

VSEngine - Similarity & Property prediction based Virtual Screening Web Services

Dragos Horvath, Gilles Marcou & Alexandre Varnek

UMR 7177 CNRS-ULP

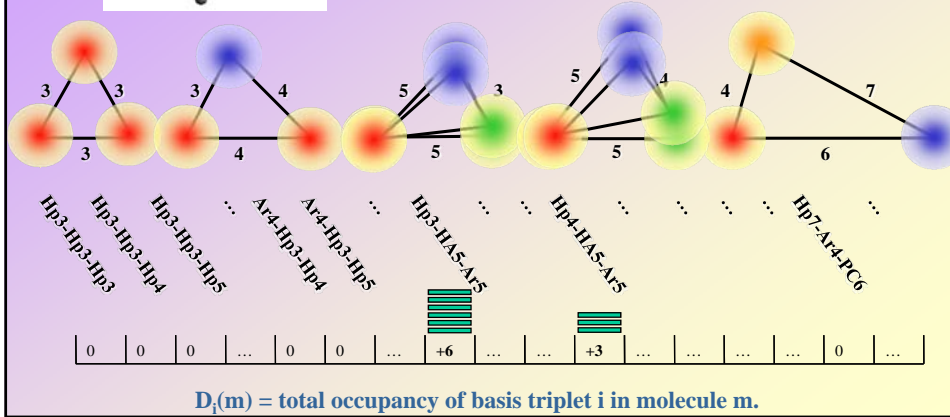
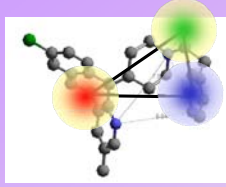
Laboratoire d'InfoChimie, Univ. Louis Pasteur
Strasbourg

horvath@chimie.u-strasbg.fr

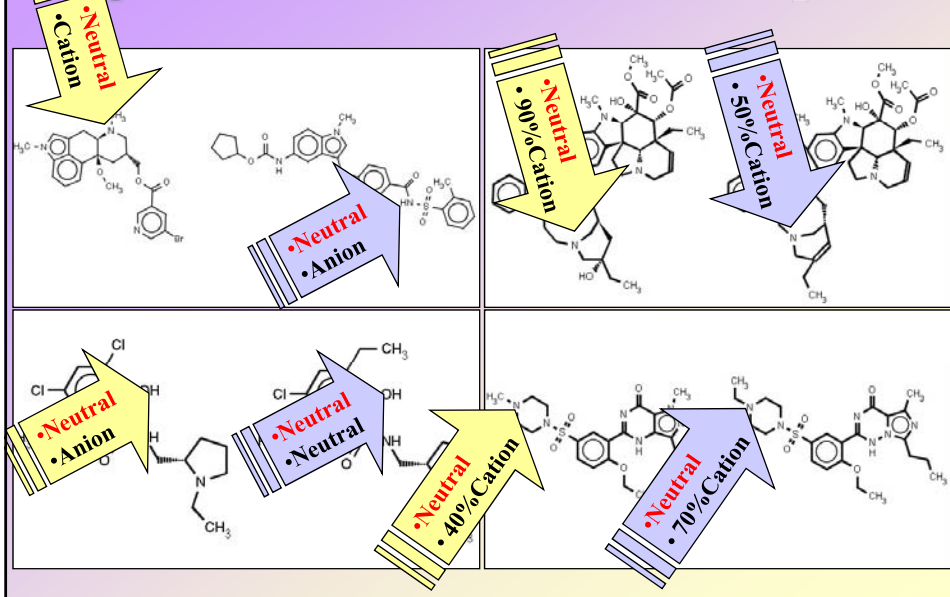
Why yet another « Virtual Lab »?

- To open up the access to the proprietary tools developed within UMR 7177:
 - Tunable Fragments (atom/bond sequence and augmented atom)
 - Fuzzy Pharmacophore Triplets & adapted similarity metrics
 - Original QSAR-building Expertise:
 - “aggressive” problem space searches, supporting non-linear model fitting
 - original applicability domain definitions...
 - potentially useful property prediction tools...
- To make all these tools, issued from different environments & backgrounds, work together
 - ChemAxon “Toolbox” welcome
- But do they work better than “Saint Tanimoto 0.85” ?
 - Don't know – finding it out is the user's part of the deal...

