Retrovirology

Review

BioAfrica's HIV-1 Proteomics Resource: Combining protein data with bioinformatics tools

Ryan S Doherty*[1], Tulio De Oliveira[1], Chris Seebregts[2], Sivapragashini Danaviah[1], Michelle Gordon[1] and Sharon Cassol[1,3]

Address: [1]Molecular Virology and Bioinformatics Unit, Africa Centre for Health and Population Studies, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa, [2]Biomedical Informatics Research Division, South African Medical Research Council, Cape Town, South Africa and [3]Department of Medical Virology, University of Pretoria, Pretoria, South Africa

Email: Ryan S Doherty* - rsd@ncf.ca; Tulio De Oliveira - tulio.deoliveira@zoology.oxford.ac.uk; Chris Seebregts - chris.seebregts@mrc.ac.za; Sivapragashini Danaviah - Siva.Danaviah@mrc.ac.za; Michelle Gordon - tarinm@nu.ac.za; Sharon Cassol - sharon.cassol@up.ac.za

* Corresponding author

Published: 09 March 2005

Retrovirology 2005, **2**:18 doi:10.1186/1742-4690-2-18

Received: 30 September 2004
Accepted: 09 March 2005

This article is available from: http://www.retrovirology.com/content/2/1/18

Abstract

Most Internet online resources for investigating HIV biology contain either bioinformatics tools, protein information or sequence data. The objective of this study was to develop a comprehensive online proteomics resource that integrates bioinformatics with the latest information on HIV-1 protein structure, gene expression, post-transcriptional/post-translational modification, functional activity, and protein-macromolecule interactions. The BioAfrica HIV-1 Proteomics Resource http://bioafrica.mrc.ac.za/proteomics/index.html is a website that contains detailed information about the HIV-1 proteome and protease cleavage sites, as well as data-mining tools that can be used to manipulate and query protein sequence data, a BLAST tool for initiating structural analyses of HIV-1 proteins, and a proteomics tools directory. The Proteome section contains extensive data on each of 19 HIV-1 proteins, including their functional properties, a sample analysis of HIV-1$_{HXB2}$, structural models and links to other online resources. The HIV-1 Protease Cleavage Sites section provides information on the position, subtype variation and genetic evolution of Gag, Gag-Pol and Nef cleavage sites. The HIV-1 Protein Data-mining Tool includes a set of 27 group M (subtypes A through K) reference sequences that can be used to assess the influence of genetic variation on immunological and functional domains of the protein. The BLAST Structure Tool identifies proteins with similar, experimentally determined topologies, and the Tools Directory provides a categorized list of websites and relevant software programs. This combined database and software repository is designed to facilitate the capture, retrieval and analysis of HIV-1 protein data, and to convert it into clinically useful information relating to the pathogenesis, transmission and therapeutic response of different HIV-1 variants. The HIV-1 Proteomics Resource is readily accessible through the BioAfrica website at: http://bioafrica.mrc.ac.za/proteomics/index.html

Background

Although the HIV-1 genome contains only 9 genes, it is capable of generating more than 19 gene products. These products can be divided into three major categories: structural and enzymatic (Gag, Pol, Env); immediate-early regulatory (Tat, Rev and Nef), and late regulatory (Vif, Vpu, Vpr) proteins. Tat, Rev and Nef are synthesized from small multiply-spliced mRNAs; Env, Vif, Vpu and Vpr are generated from singly-spliced mRNAs, the Gag and Gag-Pol precursor polyproteins are synthesized from full-length mRNA. The matrix (p17), capsid (p24) and nucleocapsid (p7) proteins are produced by protease cleavage of Gag and Gag-Pol, a fusion protein derived by ribosomal frame-shifting. Cleavage of Nef generates two different protein isoforms; one myristylated, the other non-myristylated. The viral enzymes (protease, reverse transcriptase, RNase H and integrase) are formed by protease cleavage of Gag-Pol. Alternative splicing, together with co-translational and post-translational modification, leads to additional protein variability [1].

Phylogenetic analysis, on its own, provides little information about the conformational, immunological and functional properties of HIV-1 proteins, but instead, focuses on the evolution and historical significance of sequence variants. To understand the clinical significance of genetic variation, sequence analysis needs to be combined with methods that assess change in the structural and biological properties of HIV-1 proteins. At present, information and tools for the systematic analysis of HIV-1 proteins are limited, and are scattered across a wide-range of online resources [2,3]. To facilitate studies of the biological consequences of genetic variation, we have developed an integrated, user-friendly proteomics resource that integrates common approaches to HIV-1 protein analysis (Figure 1). We are currently using this resource to better understand the structure-function relationships underlying the emergence of antiretroviral drug resistance, and to examine the process of immune escape from cytotoxic T-lymphocytes (CTLs).

We have categorized the Proteomics Resource into the following main subject headings (Figure 2 &3):

1. ***HIV Proteome*** – Information about structure and sequence, as well as references and tutorials, for each of the HIV-1 proteins (Figure 4);

2. ***HIV-1 Cleavage Sites*** – Information about the position and sequence of HIV-1 Gag, Pol and Nef cleavage sites (Figure 5);

3. ***HIV Protein Data Mining Tool*** – Application for detecting the characteristics of HIV-1 M group isolate (subtype A to K) proteins using information available in public databases and tools (Figure 6);

4. ***HIV Structure BLAST*** – Similarity search for analyzing HIV protein sequences with corresponding structural data (Figure 7);

5. ***Proteomics Online Tools*** – Directory of data resources and tools available for both protein sequence and protein structure analyses of HIV (Figure 8 &9).

The proteome link

In the HIV-1 Proteome section, each of the 19 HIV-1 proteins has a webpage that is divided into six parts: "general overview", "genomic location", "domains/folds/motifs", "protein-macromolecule interactions", "primary and secondary database entries", and "references and recommended readings" (Figure 4). The overview includes a description of the protein, a list of known isoforms, a representative tertiary structure animated image (GIF format) of the protein and its co-ordinates (PDB format), a link to chime tutorials, if available, and information about cleavage sites, localization, and functional activity. The genomic location section provides information on the location of the sequence relative to the reference sequence, HIV-1_{HXB2} [4], sequence data (fasta format), and information about the length, molecular weight and theoretical isoelectric point (pI) of the protein. The domains/folds/motifs section contains information about functional domains and predicted motifs (glycosylation, myristoylation, amidation, phosphorylation and cell attachment sites) of HIV-1_{HXB2} [4], and provides structural predictions (secondary structure, transmembrane regions, low complexity regions, and coiled-coil regions). The section on protein-macromolecule interactions includes information on protein complexes, protein-protein/DNA/RNA interactions, signal-transduction pathways, and potential interactions with other pathogens. The section on primary and secondary databases contains a list of database entries that are needed to retrieve information on protein structure, nucleotide/amino acid sequence data, protein sequence annotation, proteins with similar sequence and structure (such as Los Alamos National Laboratories HIV Sequence Database and the RCSB Protein Data Bank), as well as information on post-translational modification and protein-protein interactions. A list of key reviews and publications, used in the development of the BioAfrica HIV-1 Proteomics Resource, is provided in the references and recommended readings section. As an example, the proteome webpage for Tat, describes how this protein up-regulates HIV-1 gene expression by interacting with the long-terminal repeat (LTR) of HIV-1, promoting the elongation phase of viral transcription, allowing full-length HIV-1 mRNA transcripts to be produced [5,6] (Figure 10). The webpage also gives information on the structural organization of *tat* gene. The mRNA is derived from spliced exons encoded in two different open reading frames. In HIV-1_{HXB2}, these reading frames are separated by a distance of 2334 nucleotides. Some HIV-1 isolates, including HIV-1_{HXB2}, contain an artifact of laboratory strains consisting of a premature stop codon at position 8424 of exon 2. The presence of

