

Tuplets™

PHARMACOPHORE-BASED VIRTUAL SCREENING WITHOUT A 3D MODEL



Tripos™
Go. Discover

Tuplets¹ enables researchers to retrieve compounds from molecular structure databases that are pharmacophorically, and therefore biologically, similar to known actives without the need to derive a classical 3D pharmacophore model. Structure databases are searched with pharmacophore fingerprint hypotheses (Tuplet hypotheses). Tuplet hypotheses can be derived from known active compounds, combined sets of active and inactive compounds, or UNITY[®] queries. Compounds retrieved in this way can be of a different chemical class than those employed to generate the hypothesis, demonstrating Tuplets' lead hopping capabilities. Also, Tuplets can be used to prioritize virtual combinatorial libraries for further research.

Tuplet Fingerprints

A Tuplet fingerprint is a bitmap encoding the interfeature distances of a set of interesting pharmacophore features. Several varieties of Tuplet fingerprints can be calculated:

Pairs: Two pharmacophore features with a single distance

Triplets: Three pharmacophore features with three edge distances

Quartets: Four pharmacophore features with six edge distances and chirality

Augmented: A pair or triplet with the additional stipulation that a user-defined pharmacophoric feature must be found within a certain distance of one of the pair/triplet features

Privileged: A pair or triplet with the additional stipulation that a user defined substructure must be found in all pairs/triplets

Molecule-Based Tuplet Hypotheses

Molecules from the input training set can be clustered by their Tuplet fingerprints and subdivided into groups representing alternate binding modes prior to hypothesis generation. Tuplet hypotheses are then derived from the Tuplet fingerprints representing molecules in the training set. Tuplet hypotheses can be further refined using knowledge of newly identified active or inactive compounds. Tuplet hypotheses can be analyzed by comparing it to Tuplet fingerprints calculated for a set of reference molecules using a variety of different similarity measures. This provides insight into the selectivity of the hypothesis, and enables the user to set proper cut-off values for each similarity measure. (Figure 1)

Query-Based Tuplet Hypotheses

A Tuplet hypothesis can be constructed from UNITY² queries based on pharmacophore models or receptor-site derived queries. This enables the incorporation of important feature information derived from X-ray structure analysis, docking studies, or 3D pharmacophore models. A query-based Tuplet hypothesis can be used as a prescreen to rapidly filter unsuitable compounds and reduce the time necessary to perform a fully flexible 3D search or a docking calculation.

Searching a Database of Molecules

Molecular structure databases can be mined for compounds that are sufficiently similar to a selected Tuplet hypothesis. Compounds can be selected using either a cut-off value (specified for one or more similarity coefficients) or the top *n* or top *n* % of compounds determined by an ordered list of similarity measurements.

