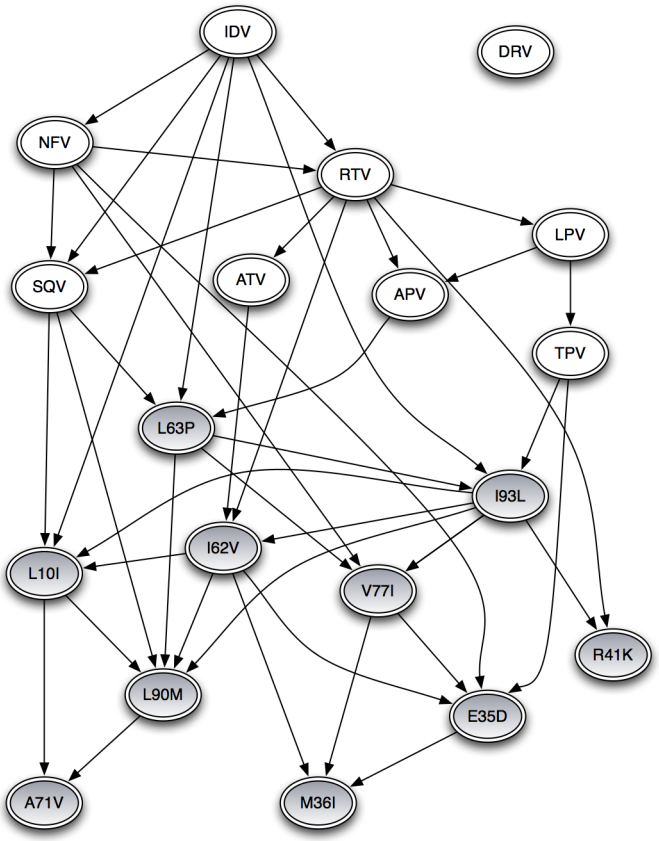

Fig. 6. A learned *static* BN with 9 Protease Inhibitors and 10 mutations that appear frequently. Drugs are indicated in white bubbles, their states are Used or Not Used. Mutations are presented in gray bubbles, their states are Appeared or Not Appeared. States and probabilities are hidden for readability.

As a final exercise, the temporal information of the mutations has been removed. Hence, the states for both the drugs and the mutations were: Appear

or Not Appeared. We used the same ordering as in the previous experiments and applied the K2 learning algorithm. The static BN learned is presented in Figure 6. The number of arcs increased. More importantly, most of the arcs obtained with the TNBN remained. Using temporal information in this particular case yields a simpler model. Further analysis of these experiments is needed, to dilucidate whether this is always the case. This last exercise while interesting computationally, is far from construct validating none of the models. In this sense we are of course short from being able to determine which one is more correct, even though intuitively the simpler TNBN seems more adequate by Occam's razor.

6 Conclusions

By using a TNBN we have been able to unveil the two more common mutational networks present in HIV evolution as response to pharmacological selective pressure, and we believe these to be previously unreported. Our model has been successful in capturing relationships between mutations and protease inhibitors critically incorporating temporal information. These results are encouraging, presenting the model as an effective tool to explain how mutations interact with each other and providing some leverage for the clinicians in interpreting clinical tests. In this sense, the success of the second model still raises more questions. For example, why are ATV and APV not related to any mutation? This demands further investigation.

Models such as ours are an initial step to facilitate treatment planning. If a certain mutation occurring early in a mutational network is observed during a sequentiation, one would expect the other mutations in the network to follow. Knowing the likely appearing of subsequent mutations gives the therapist an edge in determining the appropriate antiretroviral regimen.

Future work plans to use TNBNs models to unfold mutational networks in the Reverse Transcriptase, another important enzyme of HIV. Technically, robustness of the model may be objectively assessed by means of bootstrap [8] to check which substructures remain in the model across different subsets of data. Finally, formal validation of the approach is still pending.