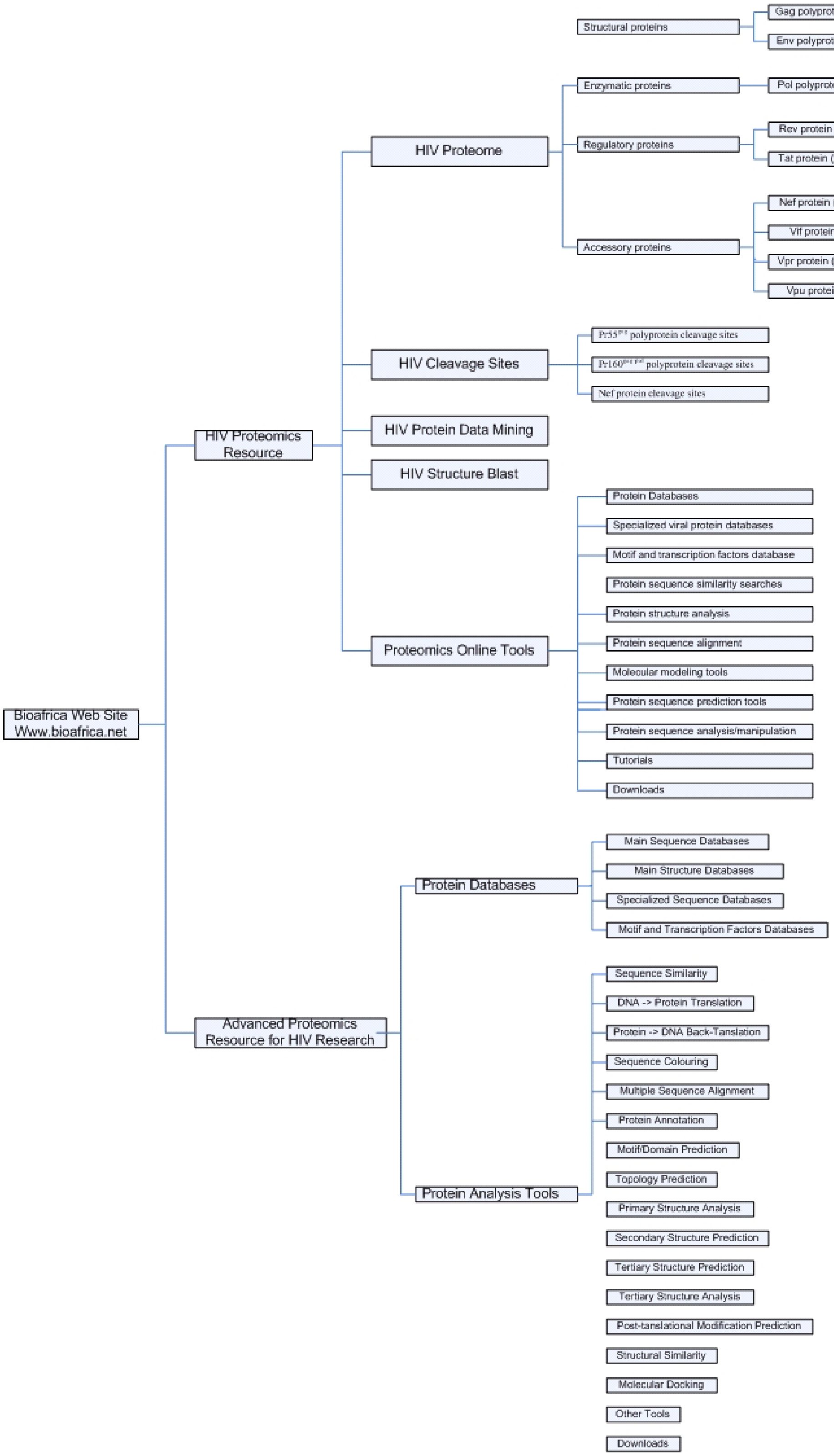


Figure 1
Site map of BioAfrica's HIV-1 Proteomics Resource, showing the separation of Beginner's and the Advanced area of the website, along with all major subject headings.

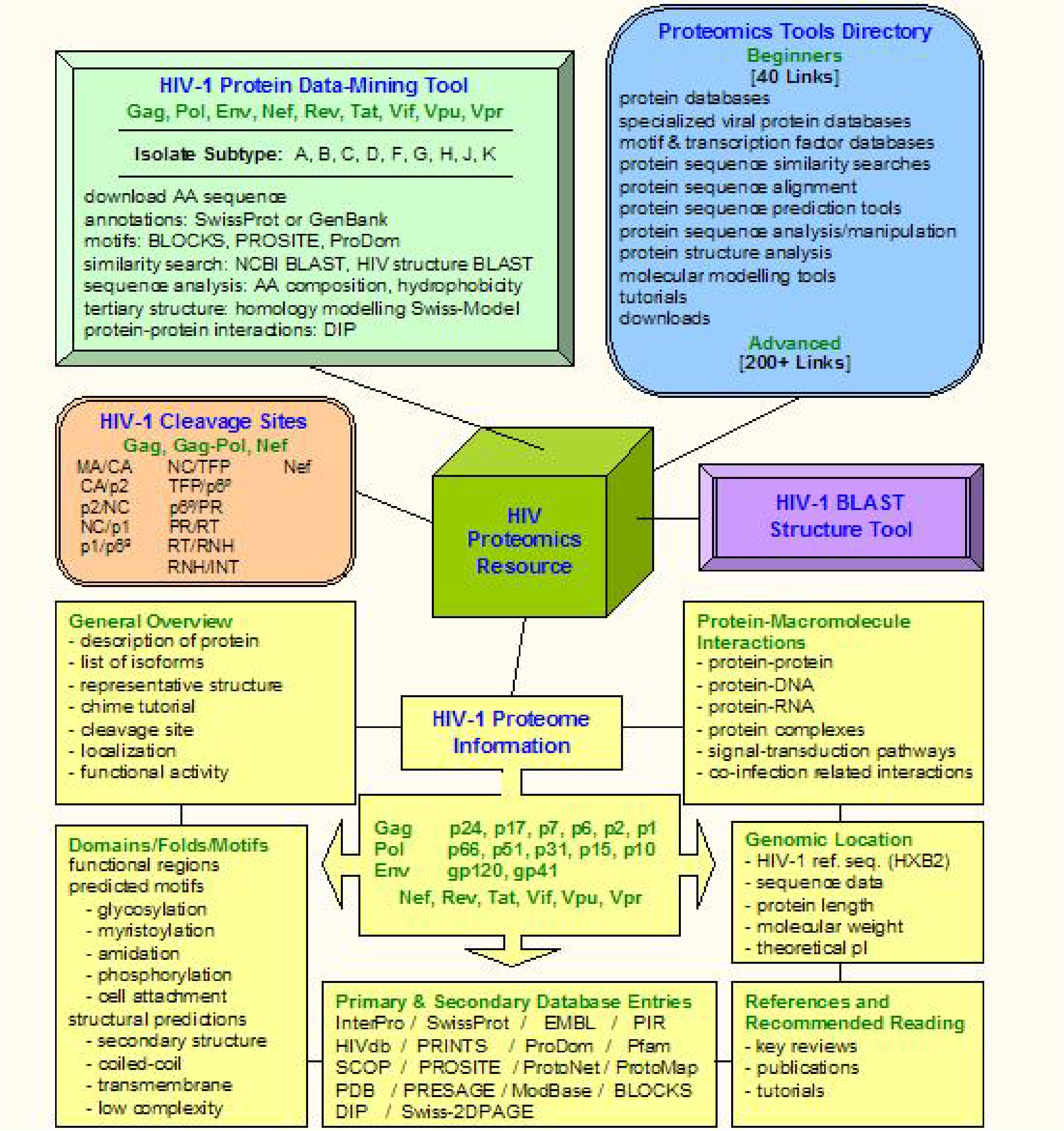
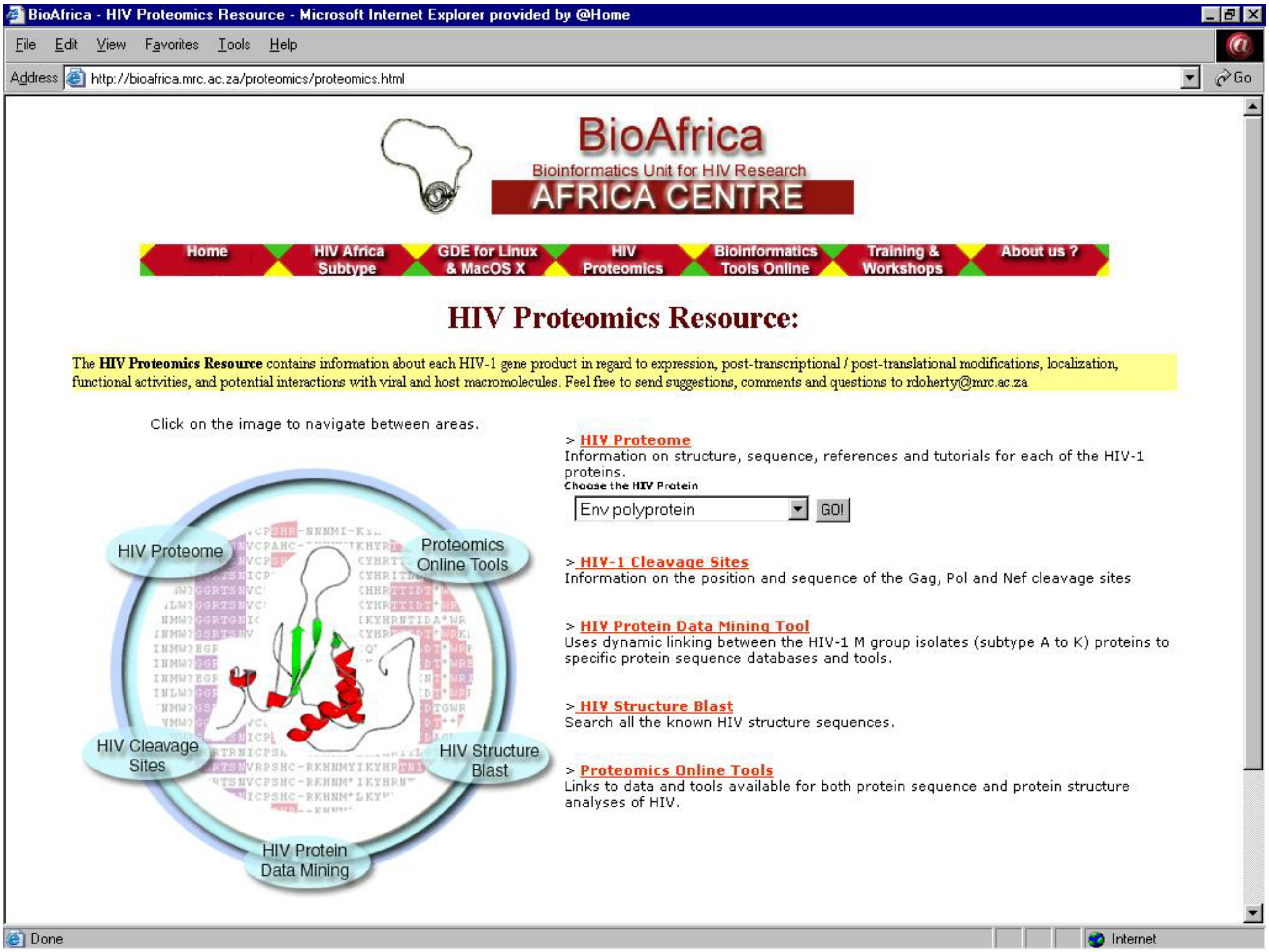