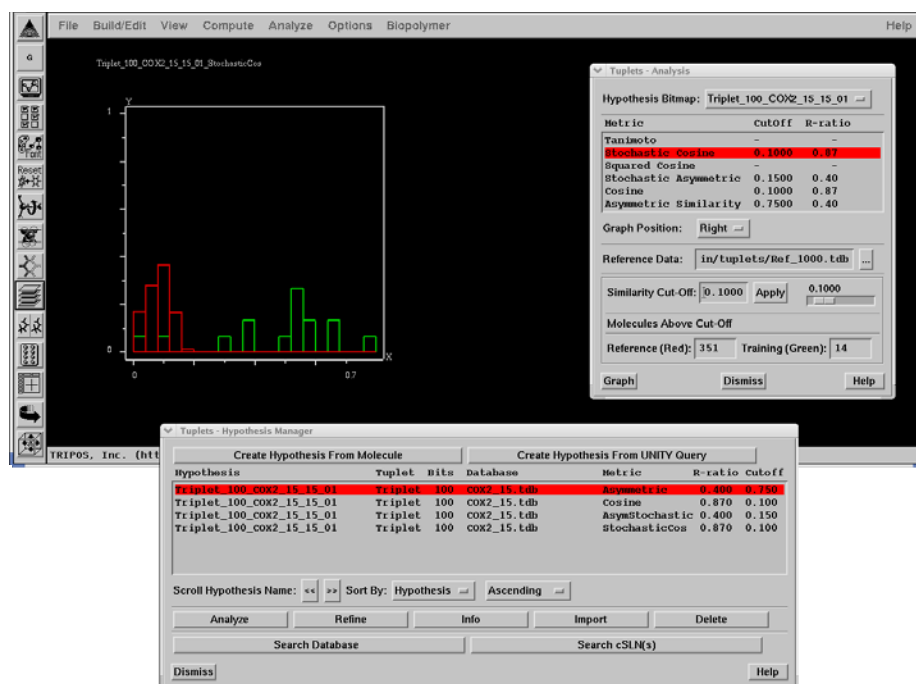


Figure 1) A histogram (top left) is used to show the results of comparing a hypothesis (selected in the Tuplets - Hypothesis Manager - bottom center) to a reference set of 1,000 randomly selected drug-like molecules (shown in red) and the training set used to create the hypothesis (shown in green). The Tuplets - Analysis interface (top right) displays the number of compounds from both the training and reference sets that have a similarity greater than or equal to the specified cut-off value, allowing the user to interactively explore and apply multiple cut-off values for one or more similarity measurements.

Advantages

- Searching a molecular structure database using Tuplets is orders of magnitude faster than traditional 3D searching
- Tuplet hypotheses can be generated from HTS hits, well characterized active and inactive compounds, and/or UNITY queries
- Tuplets allows categorization of compounds by pharmacophoric properties
- Compound sets retrieved from molecular structure databases are enriched in actives
- Tuplets allows characterization of queries by comparison with reference data sets
- Tuplets can be used without the need to derive a classical 3D pharmacophore model

Prioritizing a Set of Combinatorial Libraries

Combinatorial libraries, stored in cSLN³ format, can be prioritized using Tuplets by sampling compounds randomly from the libraries, determining their similarity to a selected Tuplet hypothesis (based on one or more similarity measurements), and generating simple statistics regarding the similarity of those compounds to the hypothesis. When comparing the results of multiple cSLNs, virtual combinatorial libraries can be prioritized for design and synthesis work.

Applications

- Database searching for molecules with desired biological activity
- Prioritization of virtual combinatorial libraries

Features

- Use pre-generated conformers or in-line computation of conformers
- Calculate Tuplet fingerprints for new compounds added to an existing database
- Select compounds used to generate a hypothesis by clustering molecules using Tuplet fingerprints
- Refine an existing Tuplet hypothesis using 3D pharmacophoric information from newly identified active or inactive compounds
- Display a human-readable information file detailing the history of a Tuplet hypothesis
- Analyze a Tuplet hypothesis and set cut-off values using interactive histograms calculated for multiple similarity measurements

- Calculate Tuplet fingerprints or perform searches on remote machines using multiple processors
- Project-based interface allows straightforward creation, manipulation, analysis, and use of Tuplet fingerprints and hypotheses
- Available through the SYBYL[®] interface or as command line executables for wider deployment

Complementary Software

- **UNITY** for rapid, flexible 3D searching of databases to identify lead compounds based on a pharmacophore hypothesis.
- **Concord**[®] for rapidly converting 2D chemical structures into high quality 3D structures.
- **Confort**[™] for generating sets of molecular conformers.
- **Legion**[™]/**CombiLibMaker**[™] for building virtual combinatorial libraries in cSLN format.
- **OptDesign**[®] for designing and editing combinatorial libraries.
- **Selector**[™] for evaluating the diversity of compound databases and selecting diverse subsets.
- **HQSAR**[™] for calculating novel 2D fingerprints (holograms) as descriptors.
- **QSAR with CoMFA**[®] for constructing predictive structure-activity models from sets of aligned molecules.
- **DISCOtech**[™] for identifying pharmacophore models from sets of pre-computed conformations.
- **GASP**[™] for automatic pharmacophore elucidation with full conformational flexibility.
- **FlexS**[™] for performing shape-based screening and automatic structural alignment of drug-like molecules.
- **VolSurf**[™] for predicting ADME properties of compounds based on pre-calculated models.