Acknowledgements

El trabajo que ha dado lugar a este invento ha recibido apoyo del CONACYT y de la Comunidad Europea a través del FONCICYT en virtud del contrato de asignación de recursos/contrato de subvención n 95185. The first author is supported by a grant 234507 from CONACYT.

References

1. Arroyo-Figueroa, G., Sucar, L.E.: A temporal Bayesian network for diagnosis and prediction. In: Proceedings of the 15th UAI Conference. pp. 13–22. Stockholm, Sweden (1999)

2. Beerenwinkel, N., Schmidt, B., Walter, H., Kaiser, R., Lengauer, T., Hoffmann, D., Korn, K., Selbig, J.: Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. Proceedings of the National Academy of Sciences of the United States of America 99(12), 8271–8276 (2002)
3. Chausa, P., Cáceres, C., Sacchi, L., León, A., García, F., Bellazzi, R., Gómez, E.: Temporal Data Mining of HIV Registries: Results from a 25 Years Follow-Up. Artificial Intelligence in Medicine pp. 56–60 (2009)
4. Cooper, G., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. Machine learning 9(4), 309–347 (1992)
5. Dagum, P., Galper, A., Horvitz, E.: Dynamic network models for forecasting. In: Proceedings of the 8th Workshop UAI. pp. 41–48. Stanford, California, USA (1992)
6. Draghici, S., Potter, R.B.: Predicting HIV drug resistance with neural networks. Bioinformatics 19(1), 98–107 (2003)
7. Freeman, S., Herron, J., Payton, M.: Evolutionary analysis. Prentice Hall Upper Saddle River, NJ: (1998)
8. Friedman, N., Goldszmidt, M., Wyner, A.: Data analysis with Bayesian networks: A bootstrap approach. In: Proceedings of the 15th UAI Conference. pp. 206–215. Stockholm, Sweden (1999)
9. Galán, S., Arroyo-Figueroa, G., Díez, F., Sucar, L.: Comparison of two types of event Bayesian networks: A case study. Applied Artificial Intelligence 21(3), 185 (2007)
10. Hemelaara, J., Gouws, E., Ghys, P.D., Osmanov, S.: Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20, W13–W23 (2006)
11. Hernandez-Leal, P., Sucar, L.E., Gonzalez, J.A.: Learning temporal nodes Bayesian networks. In: The 24th Florida Artificial Intelligence Research Society Conference (FLAIRS-24). Palm Beach, Florida, USA (2011)
12. Johnson, V., Brun-Vézinet, F., Clotet, B., Günthard, H., Kuritzkes, D., Pillay, D., Schapiro, J., Richman, D.: Update of the Drug Resistance Mutations in HIV-1: December 2010. Topics in HIV medicine 17, 138–145 (2010)
13. Liu, H., Hussain, F., Tan, C., Dash, M.: Discretization: An enabling technique. Data Mining and Knowledge Discovery 6(4), 393–423 (2002)
14. Neapolitan, R.: Learning Bayesian networks. Pearson Prentice Hall (2004)
15. Pang, B., Zhang, D., Li, N., Wang, K.: Computerized tongue diagnosis based on Bayesian networks. Biomedical Engineering, IEEE Transactions on 51(10), 1803–1810 (2004)
16. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann (1988)
17. Rambaut, A., Posada, D., Crandall, K.A., Holmes, E.C.: The causes and consequences of HIV evolution. Nature Reviews Genetics 5(1), 52–61 (2004)
18. Ramirez, J., Cook, D., Peterson, L., Peterson, D.: Temporal pattern discovery in course-of-disease data. Engineering in Medicine and Biology Magazine, IEEE 19(4), 63–71 (2000)
19. Rhee, S., Gonzales, M., Kantor, R., Betts, B., Ravela, J., Shafer, R.: Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic acids research 31(1), 298–303 (2003)
20. Rodin, A., Boerwinkle, E.: Mining genetic epidemiology data with Bayesian networks: Bayesian networks and example application (plasma apoE levels). Bioinformatics 21(15), 3273–3278 (2005)