

# AUSPYX<sup>®</sup>

STORING AND SEARCHING CHEMICAL DATA WITHIN ORACLE<sup>®</sup>



Tripos' next generation Oracle data cartridge, AusPYX, is built on Oracle's extensibility architecture and enables storage and searching of chemical structures and associated relational data within Oracle environments. Developed and validated in collaboration with our pharmaceutical partners, AusPYX provides a core framework upon which to build high performance cheminformatics applications.

### Key Benefits

- AusPYX provides the core storage and searching technology around which to build Oracle-based cheminformatic applications such as:
  - Registration, Inventory, and Ordering Systems
  - Electronic Laboratory Notebooks
  - Reagent Selection Systems
  - Virtual HTS Systems
- AusPYX's new Learned Index technology increases performance by extracting patterns of user behavior.
- AusPYX is fully integrated with the Oracle Cost Based Optimizer, leveraging Oracle's native method of query optimization.
- AusPYX-based cheminformatics systems allow accurate representation of the isomeric mixtures that occur in real-world reactions.
- AusPYX provides enhanced functionality for accurately finding duplicates within large datasets.
- Through the use of OptiSim<sup>®</sup> technology, AusPYX provides for clustering and diverse selection of compounds directly within the Oracle database.
- AusPYX can be configured to store and search 3rd party fingerprints, such as those produced by Scitegic's<sup>®</sup> Pipeline Pilot.
- Robust translation services allow migration of structures from MDL<sup>®</sup> and other file formats into AusPYX-based systems.

### Superior Stereochemistry Representation

AusPYX extends the SLN language to allow accurate representation of the isomeric mixtures that occur in real-world reactions producing chiral centers. Unresolved stereochemistry and isomeric mixtures are supported by the addition of the modifiers E (explicit), \* (relative), or M (mixture) to the standard stereo qualifiers of N, I, or U.

By adding the stereo mode modifier to the standard stereo attribute, enantiomeric mixtures and unresolved pure isomers can be described and searched. By also including a numeric index, several groups of chiral centers, whose chirality is known only relative to other atoms in the group, can be described. These extensions allow registration of any mixture that might be produced from these reactions into the database.

### Duplicate Checking Accuracy

When migrating one customer's data from legacy data into AusPYX, the AusPYX duplicate checker identified 22,000 duplicates that had not been identified by the legacy system. In many cases, the duplicates were alternate sketches of the same molecule, while some were tautomer duplicates or looked identical.

### Automatic Identification of Meso Compounds

AusPYX has the ability to automatically identify and label meso compounds upon translation from MDL format to SLN while creating an appropriate MOL file when translating back to MDL format. SLN provides a way to identify chiral centers that cannot be identified as R or S such as seen in cis/trans isomers.

### Normalization of Aromatic Structures

AusPYX automatically stores a normalized SLN representation of aromatic systems when compounds are registered. In the same way, when a search is executed, the search criteria

are converted into a similar normalized form. This ensures that searching within AusPYX yields fewer false positives or missed hits.

### Search Operators

AusPYX provides the following operators that extend SQL to allow users to create queries based on chemical structure criteria. These operators are used like any other Oracle operator to create arbitrarily complex SQL statements.

**SEARCH2D ( )** - substructure searching

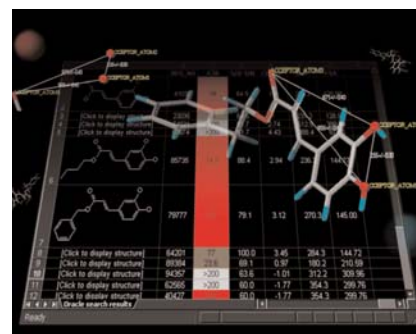
**EXACTMATCH ( )** - facilitates duplicate checking, tautomer search, locate all related salts, locate all related isomers

**SIMILARITY ( )** - locate similar structures based on molecular fingerprints; Tanimoto, asymmetric, and cosine metrics supported

**FORMULA ( )** - locate structures based on molecular formula

**POLYMER ( )** - search stored polymers

**FILTER ( )** - apply various filters and transformations to SLN tables or strings



Tripos' Oracle data cartridge, AusPYX, is built on Oracle's extensibility architecture and enables storage and searching of chemical structures and associated relational data within Oracle environments.