Figure 2
Schematic representation of BioAfrica's HIV-1 Proteomics Resource, showing its five major components: the HIV-1 Proteome (General Overview, Domains/Folds/Motifs, Genomic Location, Protein-Macromolecule Interactions, Primary and Secondary Database Entries, and References and Recommended Readings), the HIV-1 Protease Cleavage Sites section, the HIV-1 Protein Data-mining Tool, the HIV-1 BLAST Structure Tool, and the Proteomics Tools Directory (for Beginners and Advanced investigators).



HIV Proteomics Resource:

The **HIV Proteomics Resource** contains information about each HIV-1 gene product in regard to expression, post-transcriptional / post-translational modifications, localization, functional activities, and potential interactions with viral and host macromolecules. Feel free to send suggestions, comments and questions to rdoherty@mrc.ac.za

Click on the image to navigate between areas.

> **HIV Proteome**
Information on structure, sequence, references and tutorials for each of the HIV-1 proteins.
Choose the HIV Protein

> **HIV-1 Cleavage Sites**
Information on the position and sequence of the Gag, Pol and Nef cleavage sites

> **HIV Protein Data Mining Tool**
Uses dynamic linking between the HIV-1 M group isolates (subtype A to K) proteins to specific protein sequence databases and tools.

> **HIV Structure Blast**
Search all the known HIV structure sequences.

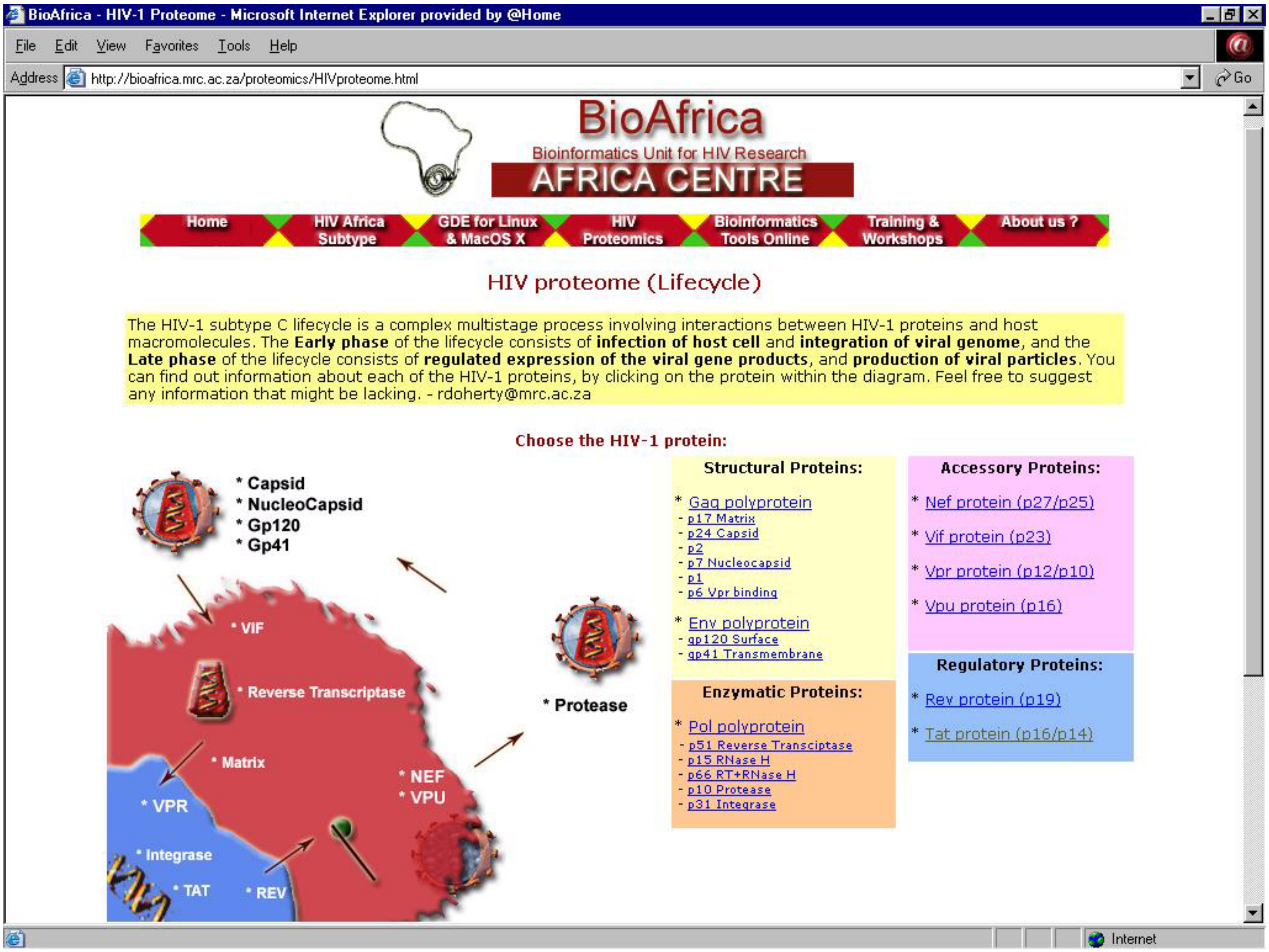
> **Proteomics Online Tools**
Links to data and tools available for both protein sequence and protein structure analyses of HIV.

Figure 3
The central webpage of BioAfrica's HIV Proteomics Resource http://bioafrica.mrc.ac.za/proteomics/index.html

this stop codon leads to the synthesis of a truncated form of Tat that is 86, rather than 101 amino acids in length. The protein has two different isoforms – one translated from early-stage multiply spliced mRNA (p14); the other from singly-spliced mRNA (p16) [7]. Important functional domains include the acidic, amphipathic region (1-**M***E***PV***D***P***R***L***E***PW***KH***P**G*SQ***P***KT***A**-21; the hydrophobic residues are highlighted in bold, and polar residues are italicized) at the N-terminus of the protein; the cysteine-rich disulphide bond region (22-**C**TN**C**Y**C**KK**CC**FH**C**QV**C**-37); the core, basic and glutamine-rich region (49-**R**KK**RRQRRR**AH**Q**NS**Q**TH**Q**ASLSK**Q**-72) that is important for nuclear localization and TAR-binding activity, and the RGD cell-attachment site that binds to cellular integrins. In addition to being expressed in HIV-1-infected cells, Tat is also released into the extracellular fluid where it acts as a growth factor for the development of Kaposi's Sarcoma. Additional information about Tat and its protein-protein interactions can be found on the proteome page of the BioAfrica website located at http://bioafrica.mrc.ac.za/proteomics/TATprot.html.

