

Algorithm Specification Interface for Human Immunodeficiency Virus Type 1 Genotypic Interpretation

Human immunodeficiency virus type 1 (HIV-1) genotypic resistance testing is frequently done to help select therapy for HIV-1-infected persons. The manual interpretation of such assays is difficult, because a large number of reverse transcriptase (RT) and protease drug resistance mutations interact and emerge in complex patterns. Several computerized interpretation algorithms have been developed that use RT and protease sequences to assess HIV-1 drug susceptibility and likely virologic response to a new antiretroviral regimen. Most of these algorithms contain rules assigning a level of resistance to a drug on the basis of a pattern of mutations.

Different algorithms for HIV-1 genotypic interpretation, however, not only contain different rules but also lack consistency in their programming languages and logical designs. Both of these factors may result in different interpretations for the same sequence, leading to confusion among clinicians and patients. Indeed, several groups have reported various levels of discordance between several of the commonly used drug resistance interpretation algorithms (2, 4, 5).

Although HIV-1 drug resistance interpretation algorithms will evolve with the publication of new data and introduction of new antiretroviral inhibitors, the software needed to implement these algorithms should remain stable to allow direct comparisons among interpretation algorithms and their component rules. We have developed a platform, which we call an Algorithm Specification Interface (ASI), that consists of an XML format for specifying an algorithm and a compiler that transforms the XML into executable code.

Here we describe the use of the ASI to encode five diverse publicly available HIV-1 genotypic resistance interpretation algorithms: Agencie Nationale de Recherches sur le Sida (3, 6), HIV RT and Protease Sequence Database (7), Resistance Collaborative Group (1), Rega Institute (8), and Visible Genetics Inc. (Toronto, Canada) (C. Reid, R. Bassett, S. Day, B. Larder, V. DeGruttola, and D. Winslow, abstr. from the 5th International Workshop on HIV Drug Resistance and Treatment Strategies, 4 to 8 June 2001, Scottsdale, Ariz., *Antivir. Ther.* 7:S91, 2002).

Figure 1 shows an example of an algorithm that contains each of the key elements, although for brevity in this exposition it has only two rules. The first part of the algorithm consists of an algorithm name, a version number, and a list of definitions. The definitions include the levels of drug resistance used by the algorithm, the list of drugs for which resistance levels are

generated, an optional reference sequence, and an optional list of comment strings (the last two items in this list are not shown in Fig. 1). The second part of the algorithm consists of a list of rules grouped by drug.

Each rule contains a condition and a set of actions. Conditions consist of statements in a language we developed for this purpose. The actions consist of (i) assigning a score to a drug, (ii) assigning a resistance level to a drug, or (iii) reporting a comment. Each condition contains at least one mutation (indicated by a numbered position followed by an amino acid abbreviation, e.g., 215Y) and none or one or more of the following words: SELECT, SCORE, NOTMORETHAN, ATLEAST, AND, OR, NOT.

The ASI Document Type Definition, XML coding for the five HIV genotypic drug resistance algorithms, the compiler, and detailed information about the syntax for encoding algorithms are available online (<http://hivdb.stanford.edu>). At this site, users can submit a list of HIV RT or protease mutations or nucleotide sequences for processing by as many of the five algorithms as they choose. They may also provide their own ASI-encoded algorithms by uploading them from their computers. The ASI is written in Perl and can be downloaded from our web site. If the ASI software is installed, algorithms can be implemented on the user's machine as well as online.

Developing a common platform for writing algorithms does not guarantee their validity. Rather, the ASI facilitates the encoding and modification of algorithms by clinical virologists and makes it feasible to compare the results of algorithms using large numbers of sequences or lists of mutations. The five algorithms we encoded were chosen because they are publicly available and widely used. Doubtlessly other algorithms can also take immediate advantage of the ASI. As new algorithms are released, or as existing algorithms are modified, the ASI Document Type Definition, the syntax of our language, and the compiler may need to evolve to accommodate new types of rules.