(ChemAxon-Powered) Fuzzy Pharmacophore Triplets

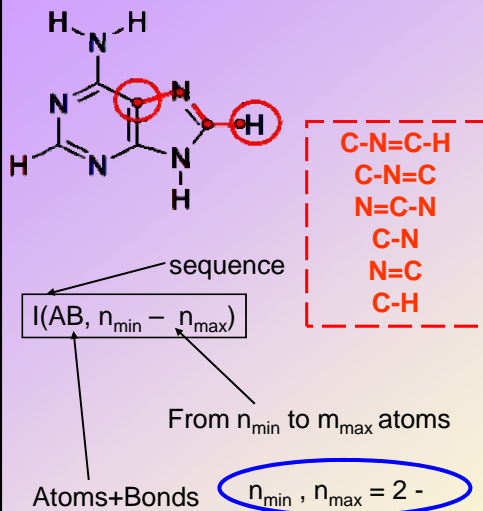


Some 'activity cliffs' in rule-based descriptor space are smoothed out in 2D-FPT-space

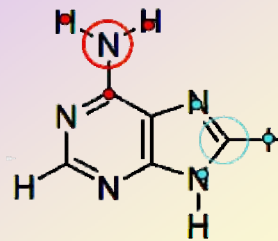


Fragment Descriptors (ISIDA)