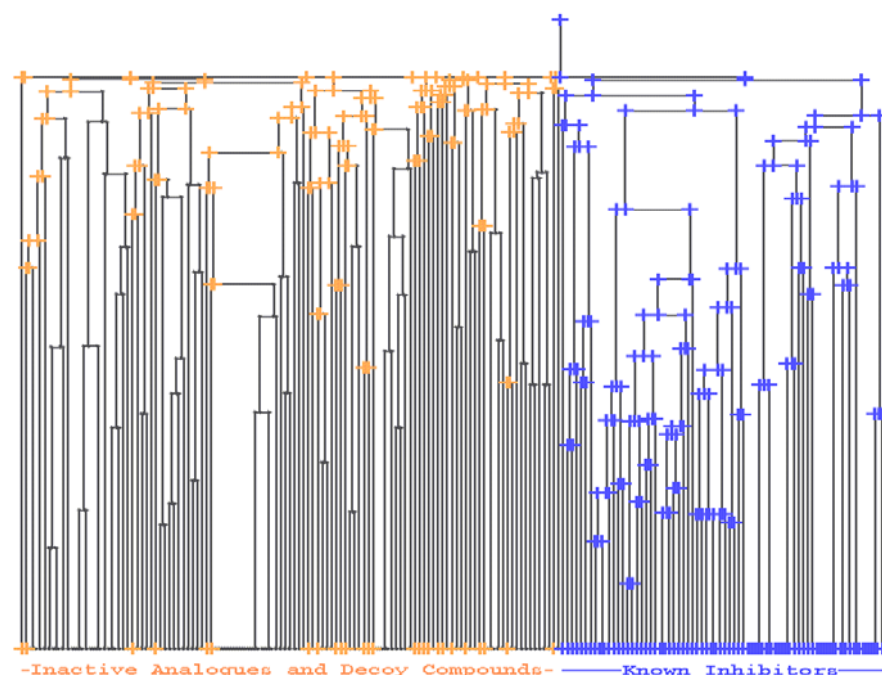


Figure 2) Hierarchical clustering of Tuplet fingerprints for 96 compounds from a GSK3 QSAR data set⁴ together with 91 structurally similar compounds drawn from the NCI anticancer screening database.⁵ Basal nodes corresponding to the GSK training set are highlighted: blue crosses indicate inhibitors with characterized IC50s, whereas orange crosses indicate inactive analogues. Basal nodes not highlighted correspond to NCI decoy compounds.