Algorithm specification interfaces that provide a common platform for encoding, implementing, and comparing genotypic interpretation algorithms will become necessary for other genes—particularly microbial genes with complex patterns of drug resistance mutations—if the sequences of these genes are used as the basis for medical decisions. Such interfaces will prevent the implementation of algorithms from becoming

```

<?xml version="1.0" standalone="no"?>
<!DOCTYPE ALGORITHM SYSTEM
"http://hivdb2.stanford.edu/asi/hivProduction/xml/ASI.dtd">

<ALGORITHM>
  <ALGNAME>Example Algorithm For Paper</ALGNAME>
  <ALGVERSION>2002.06</ALGVERSION>

  <DEFINITIONS>
    <LEVEL_DEFINITION>
      <ORDER>1</ORDER>
      <ORIGINAL>Drug is Active</ORIGINAL>
      <SIR>S</SIR>
    </LEVEL_DEFINITION>

    <LEVEL_DEFINITION>
      <ORDER>2</ORDER>
      <ORIGINAL>Drug is Partially Active</ORIGINAL>
      <SIR>I</SIR>
    </LEVEL_DEFINITION>

    <LEVEL_DEFINITION>
      <ORDER>3</ORDER>
      <ORIGINAL>Drug is Inactive</ORIGINAL>
      <SIR>R</SIR>
    </LEVEL_DEFINITION>

    <DRUGCLASS>
      <NAME>NRTI</NAME>
      <DRUGLIST>EXAMPLE_DRUG1</DRUGLIST>
    </DRUGCLASS>
  </DEFINITIONS>

  <DRUG>    <!-- Based on DDI drug in RCG -->
    <NAME>EXAMPLE_DRUG1</NAME>
    <RULE>
      <CONDITION>
        SELECT ATLEAST 1 FROM (65R,74V,151M,184VI,69i)
      </CONDITION>
      <ACTIONS>
        <LEVEL>3</LEVEL>
      </ACTIONS>
    </RULE>

    <RULE>
      <CONDITION>
        215YF AND SELECT ATLEAST 2 FROM (41L,67N,70R,210W,219QE)
      </CONDITION>
      <ACTIONS>
        <LEVEL>2</LEVEL>
      </ACTIONS>
    </RULE>
  </DRUG>
</ALGORITHM>

```

FIG. 1. An algorithm encoded using the XML template for the ASI for HIV drug resistance. Although it is syntactically correct, this algorithm contains only two rules and was therefore only designed as an example. This file shows the required XML tags, including the introductory tags for algorithm name, level definitions, drug classes, and drugs, as well as the tags for rules, conditions, and actions. The complete XML specification of five algorithms can be found at <http://hivdb.stanford.edu>.

locked within inaccessible proprietary formats and will allow experts to focus on developing, testing, and modifying algorithms rather than on developing software to encode them.

REFERENCES

1. DeGruttola, V., L. Dix, R. D'Aquila, D. Holder, A. Phillips, M. Ait-Khaled, J. Baxter, P. Clevenbergh, S. Hammer, R. Harrigan, D. Katzenstein, R. Lanier, M. Miller, M. Para, S. Yerly, A. Zolopa, J. Murray, A. Patick, V. Miller, S. Castillo, L. Pedneault, and J. Mellors. 2000. The relation between baseline HIV drug resistance and response to antiretroviral therapy: re-analysis of retrospective and prospective studies using a standardized data analysis plan. *Antivir. Ther.* 5:41-48.
2. Kijak, G. H., A. E. Rubio, S. E. Pampuro, C. Zala, P. Cahn, R. Galli, J. S. Montaner, and H. Salomon. 2003. Discrepant results in the interpretation of HIV-1 drug-resistance genotypic data among widely used algorithms. *HIV Med.* 4:72-78.
3. Meynard, J. L., M. Vray, L. Morand-Joubert, E. Race, D. Descamps, G. Peytavin, S. Matheron, C. Lamotte, S. Guiramand, D. Costagliola, F. Brun-Vizinet, F. Clavel, and P.-M. Girard. 2002. Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. *AIDS* 16:727-736.
4. Puchhammer-Stockl, E., C. Steininger, E. Geringer, and F. X. Heinz. 2002.

- Comparison of virtual phenotype and HIV-SEQ program (Stanford) interpretation for predicting drug resistance of HIV strains. *HIV Med.* **3**:200–206.
5. **Ravela, J., B. J. Betts, F. Brun-Vezinet, A. M. Vandamme, D. Descamps, K. Van Laethem, K. Smith, J. M. Shapiro, D. Winslow, C. Reid, and R. W. Shafer.** 2003. HIV-1 protease and reverse transcriptase mutation patterns causing discordances between drug resistance interpretation algorithms. *J. Acquir. Immun. Defic. Syndr.* **33**:8–14.
 6. **Rousseau, M. N., L. Vergne, B. Montes, M. Peeters, J. Reynes, E. Delaporte, and M. Segondy.** 2001. Patterns of resistance mutations to antiretroviral drugs in extensively treated HIV-1-infected patients with failure of highly active antiretroviral therapy. *J. Acquir. Immune Defic. Syndr.* **26**:36–43.
 7. **Shafer, R. W., D. Stevenson, and B. Chan.** 1999. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **27**:348–352.
 8. **Van Laethem, K., A. De Luca, A. Antinori, A. Cingolani, C. F. Perno, and A.-M. Vandamme.** 2002. A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1-infected patients. *Anti-vir. Ther.* **7**:123–129.

Bradley J. Betts
Robert W. Shafer*
Department of Medicine
Stanford University
Stanford, CA 94305

*Phone: (650) 725-2946
Fax: (650) 723-8596
E-mail: rshafer@stanford.edu