Sequences

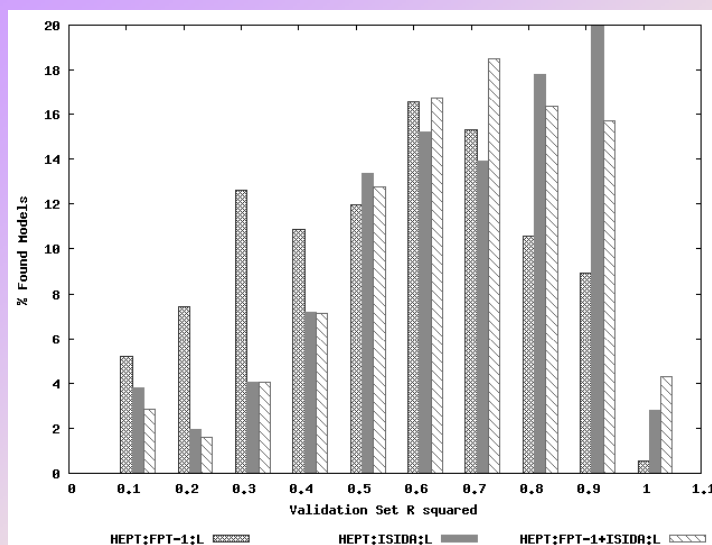


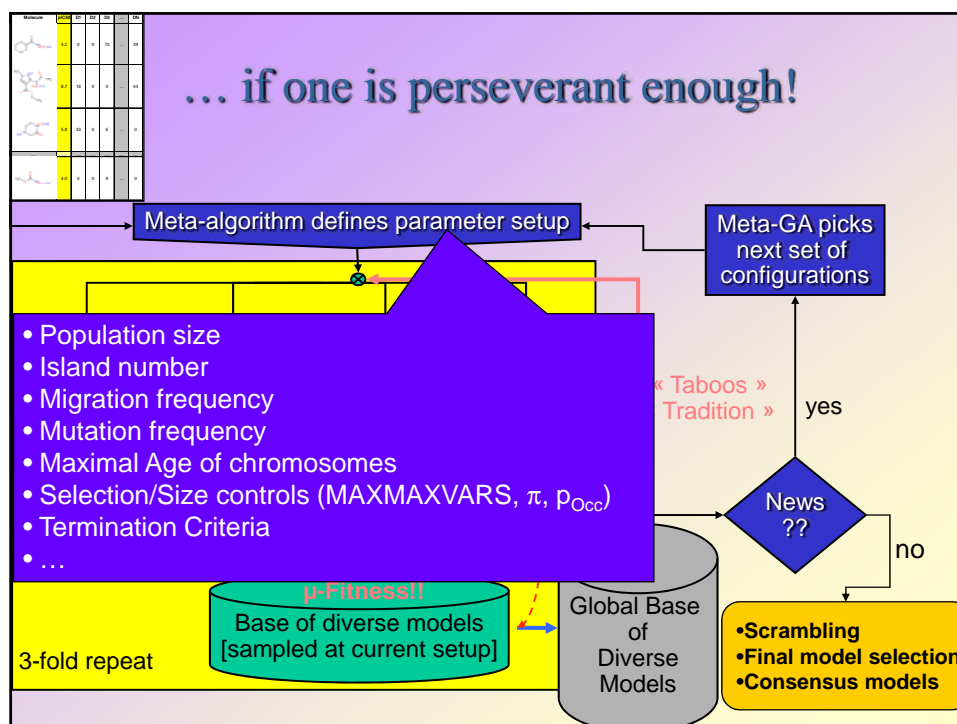
Augmented Atoms



- II(A)** (no hybridization)
- II(Hy)** (hybridization of neighbours is taken into account)

Sharp Fragments and Fuzzy Pharmacophores, together, “crack” many QSAR problems...



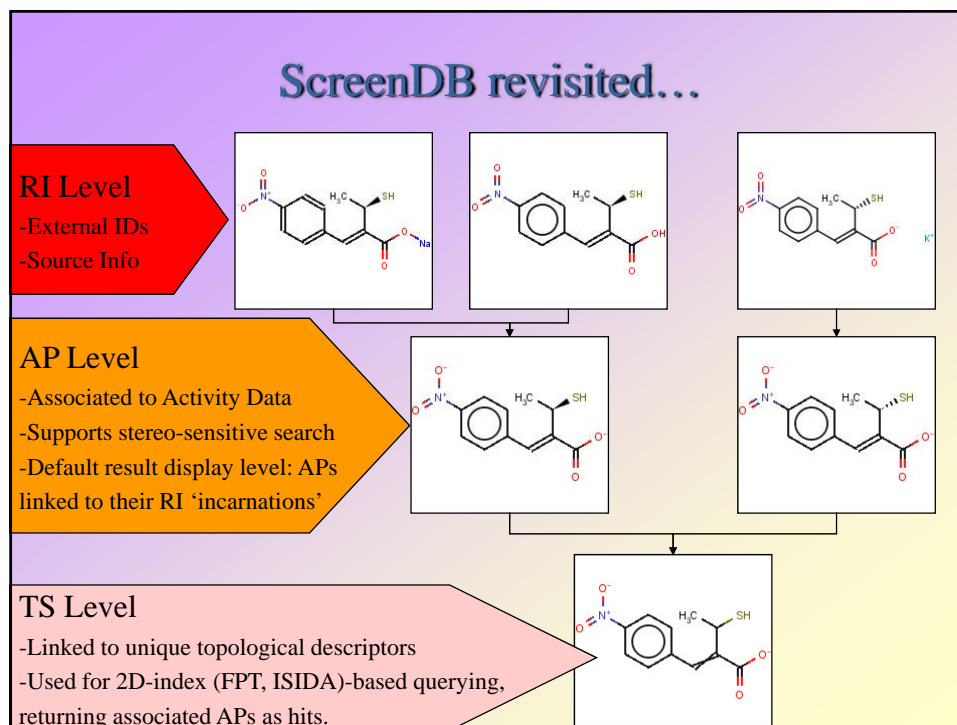


Software integration issues...

- **ISIDA Fragment Descriptors & Tools** originally stem from a Windows/Delphi world
 - ... but Unix standalone versions have been created to facilitate integration
- **The FPT generator** is a standalone tool employing the ChemAxon toolkit...
 - ... though not exploiting any database access facilities offered by the latter
- **Don't want to risk the death of my one & only neuron while learning JSP or other weird Web freak languages...**
- **Perl/cgi and Unix shell scripts** pilot **standalone tools** (Java, Delphi, F77, Perl, awk) having access to a central **MySQL/JChem database**.

