Three-dimensional and Two-dimensional Searching of Databases

Conventional 3D database searching finds only molecules whose stored conformations match the constraints of the query. UNITY's conformationally flexible 3D searching finds molecules that can achieve a matching conformation regardless of the stored conformation. The Directed Tweak algorithm in UNITY quickly finds molecules that could match the constraints of the query. Important molecules in compound classes never before considered are often identified.

UNITY allows you to build structural queries based on molecules, molecular fragments, pharmacophore models, or receptor sites. In addition to atoms and bonds, 3D queries can include features such as lines, planes, centroids, extension points, hydrogen bond sites, and hydrophobic sites. Distance, angle, excluded volume, surface volume, and spatial constraints define the geometric relationships between features. In receptor-based queries, UNITY additionally has the ability to specify multiple excluded volumes that represent the binding site structure, or to define the containing volume of the receptor cavity.

Partial match directives (for example, match at least four but no more than seven of 10 possible H-bond donor sites) allow a search for compounds containing only some of the features specified in the query.

Two-dimensional database searching looks for connectivity patterns in molecules as specified by the query. UNITY's 2D search options include exact and substructure searching. A unique option of UNITY provides substructure searching of combinatorial libraries. Queries may also specify variable elements such as Markushes. Precomputed fingerprints that encode molecular features optimize search times by quickly eliminating compounds that cannot match a query. Similarity searching in UNITY is based on a comparison of these fingerprints.

Features
- Directed Tweak provides rapid, conformationally-flexible 3D searching
- An extensive set of features and constraints are available to construct 3D queries
- Partial matching of query features is allowed for any 3D search
- Full Markush capabilities enable variable query specification
- Fingerprints are used as filters to speed 2D and 3D searching
- Queries contain realistic hydrogen bond donor/acceptor site representations
- Fingerprints can be customized and edited
Storage of Compound Structures and Associated Data

UNITY creates Tripos databases that store compound structural data such as name, RegID, fingerprints, and 3D coordinates. Standard or custom fingerprints are calculated during database creation. UNITY uses SYBYL Line Notation (SLN), a flexible and concise language for storing molecular structures and for creating search queries. An extension of SLN, combinatorial SLN (cSLN), can specify a library of compounds as a single searchable entry.

UNITY utilities convert a file or database in one format to another. UNITY can transfer structures between Tripos databases and MOL2, MOL, and SD files, as well as Daylight SMILES strings. Commercial databases are easily converted to Tripos databases, and associated property data can be stored in Oracle. Large 2D databases can be converted to 3D Tripos databases with Concord (licensed separately), the industry standard for rapid conversion of 2D input to accurate, geometry-optimized 3D structures.

Enterprise-wide Access to Structures and Data

UNITY can access databases anywhere on a network. Security locks may be implemented at the level of workgroups or individuals, and can be applied to specific sets of compounds.

Features
- Structures can be stored with 3D coordinates, fingerprints, and associated data
- Concise combinatorial library storage is made possible by combinatorial SLN
- UNITY 3D databases can be created from 2D databases using Concord and StereoPlex
- UNITY is compatible with other vendors’ databases

Applications
- Conformationally flexible searching of compound databases for ligands that fit a receptor site
- Exploration of databases for compounds consistent with a pharmacophore hypothesis
- Lead explosion by retrieving similar compounds
- Virtual screening of compound databases to discover lead compounds
- Determining reagents in commercial databases that support combinatorial chemistry synthesis

Features
- Multiple databases residing anywhere on a network can be searched simultaneously
- Network-wide accessibility eliminates the need to keep and synchronize multiple copies of databases

The UNITY Hitlist Manager, showing the results of a sub-structure search of the NCI database.
Hardware and Software Requirements
UNITY is accessible through the SYBYL expert molecular modeling environment. SYBYL requires a separate license. Licensing options for UNITY include: UNITY/Base for 2D searching, UNITY 3D for 3D searching, UNITY extra search engine for running multiple searches concurrently, and UNITY RDBMS for searching relational data in Oracle. UNITY also available standalone. UNITY and SYBYL run on workstations operating under Mac OS X or Linux® (x86).

Complementary Software
- Advanced Computation for exploring the conformational properties of compounds.
- AMPAC™ for calculating transition states and spectral properties using semiempirical quantum mechanical methods.

Features
- Results may be viewed in a Molecular Spreadsheet or the UNITY Hitlist Manager.
- Boolean operations can be used to combine different hitlists.
- Hitlists or Molecular Spreadsheets can also be searched.
- Hits returned by UNITY are aligned to the query, facilitating analysis by QSAR with CoMFA.
- Hits returned from receptor-site queries are docked to the receptor.
- Post-processing of UNITY hits includes torsional minimization constrained by the query.

References