Protease cleavage sites link

Post-translational cleavage of the Gag, Gag-Pol and Nef precursor proteins occurs at the cell membrane during virion packaging, and is essential to the production of infectious viral particles. Drugs that inhibit this process, the protease inhibitors (PIs), are the most potent antiretroviral agents currently available. Thus it is important to collect information, not only on the sequence of protease enzymes from different HIV-1 subtypes, but also on the natural polymorphisms and resistance mutations that



HIV proteome (Lifecycle)

The HIV-1 subtype C lifecycle is a complex multistage process involving interactions between HIV-1 proteins and host macromolecules. The **Early phase** of the lifecycle consists of **infection of host cell** and **integration of viral genome**, and the **Late phase** of the lifecycle consists of **regulated expression of the viral gene products**, and **production of viral particles**. You can find out information about each of the HIV-1 proteins, by clicking on the protein within the diagram. Feel free to suggest any information that might be lacking. - rdoherty@mrc.ac.za

Choose the HIV-1 protein:

Structural Proteins:

* Gag polyprotein
- p17 Matrix
- p24 Capsid
- p2
- p7 Nucleocapsid
- p1
- p6 Vpr binding

* Env polyprotein
- gp120 Surface
- gp41 Transmembrane

Enzymatic Proteins:

* Pol polyprotein
- p51 Reverse Transciptase
- p15 RNase H
- p66 RT+RNase H
- p10 Protease
- p31 Integrase

Accessory Proteins:

* Nef protein (p27/p25)

* Vif protein (p23)

* Vpr protein (p12/p10)

* Vpu protein (p16)

Regulatory Proteins:

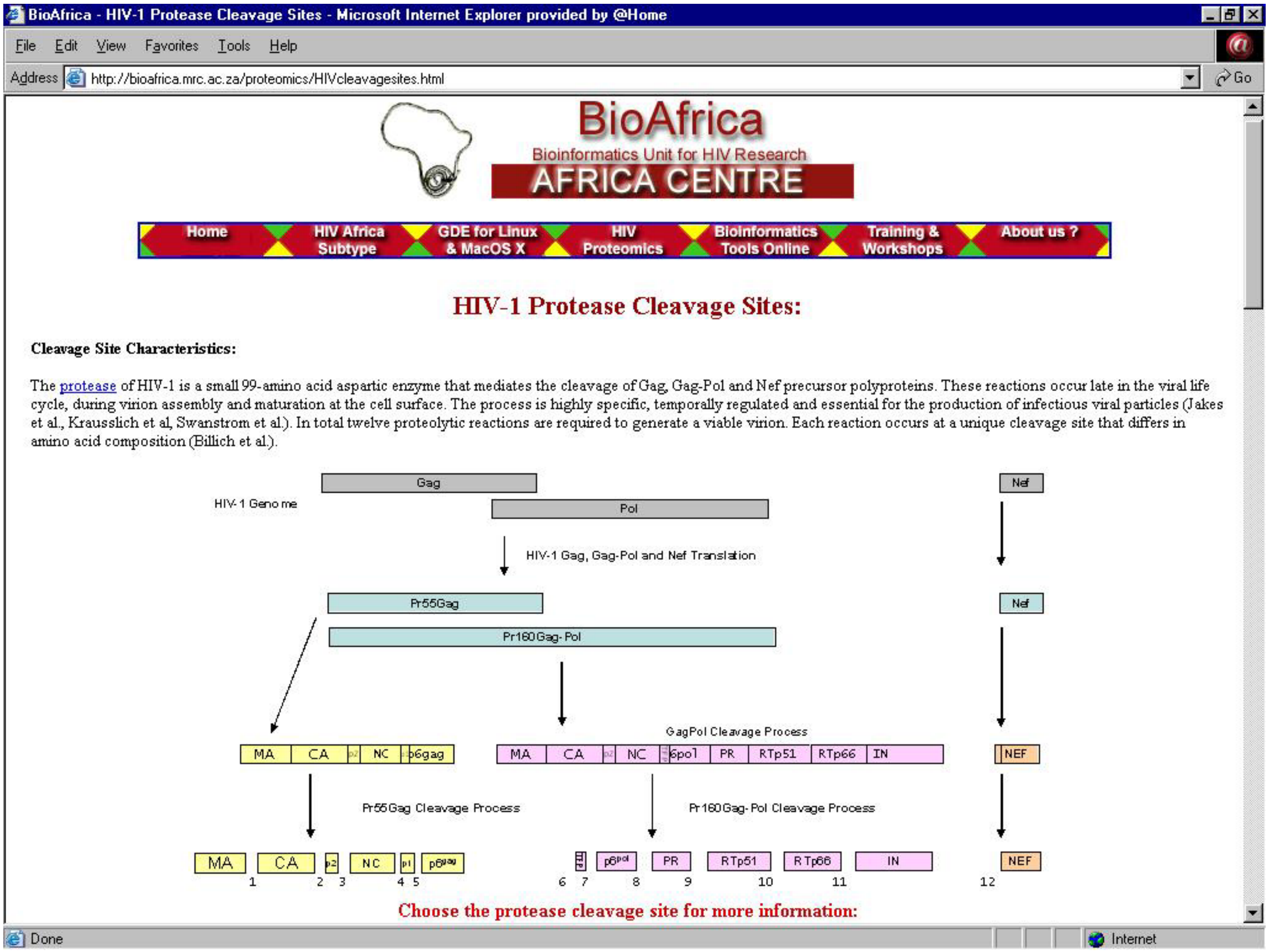
* Rev protein (p19)

* Tat protein (p16/p14)

Figure 4
The central webpage of the HIV-1 Proteome section of the BioAfrica website http://bioafrica.mrc.ac.za/proteomics/HIVproteome.html.

may effect their catalytic activities, drug responsiveness, substrate specificities, and cleavage site characteristics. Studies have shown that resistance mutations in the protease of subtype B are associated with impaired proteolytic processing and decreased enzymatic activity, and that compensatory mutations at Gag and Gag-Pol cleavage sites can partially overcome these defects [8]. These findings suggest that variation at protease cleavage sites may play an important role, not only in regulation of the viral life cycle, but also in disease progression and response to therapy.

The cleavage site section of the BioAfrica webpage is the direct extension of a recent publication in the Journal of Virology describing the location and variability of protease cleavage sites [9] (Figure 5). Together, these two resources provide information on the structure, amino acid composition, genetic variation and evolutionary history of protease cleavage sites, and on the natural selection pressures exerted on these sites. The section also serves as a baseline for understanding the impact of natural polymorphisms and resistance mutations on the catalytic efficiency of the protease enzyme, and on its ability to recognize and cleave individual Gag, Gag-Pol and Nef substrates. Such studies are important for understanding the mechanisms underlying the emergence of PI-induced drug resistance, and for designing alternative, optimized therapies.



HIV-1 Protease Cleavage Sites:

Cleavage Site Characteristics:

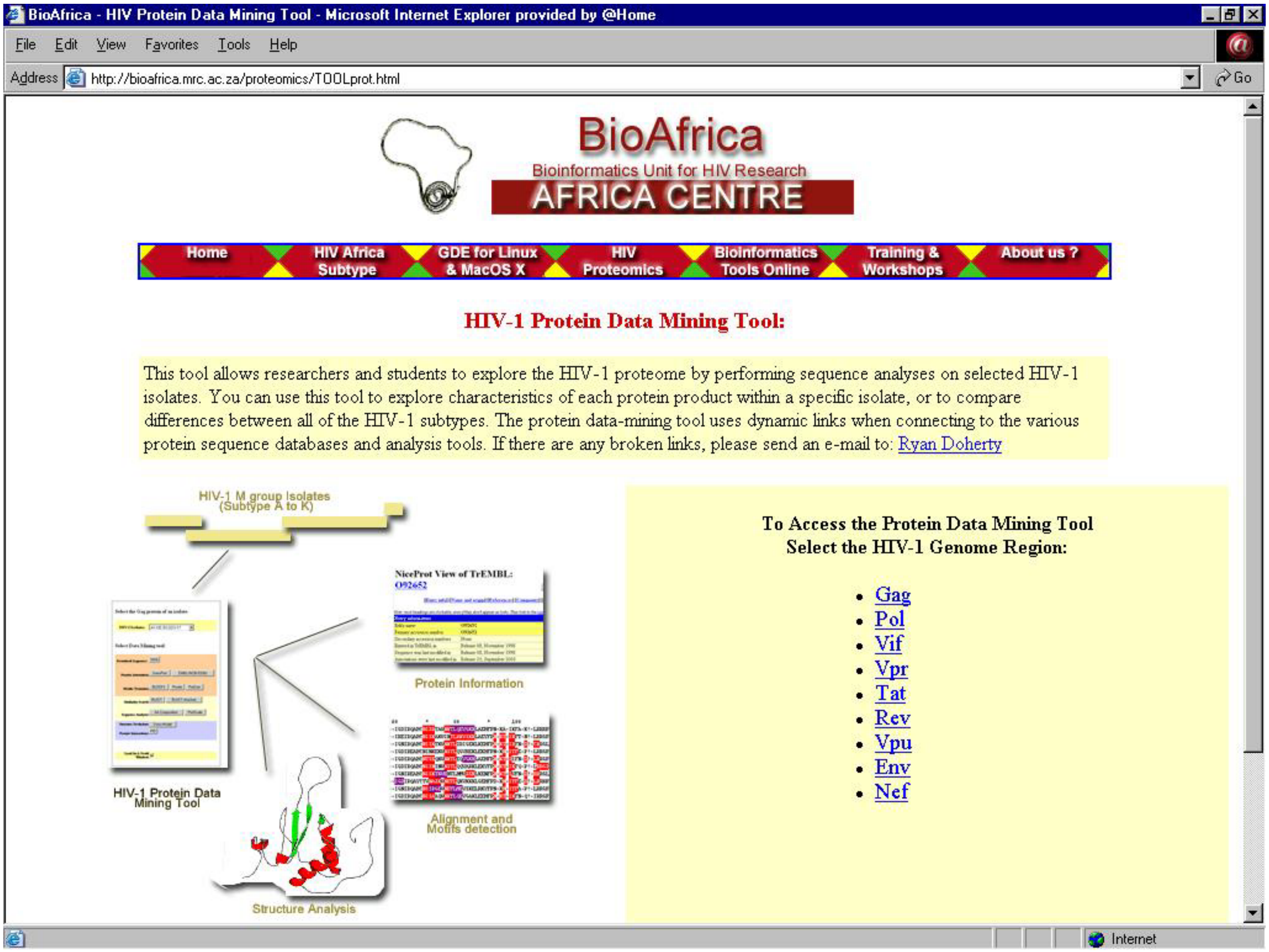
The protease of HIV-1 is a small 99-amino acid aspartic enzyme that mediates the cleavage of Gag, Gag-Pol and Nef precursor polyproteins. These reactions occur late in the viral life cycle, during virion assembly and maturation at the cell surface. The process is highly specific, temporally regulated and essential for the production of infectious viral particles (Jakes et al., Krausslich et al, Swanstrom et al.). In total twelve proteolytic reactions are required to generate a viable virion. Each reaction occurs at a unique cleavage site that differs in amino acid composition (Billich et al.).

Choose the protease cleavage site for more information:

Figure 5
The HIV-1 Protease Cleavage Sites section of the BioAfrica website http://bioafrica.mrc.ac.za/proteomics/HIVcleavagesites.html.

Protein data-mining tools link

The HIV-1 Protein Data-Mining Tool contains twelve sequence analysis techniques for assessing protein variability among different strains of HIV-1 (Figure 6). These tools allow the user to manipulate, analyze and compare published [9-12] and newly-acquired data in a user-friendly, hands-on manner. The analysis is initiated by selecting a particular subset of HIV-1 proteins, either from the user's database, or from the representative dataset of group M viruses (subtypes A through K). Using this dataset, the investigator can then perform a variety of protein-specific analyses. With a single click of the mouse, users can download the amino acid sequence in fasta format; obtain sequence annotations from SwissProt [13] or GenBank [14]; identify functional motifs using BLOCKS [15], PROSITE [16] or ProDom [17]; perform similarity searches using the BLAST program available at Genbank [18], conduct structural comparisons using the BioAfrica BLAST Structure program; determine amino acid composition, predict hydrophobicity and tertiary structure using the Swiss-Model homology modelling server [19], and obtain a list of potential protein-macromolecule interactions from the Database of Interacting Proteins (DIP) [20]. A representative analysis of HIV-1 Tat is shown in Additional file 1. The selected dataset, consisting of eight reference strains – four subtype B (HXB2-1983-France, RF-1983-US, JRFL-1986-US, WEAU160-1990-US) and four subtype C (92BR025-1992-Brazil, 96BW0502-1996-Botswana, TV002c12-1998-SouthAfrica, TV001c8.5-1998-SouthAfrica) isolates – were analyzed using PROSITE



HIV-1 Protein Data Mining Tool:

This tool allows researchers and students to explore the HIV-1 proteome by performing sequence analyses on selected HIV-1 isolates. You can use this tool to explore characteristics of each protein product within a specific isolate, or to compare differences between all of the HIV-1 subtypes. The protein data-mining tool uses dynamic links when connecting to the various protein sequence databases and analysis tools. If there are any broken links, please send an e-mail to: Ryan Doherty

To Access the Protein Data Mining Tool
Select the HIV-1 Genome Region:

- Gag
- Pol
- Vif
- Vpr
- Tat
- Rev
- Vpu
- Env
- Nef

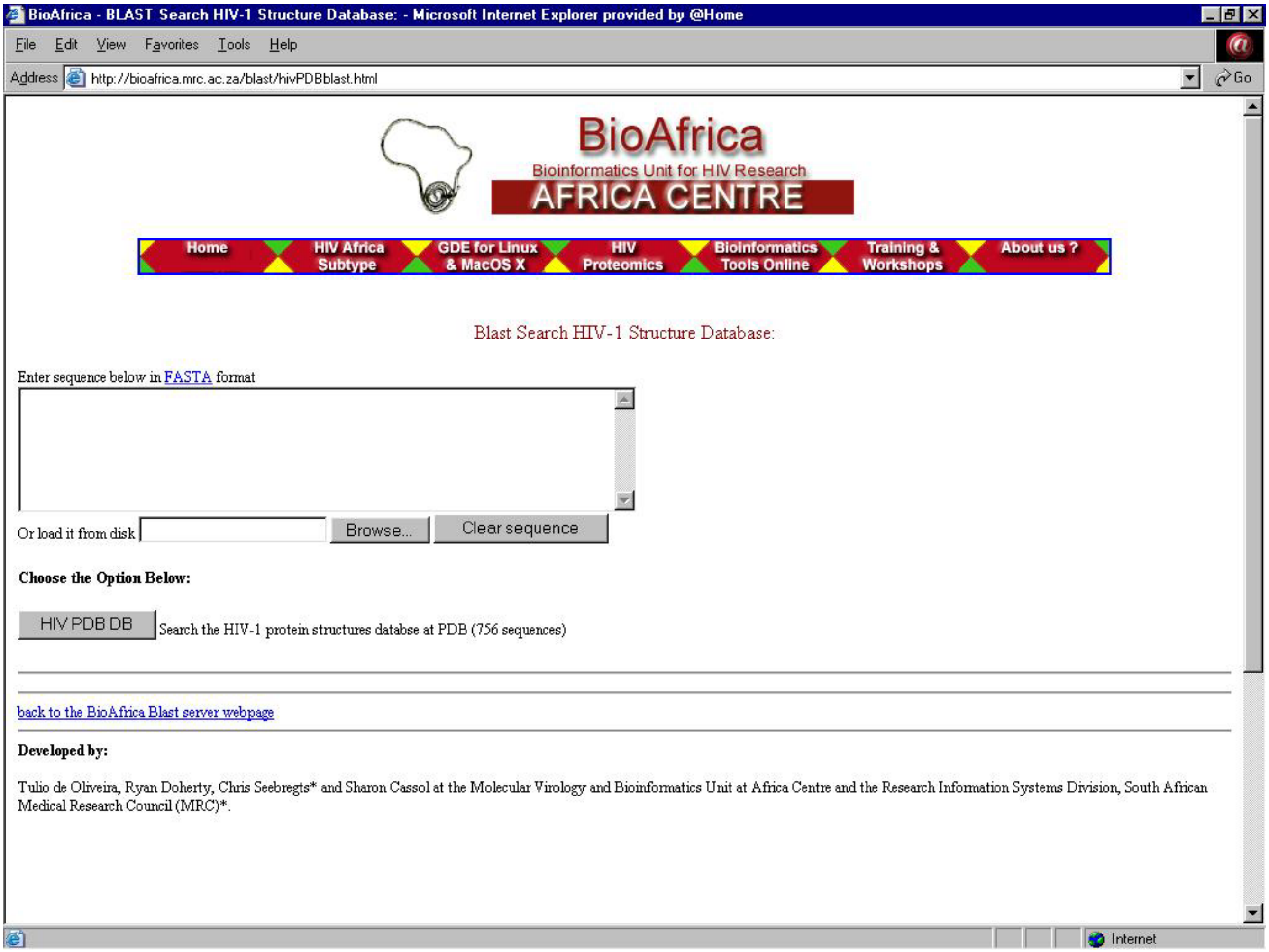
Figure 6

The central webpage of the HIV-1 Protein Data Mining Tool section of the BioAfrica website, where a specific HIV-1 genomic region is selected to be analyzed http://bioafrica.mrc.ac.za/proteomics/TOOLprot.html.