ScreenDB, the JChem Database: Standardization, Traceability & Property Calculations

- Three levels of structure representation cohabitate:
 - **RI: Raw Input:** structures as entered by chemists
 - **AP:** Active principles: major microspecies of default tautomer – standardized, without counterions
 - **TS:** Topological Skeletons - AP stripped of stereochemical info (thanks smiles:0!)
- ChemAxon calculated properties (logP, logD, RotBonds, HBA, HBD) entered in the database upon registration
- **Import & consistency check of experimental data fields**
 - QSAR set repository & Hit Set Property Alerts
- **ISIDA and FPT descriptors** are stored in .gzipped files indexed by the JChem cd_id



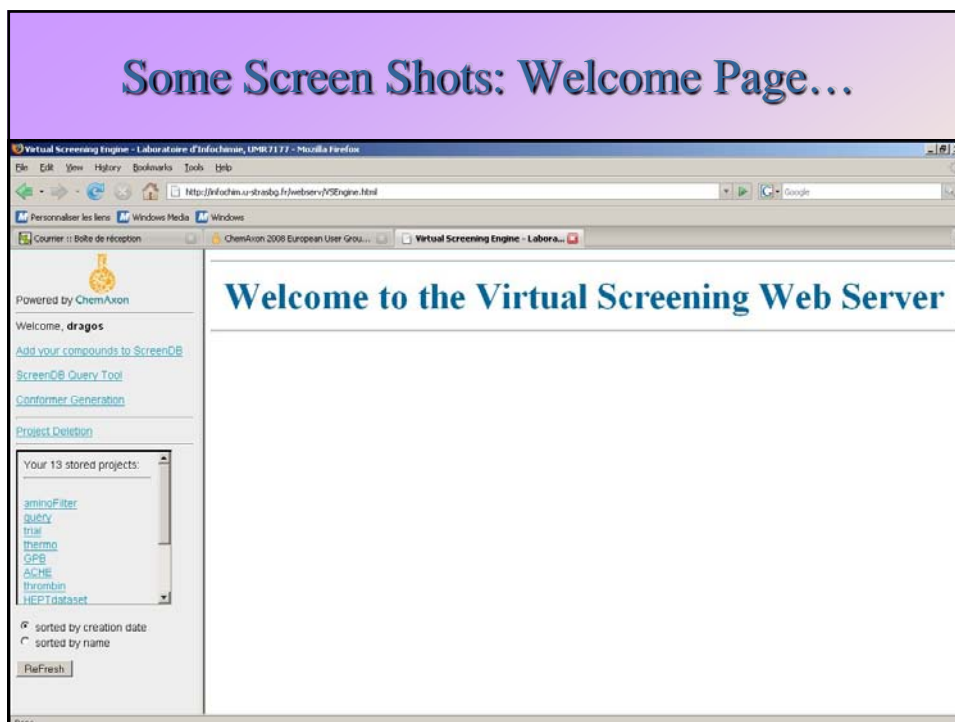
General Site Architecture...

- Apache/CGI server, with user-specific directories (pseudo-accounts)
- Users create server-side project directories, to harbor input requests (query structures, *etc.*), temporary files, flags and results.
- A project is a perennial entry, dedicated to a specific task:
 - Molecule upload into the data base
 - Database Querying
 - *Online property prediction (bypassing database storage)
 - **QSAR model building & predictions
- and associated to a stance-sensitive task-specific .cgi script
 - Interactive parameter input/instantiation of background computing jobs
 - Project job in progress...
 - Project results available or job disaster report...