**Complementary Products**

■ **Benchware® Notebook**

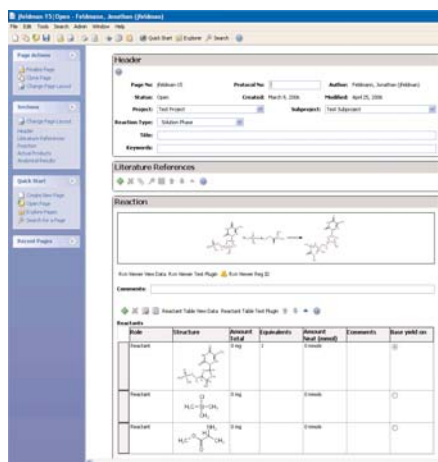
Benchware Notebook is an enterprise electronic laboratory notebook already in use by hundreds of discovery scientist. The system provides an open, service-oriented architecture that is designed to integrate with existing enterprise systems

■ **Benchware Dock**

Benchware Dock is a Windows®-based system that allows laboratory chemists to test compound ideas in real time against a 3D model of the target protein structure created and validated by expert molecular modelers.

**Performance of EXACTMATCH Searches**

AUSPYX implements high performance EXACTMATCH search algorithms with the ability to match salts or tautomers of a target compound. This operator averages approximately 50 searches per second or .02 seconds per search on dual 650 Mhz CPU PA-RISC HP.



Benchware Notebook, Tripos' enterprise-class electronic laboratory notebook, is an example of an informatics system built on the AUSPYX framework.

**Tautomer Searching Accuracy**

Tripos chemists and software developers have verified that many tautomer matches correctly identified by AUSPYX cannot be reproduced by other Oracle data cartridges.

**Standardization Capabilities**

AUSPYX provides the ability for customers to implement their own specific rules to convert compounds drawn with certain chemical idioms to a standard form to enforce data consistency. Many standardization rules exist in the default AUSPYX configuration that enforces this data integrity.

**Robust Validation Capabilities**

AUSPYX provides the ability to identify compounds that might require human intervention prior to registration. Validation is based on identification of valid valence states, isotopes, and specific patterns.

**Importing Data Into Auspyx**

AUSPYX includes PL/SQL APIs that allows users to load their structure data into AUSPYX datasets. These APIs may also be used for loading data and providing search functionality on existing schemas that do not have predefined AUSPYX datasets.

**Exporting Data From Auspyx**

In addition to Oracle export utilities, AUSPYX provides a command line client application that allows users to export data from an AUSPYX data set into SLN hitlists similar to those generated by SYBYL® and UNITY.

**Utilities**

AUSPYX provides a full set of utility functions for handling and manipulating SLNs. SLN OPS utilities provide a rich set of molecule descriptors including functions to return molecular weight, molecular formulae, hydrogen bond donor, and acceptor counts. The SLN OPS functions are suitable for creating function-based indices on SLN database columns. In addition, the SLN OPS package provides functions to translate to and from SMILES strings as well as to parse CLOBs containing MDL molecules into SLNs.

**Storing and Searching 3rd Party Fingerprints**

The FPRINT\_INDEX\_TYPE allows AUSPYX to store and search 3rd party fingerprints. This indextype provides a new set of operators for Tanimoto, Cosine, Tversky, Hamming, Asymmetric, Modified Tanimoto, and Dice methods of similarity. This fingerprint index

**New!! Learned Index Technology**

AUSPYX's Learned Index technology introduces a boost to AUSPYX performance. By extracting patterns of user behavior from submitted queries, AUSPYX is able to accelerate future queries based upon prior requests.

In this way, communities of AUSPYX users will benefit from one another's work, even if they are not necessarily working on the same project.

type supports both traditional bitstring fingerprints as well as SciTegic fingerprints.

**OptiSim™ Integration**

AUSPYX provides the ability to select diverse yet representative subsets of the database using Tripos OptiSim™ technology. This is the same technology available in Tripos' Benchware HTS DataMiner product. AUSPYX integrates OptiSim with its FPRINT OPS functions to support the use of OptiSim with third party fingerprint data. In addition to diversity selection, AUSPYX supports compound clustering based on the OptiSim algorithm.

**System Requirements**

- Oracle 10g R2
- Available for multiple operating systems, including:
  - Red Hat Enterprise Linux AS/ES 4.0
  - HP-UX 11i: 64-bit
  - Windows Server 2003
  - Solaris 10: 64-bit



WWW.TRIPOS.COM				CONTACT_US@TRIPOS.COM			
<b>AUSTRALIA</b> +61 (7) 5439 9775	<b>CANADA</b> +1 450 4334500	<b>FRANCE</b> +33 1 69 59 29 49	<b>GERMANY</b> +49 89 45 10 300	<b>JAPAN</b> +81 3 5166 1721	<b>UNITED KINGDOM</b> +44 1 908 650000	<b>UNITED STATES</b> 800 323 2960 +1 314 647 1099	