[16]. As shown in Additional file 1, all eight isolates had identical amidation, cysteine-rich and myristylation motifs at amino acid codons 47–50, 22–37 and 44–49, respectively. Three (75%) of the B isolates contained a second myristylation site at codons 42–47, as did three (75%) subtype C viruses. One (25%) of the C viruses carried an extra GNptGS myristylation motif at position 79–84. In addition, all four (100%) C isolates contained a novel myristylation motif, GSeeSK, at amino acid position 83–88, that was not present in four B viruses selected for study. However, the most striking difference between the two subtypes was the increased number of phosphorylation motifs in subtype C relative to B viruses. This increase, which occurs in cAMP/cGMP-dependent kinase, protein kinase C (PKC) and casein kinase II (CKII) phosphorylation sites, has been reported previously [21], but the significance of these findings remain to be established. The analysis also highlighted the atypical nature of the HIV-1_{HXB2} isolate, which, in addition to a premature stop codon, contained no cAMP/cGMP, PKC or CKII phosphorylation sites.

The blast structure tool link

The HIV-1 BLAST Structure Tool facilitates the analysis of HIV-1 protein structure by allowing for rapid retrieval of archived structural data stored in the public databases (Figure 7). Users may input any HIV-1 amino acid sequence and obtain a list of similar HIV protein sequences for which structural data have been experimentally determined and deposited into the Protein Data



Blast Search HIV-1 Structure Database:

Enter sequence below in FASTA format

Or load it from disk

Choose the Option Below:

Search the HIV-1 protein structures databse at PDB (756 sequences)

back to the BioAfrica Blast server webpage

Developed by:

Tulio de Oliveira, Ryan Doherty, Chris Seebregts* and Sharon Cassol at the Molecular Virology and Bioinformatics Unit at Africa Centre and the Research Information Systems Division, South African Medical Research Council (MRC)*.

Figure 7

The BLAST HIV-1 protein structure similarity search is an online tool that searches for all protein structure data within the PDB that have an amino acid sequence similar to the query sequence http://bioafrica.mrc.ac.za/blast/hivPDBblast.html.

Bank (PDB) [22]. After downloading the data from the PDB, subsequent structural analyses can be performed using the software programs and web-servers listed in the Proteomics Tools Directory. For example, a query using an amino acid sequence of HIV-1 Integrase protein from NCBI (gi|15553624|gb|AAL01959.1) results in a list of 54 structural models (ie. PDB_ID|1K6Y) within the PDB. Each of these structural models can be retrieved from the PDB, and the most appropriate structural model could be used for generating a homology model using the query protein sequence.

The proteomics tools directory link

The HIV-1 Proteomics Tools Directory is divided into two web pages. The initial webpage is a concise compilation of some of the most commonly used protein-specific Internet resources (Figure 8). This "beginners" page displays a short list of websites for each of the following twelve categories: "protein databases", "specialized viral-protein databases", "motif and transcription factor databases", "protein sequence similarity searches", "protein sequence alignment", "protein sequence prediction tools", "protein sequence analysis", "protein sequence manipulation", "protein structure analysis", "molecular modelling tools", "tutorials", and "downloads". In addition, the Proteomics Tools Directory has an advanced web page for users who are looking for alternative, or more specialized, protein analysis tools (Figure 9). The advanced webpage displays a list of more than 200 links to different websites and web-servers. These data sources contain a variety of

Proteomics Resources for HIV Research

The **BioAfrica Proteomics Resource** has been created to direct HIV clinicians and scientists to data and tools available for both protein sequence and protein structure analyses of HIV. This page contains an introductory-level list of tools and databases. Please click on the **advanced** link if you would like to see a list of additional tools for performing specific bioinformatics tasks. Feel free to contact me if you have any questions, suggestions or remarks; Ryan Doherty (rdoherty@mrc.ac.za)

Protein Databases:

- UniProt - the central repository of protein sequence and function data, created by combining the information contained in Swiss-Prot, TrEMBL, and PIR
- DSSP - Database of Secondary Structure
- SCOP - Structural Classification Of Protein
- PDB - Protein structure Data Bank [Molecules R Us | PDBsum]

Specialized Virus-related Protein Databases:

- Los Alamos Nat'l Labs - HIV Sequence Database - All publicly available HIV sequence data.
- HIV-1/Human Protein Interaction Database - Literature and data relating to all known protein interactions involving each HIV-1 protein.
- VOG - Viral Clusters of Orthologous Groups of proteins at NCBI.
- dbMHC - Major Histocompatibility Complex database at NCBI
- MEROPS - Peptidase database.
- HIV Drug Resistance Database - All published drug-related variations in Protease and Reverse Transcriptase

Motif and Transcription Factors Databases:

- Pfam - Protein Family classification
- ProDom - Protein Domain database
- MOTIF - Dictionary of Protein Sites and Patterns
- PROSITE - Protein Motif Fingerprint Database

Proteomics site map, with external links to main databases and tools.

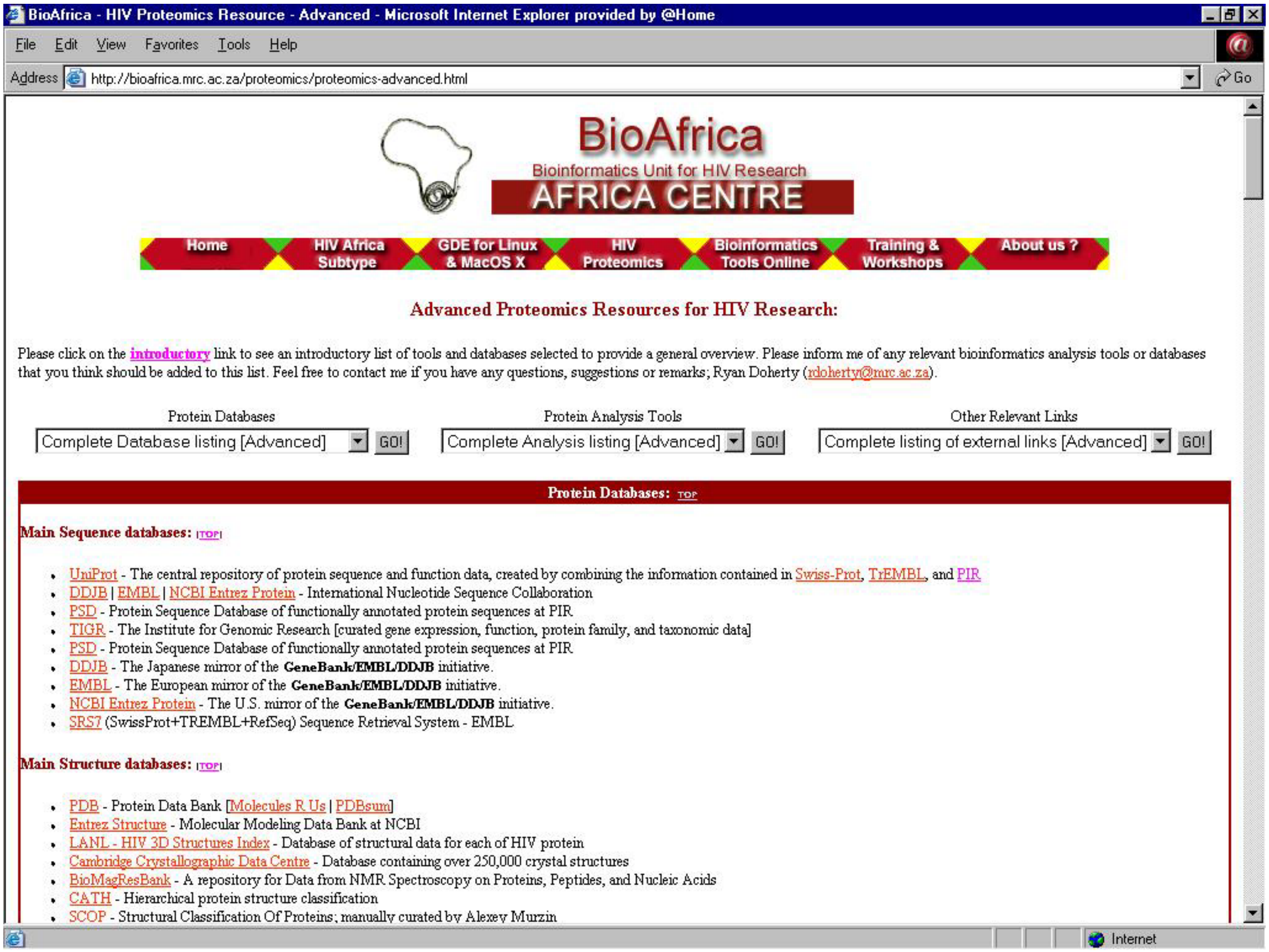
Figure 8
The introductory listing of proteomics resources for HIV research chosen to give a general overview of online tools and databases relevant for the analysis of HIV protein data http://bioafrica.mrc.ac.za/proteomics/proteomicstools.html.

information ranging from specialized protein sequence databases to software programs capable of performing rigid body protein-protein molecular docking simulations.