ScreenDB Querying

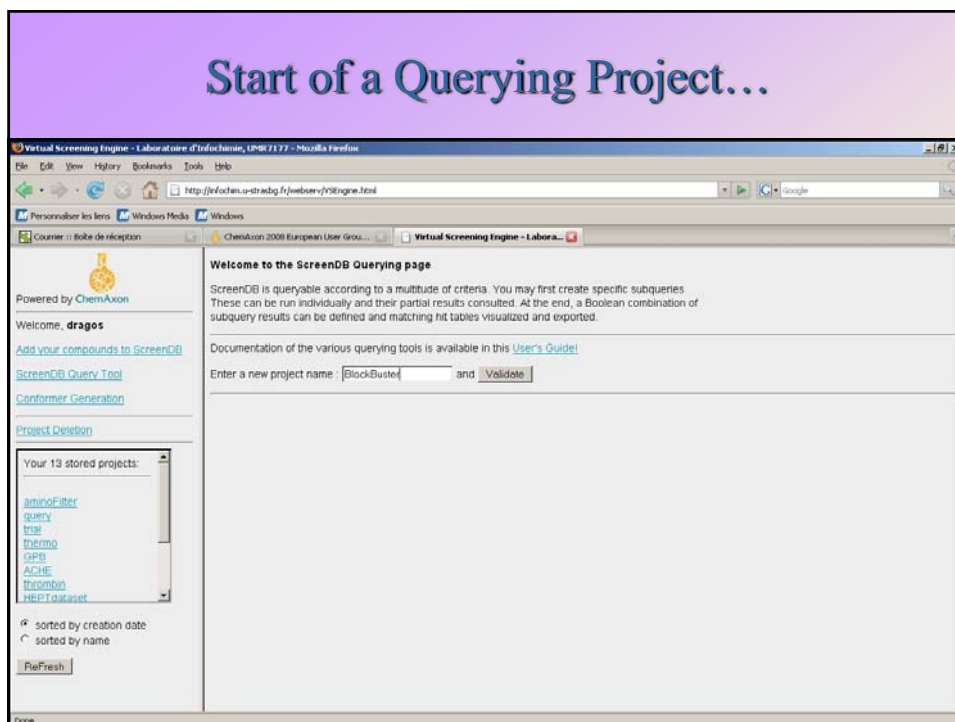
- Data field-based queries (in progress – including on-the-fly predicted properties from external QSARs)
- ChemAxon-driven tools (.cgi-controlled standalone java application)
 - Substructure & Unique (!) Substructure (SMARTS)
 - Exact match or ChemAxon structural fingerprint-based default similarity search (SMILES).
- ISIDA fragment- or Fuzzy Triplet-based similarity queries
 - Support multiple molecules as a query (“AND” fingerprint), in conjunction to various descriptor instances and dissimilarity metrics (background similarity calculations using specific Unix command-line tools)
- A hit post-processing tool converts cd_id lists produced by these applications into hit display pages and hit property alerts...

Some Screen Shots: Welcome Page...



The screenshot shows a web browser window displaying the 'Welcome to the Virtual Screening Web Server' page. The browser's address bar shows the URL <http://infocim.u-strasbg.fr/webse/vsEngine.html>. The page features a sidebar on the left with the ChemAxon logo and a list of 13 stored projects: aminoFilter, query, trial, thermo, GPE, ACHE, thrombin, and HEPTdataset. The main content area displays the title 'Welcome to the Virtual Screening Web Server'.

Start of a Querying Project...



The screenshot shows the 'Start of a Querying Project' page. The browser address bar is the same as in the previous screenshot. The sidebar on the left is identical, showing the 13 stored projects. The main content area is titled 'Welcome to the ScreenDB Querying page' and contains the following text: 'ScreenDB is queryable according to a multitude of criteria. You may first create specific subqueries. These can be run individually and their partial results consulted. At the end, a Boolean combination of subquery results can be defined and matching hit tables visualized and exported.' Below this text, there is a link to the 'User's Guide!' and a form with the label 'Enter a new project name:' followed by an input field containing 'BlockBuster' and a 'Validate' button.

Initiate a Substructure SubQuery

The screenshot shows the Virtual Screening Engine web interface. The browser address bar displays <http://infocten.u-strasbg.fr/webserve/vsengine.html>. The page title is "Virtual Screening Engine - Laboratoire d'Informatique, UPM 7177 - Mozilla Firefox".

On the left sidebar, there is a "Your 13 stored projects:" list containing:

- aminoFilter
- query
- trial
- thermo
- gsp
- ACHE
- thrombin
- HERF dataset

Below the list, there are radio buttons for "sorted by creation date" (selected) and "sorted by name", and a "Refresh" button.

The main content area shows a form for adding a new query. The text reads: "Add new Query entitled [BZO-sub] of type: [DateField] -- [GO!]. A dropdown menu is open, showing options: "DateField", "SubstructureQuery" (highlighted), and "Similarity". Above this form, there is a control: "Allow at most 10 hits per/subquery to be visualized".

Specify the Substructure SubQuery

The screenshot shows the Virtual Screening Engine web interface for specifying a substructure query. The browser address bar displays <http://infocten.u-strasbg.fr/webserve/ChemAxonQuery.cgi?user=dragosproject=Blockbuster&label=BZOsub>. The page title is "BZOsub - Mozilla Firefox".

The main content area shows a form for specifying the query. The text reads: "Select Substructure, Unique Substructure, Exact Match or Dissimilarity Query < 0.2". Below this, there is a dropdown menu for "out of" with options: "All molecules in ScreenDB" (selected), "Existing (Commercial/Academic) only", and "Tested (with experimental data) only".

The central part of the interface features a chemical structure editor. It displays a chemical structure of a pyrimidine derivative (SMILES: C1=NC(=O)NC(=O)N1). The editor includes a toolbar with various tools (select, delete, add, etc.) and a vertical element palette on the right with buttons for H, C, N, O, S, F, P.

At the bottom of the editor, there is a search bar: "Search for molecules matching or NOT matching the Query".

SubQuery in Progress (go start the next one while it completes)...

ChemZoon 2008 European User Group... Virtual Screening Engine - Laboratoire... RZDoub

Select Substructure, Unique Substructure, Exact Match or Dissimilarity Query < 0.2

out of | All molecules in ScreenDB

File Edit View Insert Atom Bond Structure Tools Help

100 %

** Query started on Wed Apr 23 16:02:29 UTC 2008 - Refresh

Applet MSketch started

SubQuery done...

ChemZoon 2008 European User Group... Virtual Screening Engine - Laboratoire... RZDoub in BioBuster : HES RZDoub

Select Substructure, Unique Substructure, Exact Match or Dissimilarity Query < 0.2

out of | All molecules in ScreenDB

File Edit View Insert Atom Bond Structure Tools Help

100 %

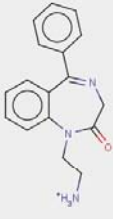
Search for molecules matching or NOT matching the Query

>>> View Results

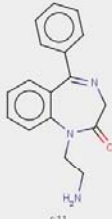
Applet MSketch started

Hit display...

Species ID=697



Alternative Representations/Formulations:



Activity Values associated to species nr. 697

Target	Measure	Value	Source	Entered by	Entered on
BZP	pIC50	6.59	JMedChem 38 715-724 (1995)	dragos	2008-02-21

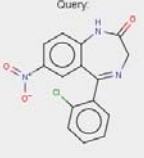

Molecular Identifiers associated to species nr. 697 and its possible alternative formulations

ID	Data Field Type	Data Field Value	Entered by	Entered on
677	CURR_ID	dragos.bzr.74	dragos	2008-02-21

Right click for menu

A Similarity SubQuery ...

80 Hits displayed out of 2589:

Score=0.57 ID=10 >>> [More info](#)

[UPLOAD](#) the .sd file of hits (note - this may be VERY long for > 500 hits - we are mining for all relevant info related to hits through ScreenDB)

Properties differing significantly from averages within the selection:

Property	Global Occurrences	Global average	Occurrences in Selection	Selection average	T Factor
HBD	59417	0.89+-0.94	2589	0.60+-0.58	23.85
BZP-pIC50	147	5.01+-0.14	114	5.12+-0.60	9.90
HBA	59417	5.09+-2.28	2589	4.78+-1.64	9.35
logD74	59417	2.98+-2.21	2589	3.36+-1.22	14.99
RotBonds	59417	4.21+-2.30	2589	3.55+-1.67	19.53

Done

The Complex Query Builder...

80 Hits displayed out of 2522.

FactSheet of Molecule ID=31342

UPLOAD the .sd file of hits (note - this may be VERY long for > 500 hits: tool is mining for all relevant info related to hits through ScreenDB)

Properties differing **significantly** from averages within the selection:

Property	Global Occurrences	Global average	Occurrences in Selection	Selection average	T Factor
HBD	59417	0.89+-0.94	2522	0.59+-0.58	24.08
BZR-piC50	147	5.01+-0.14	55	5.07+-0.46	6.51
HBA	59417	5.09+-2.28	2522	4.80+-1.65	8.48
logD74	59417	2.98+-2.21	2522	3.37+-1.23	15.04
RotBonds	59417	4.21+-2.30	2522	3.60+-1.65	17.94

Conclusions, Perspectives...

- Buzzword FLEXIBILITY: The ChemAxon API is a convenient way to get a heterogeneous bunch of tools talk to a MySQL chemical database – be it by means of temporary files!
 - Standardization & file format conversions are heavily used
- Buzzword CHEMICAL SENSE: structure representation, substructure querying, ionization
- Speed is not exactly a bottleneck, since we are looking at more elaborate and therefore more costly virtual screening options...
- Unix-shell centric approach has the merit to be easily deployable on GRIDS!