Conclusion

The impending rollout of antiretroviral therapy to millions of HIV-1-infected people in sub-Saharan Africa provides a unique opportunity to monitor the efficacy of non-B treatment programs from their very inception, and to obtain critical new information for the optimization of treatment strategies that are safe, affordable and appropriate for the developing world. An integral part of this massive humanitarian effort will be the collection of large amounts of clinical and laboratory data, including genetic information on viral subtype and resistance mutations, as well as routine CD4+ T-cell counts and viral load measurements. The mere collection of this data, however, does not ensure that it will be used to its maximum potential. To achieve full benefit from this explosive source of new information, the data will need to be appropriately collated, stored, analyzed and interpreted.

The rapidly emerging field of Bioinformatics has the capacity to greatly enhance treatment (and vaccine) efforts by serving as a bridge between Medical Informatics and Experimental Science. By correlating genetic variation and potential changes in protein structure with clinical risk factors, disease presentation, and differential response to treatment and vaccine candidates, it may be possible to



Advanced Proteomics Resources for HIV Research:

Please click on the introductory link to see an introductory list of tools and databases selected to provide a general overview. Please inform me of any relevant bioinformatics analysis tools or databases that you think should be added to this list. Feel free to contact me if you have any questions, suggestions or remarks; Ryan Doherty (rdoherty@mrc.ac.za).

Protein Databases: Complete Database listing [Advanced] GO!

Protein Analysis Tools: Complete Analysis listing [Advanced] GO!

Other Relevant Links: Complete listing of external links [Advanced] GO!

Protein Databases: TOP

Main Sequence databases: [TOP]

- UniProt - The central repository of protein sequence and function data, created by combining the information contained in Swiss-Prot, TrEMBL, and PIR
- DDJB | EMBL | NCBI Entrez Protein - International Nucleotide Sequence Collaboration
- PSD - Protein Sequence Database of functionally annotated protein sequences at PIR
- TIGR - The Institute for Genomic Research [curated gene expression, function, protein family, and taxonomic data]
- PSD - Protein Sequence Database of functionally annotated protein sequences at PIR
- DDJB - The Japanese mirror of the **GeneBank/EMBL/DDJB** initiative.
- EMBL - The European mirror of the **GeneBank/EMBL/DDJB** initiative.
- NCBI Entrez Protein - The U.S. mirror of the **GeneBank/EMBL/DDJB** initiative.
- SRS7 (SwissProt+TREMBL+RefSeq) Sequence Retrieval System - EMBL

Main Structure databases: [TOP]

- PDB - Protein Data Bank [Molecules R Us | PDBsum]
- Entrez Structure - Molecular Modeling Data Bank at NCBI
- LANL - HIV 3D Structures Index - Database of structural data for each of HIV protein
- Cambridge Crystallographic Data Centre - Database containing over 250,000 crystal structures
- BioMagResBank - A repository for Data from NMR Spectroscopy on Proteins, Peptides, and Nucleic Acids
- CATH - Hierarchical protein structure classification
- SCOP - Structural Classification Of Proteins; manually curated by Alexey Murzin

Figure 9
The advanced listing of online tools and databases relevant for the analysis of HIV protein data http://bioafrica.mrc.ac.za/proteomics/proteomics-advanced.html.

obtain valuable new insights that can be used to support and guide rationale decision-making, both at the clinical and public health levels. The HIV-1 Proteomics Resource, described in this report, is an initial first step in the development of improved methods for extracting and analyzing genomics data, converting it into biologically useful information related to the structure, function and physiology of HIV-1 proteins, and for assessing the role these proteins play in disease progression and response to therapy. The Resource, developed at the Molecular Virology and Bioinformatics Unit of the Africa Centre of Health and Population Studies, is a centralized user-friendly database that is easily accessed through the BioAfrica website at http://bioafrica.mrc.ac.za/proteomics[23].

List of abbreviations used

AA – Amino Acid

BLAST – Basic Local Alignment Search Tool

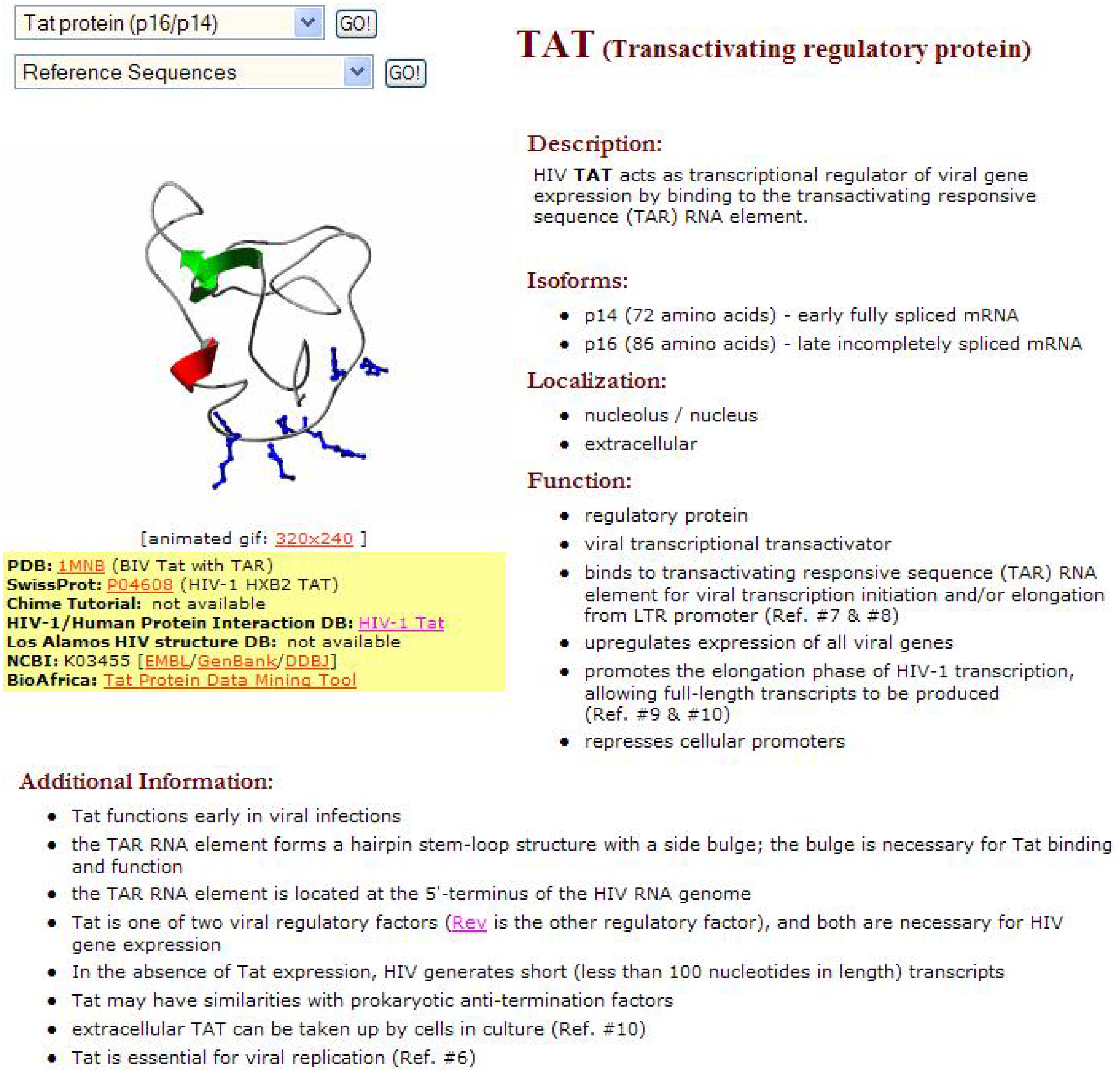
CKII – casein kinase II

CTLs – cytotoxic T-lymphocytes

DIP – Database of Interacting Proteins

DNA – deoxyribonucleic acid

Env – envelope glycoprotein



TAT (Transactivating regulatory protein)

Description:

HIV **TAT** acts as transcriptional regulator of viral gene expression by binding to the transactivating responsive sequence (TAR) RNA element.

Isoforms:

- p14 (72 amino acids) - early fully spliced mRNA
- p16 (86 amino acids) - late incompletely spliced mRNA

Localization:

- nucleolus / nucleus
- extracellular

Function:

- regulatory protein
- viral transcriptional transactivator
- binds to transactivating responsive sequence (TAR) RNA element for viral transcription initiation and/or elongation from LTR promoter (Ref. #7 & #8)
- upregulates expression of all viral genes
- promotes the elongation phase of HIV-1 transcription, allowing full-length transcripts to be produced (Ref. #9 & #10)
- represses cellular promoters

[animated gif: 320x240]

PDB: 1MNB (BIV Tat with TAR)
SwissProt: P04608 (HIV-1 HXB2 TAT)
Chime Tutorial: not available
HIV-1/Human Protein Interaction DB: HIV-1 Tat
Los Alamos HIV structure DB: not available
NCBI: K03455 [EMBL/GenBank/DDBJ]
BioAfrica: Tat Protein Data Mining Tool

Additional Information:

- Tat functions early in viral infections
- the TAR RNA element forms a hairpin stem-loop structure with a side bulge; the bulge is necessary for Tat binding and function
- the TAR RNA element is located at the 5'-terminus of the HIV RNA genome
- Tat is one of two viral regulatory factors (Rev is the other regulatory factor), and both are necessary for HIV gene expression
- In the absence of Tat expression, HIV generates short (less than 100 nucleotides in length) transcripts
- Tat may have similarities with prokaryotic anti-termination factors
- extracellular TAT can be taken up by cells in culture (Ref. #10)
- Tat is essential for viral replication (Ref. #6)

Figure 10
A general overview of the HIV-1 Proteome section of the BioAfrica website, as exemplified by the Tat web page http://bioafrica.mrc.ac.za/proteomics/TATprot.html.

Gag – group-specific antigen polyprotein

GIF – Graphics Interchange Format

HIV – Human Immunodeficiency Virus

HIV-1 – Human Immunodeficiency Virus Type-1

HTTP – Hypertext Transfer Protocol

LTR – long-terminal repeat

mRNA – messenger RNA

NCBI – National Center for Biotechnology Information

Nef – negative factor

PDB – Protein Data Bank

pI – isoelectric point

PIs – protease inhibitors

PKC – protein kinase C

Pol – polymerase polyprotein

Rev – ART/TRS anti-repression transactivator protein

RNA – ribonucleic acid

RNase H – ribonuclease H

Tat – transactivating regulatory protein

Vif – virion infectivity factor

Vpr – viral protein R

Vpu – viral protein U

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

RSD created and maintains BioAfrica's HIV proteomics resource, HIV proteome section, proteomics tools directory, HIV-1 protein data-mining tool and HIV structure BLAST tool; performed protein sequence and structural model analyses; and wrote the manuscript.

TDO conceived and maintains the BioAfrica website, and continues to oversee its rapid expansion; created the cleavage sites section; and participated in the design and implementation of the HIV proteomics resource.

CS participated in the design of the HIV proteomics resource, with an emphasis on the proteomics tools directory.

SD participated in the design and creation of the HIV proteome section, with an emphasis on the HIV-1 Tat protein.

MG participated in the design of the HIV proteomics resource, with an emphasis on the HIV proteome section.

SC supervised the project, and participated in the design and implementation of the HIV proteomics resource.

All authors read and approved the final manuscript.

Additional material

Additional File 1

A table containing a comparative summary of potential functional motifs (cysteine-rich region, myristoylated Asparagine, amidation, cAMP- and cGMP- dependent kinase phosphorylation, Protein Kinase C phosphorylation, and Casein Kinase II phosphorylation) in the HIV-1 Tat proteins of subtypes B and C, as identified using PROSITE.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-2-18-S1.jpeg]

Acknowledgements

Development of the Bioafrica HIV-1 Proteomics Resource was supported by a program grant from the Wellcome Trust U.K. (#061238). The website is hosted by the South African Medical Research Council (MRC).

References

1. Freed EO: **HIV-1 replication.** *Somat Cell Mol Genet* 2001, **26:**13-33.
2. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LSL: **UniProt: The Universal Protein knowledgebase.** *Nucleic Acids Res* 2004, **32:**D115-119.
3. Kuiken C, Korber B, Shafer RW: **HIV sequence databases.** *AIDS Rev* 2003, **5:**52-61.
4. Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K: **Complete nucleotide sequence of the AIDS virus, HTLV-III.** *Nature* 1985, **313:**277-284.
5. Kao SY, Calman AF, Luciw PA, Peterlin BM: **Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product.** *Nature* 1987, **330:**489-493.
6. Feinberg MB, Baltimore D, Frankel AD: **The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation.** *Proc Natl Acad Sci USA* 1991, **88:**4045-4049.
7. Cullen BR: **Human Immunodeficiency Virus as a Prototypic Complex Retrovirus.** *J Virol* 1991, **65:**1053-1056.
8. Mammano F, Petit C, Clavel F: **Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients.** *J Virol* 1998, **72:**7632-7637.
9. de Oliveira T, Engelbrecht S, van Rensburg EJ, Gordon M, Bishop K, zur Megede J, Barnett SW, Cassol S: **Variability at Human Immunodeficiency Virus Type 1 Subtype C Protease Cleavage Sites: an Indication of Viral Fitness?** *J Virol* 2003, **77:**9422-9430.
10. zur Megede J, Engelbrecht S, de Oliveira T, Cassol S, Scriba TJ, van Rensburg EJ, Barnett SW: **Novel evolutionary analyses of full-length HIV type 1 subtype C molecular clones from Cape Town, South Africa.** *AIDS Res Hum Retroviruses* 2002, **18:**1327-1332.
11. Morgado MG, Guimaraes ML, Galvao-Castro B: **HIV-1 polymorphism: a challenge for vaccine development – a review.** *Mem Inst Oswaldo Cruz* 2002, **97:**143-150.
12. Burns CC, Gleason LM, Mozaffarian A, Giachetti C, Carr JK, Overbaugh J: **Sequence variability of the integrase protein from a diverse collection of HIV type 1 isolates representing several subtypes.** *AIDS Res Hum Retroviruses* 2002, **18:**1031-1041.
13. Bairoch A, Apweiler R: **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Res* 2000, **28:**45-48.
14. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2005, **33(Database Issue):**D34-38.

15. Henikoff JG, Greene EA, Pietrokovski S, Henikoff S: **Increased coverage of protein families with the BLOCKS database servers.** *Nucleic Acids Res* 2000, **28:**228-230.
16. Hulo N, Sigrist CJA, Saux VL, Langendijk-Genevaux PS, Bordoli L, Gattiker A, de Castro E, Bucher P, Bairoch A: **Recent improvements to the PROSITE database.** *Nucleic Acids Res* 2004, **32(Database Issue):**134-137.
17. Servant F, Bru C, Carrere S, Courcelle E, Gouzy J, Peyruc D, Kahn D: **ProDom: Automated clustering of homologous domains.** *Brief Bioinform* 2002, **3:**246-251.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
19. Schwede T, Kopp J, Guex N, Peitsch MC: **SWISS-MODEL: an automated protein homology-modeling server.** *Nucleic Acids Res* 2003, **31:**3381-3385.
20. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: **The Database of Interacting Proteins: 2004 update.** *Nucleic Acids Res* 2004, **32(Database Issue):**449-451.
21. de Oliveira T, Salemi M, Gordon M, Vandamme AM, van Rensburg EJ, Engelbrecht S, Coovadia HM, Cassol S: **Mapping Sites of Positive Selection and Amino Acid Diversification in the HIV Genome: An Alternative Approach to Vaccine Design?** *Genetics* 2004, **167:**1047-1058.
22. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28:**235-242.
23. De Oliveira T, Doherty RS, Seebregts C, Monosi B, Gordon M, Cassol S: **The BioAfrica Website: An Integrated Bioinformatics Website for Studying the Explosive HIV-1 Subtype C Epidemic in Africa.** In *Digital Biology: The Emerging Paradigm Conference, NIH: 6 – 7 November 2003* Maryland, USA.