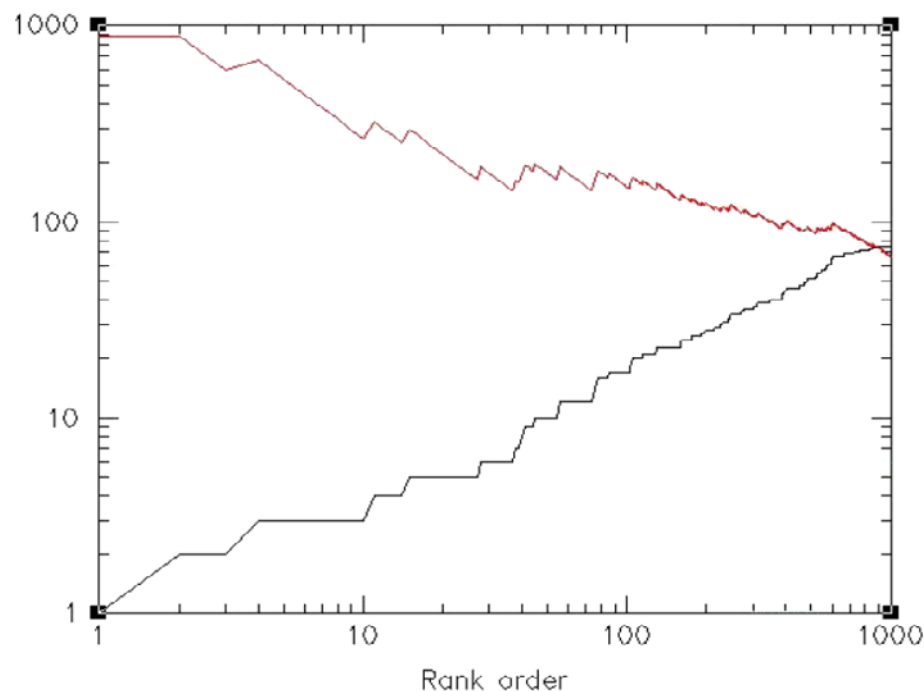


Figure 3) Log scale recovery (red) and enrichment (black) curves for 434 known estrogen antagonists seeded into a 385,000-compound data set for a search using a Tuplet hypothesis obtained from the training set of eight compounds.⁴ These curves demonstrate that 25% of the active compounds can be found in the first 0.26% (1,000 compounds) of the ranked database. This represents an enrichment rate of ~100 fold over random sampling.

Validation

An analysis was carried out on a set of 96 glycogen synthase kinase-3 (GSK3) inhibitors and analogues originally compiled for a quantitative structure activity relationship (QSAR) analysis.⁴ This set was comprised of 70 active compounds and 26 less active analogues. An additional 91 were drawn from the NCI anticancer screening database⁵ that exhibited a Tanimoto similarity of 0.65 or greater to the compounds in the training set with respect to standard UNITY substructural fingerprints. Twenty-five conformers were produced for each of the 187 compounds, and the generated Tuplet fingerprints were used for

hierarchical clustering, with the results shown in (Figure 2). The actives (blue crosses) all cluster together to the right in the dendrogram, whereas the inactive analogues (orange crosses) are distributed among the nodes that are not highlighted, corresponding to decoy compounds from the NCI database. Additionally, Tuplets has demonstrated the ability to retrieve compounds from a molecular structure database enriched in actives (Figure 3).

Acknowledgements

Tuplets is the result of a research collaboration performed between Novo Nordisk and Tripos.

Hardware and Software Requirements

Tuplets requires Tuplets, UNITY, and ChemEnlightenUtility licenses. Virtual combinatorial library prioritization requires Legion and Concord licenses. Tuplets is accessible from SYBYL and runs on workstations operating under IRIX® (SGI®) or Linux® (x86).

References

1. Abrahamian, E.; Fox, P.C.; Nærum, L.; Christensen, I.T.; Thøgersen, H.; Clark, R.D. "Efficient Generation and Manipulation of Fully Flexible Pharmacophore Multiplets and Their Use in 3-D Similarity Searching." *J. Chem. Inf. Comput. Sci.* **2003**, *43*: 458-468.
2. Martin, Y. "3D Database Searching in Drug Design." *J. Med. Chem.* **1992**, *35*: 2145-2154.
3. Ash, S.; Cline M.; Homer, R.; Hurst, T.; Smith, G. "SYBYL Line Notation (SLN): A Versatile Language for Chemical Structure Representation." *J. Chem. Info. Comp. Sci.* **1997**, *37*: 71-79.
4. Nærum, L.; Nørskov-Lauritsen, L.; Olesen, P. H. "Scaffold Hopping and Optimization Towards Libraries of Glycogen Kinase-3 Inhibitors." *Bioorg. Med. Chem. Lett.* **2002**, *12*: 1525-1528.
5. National Cancer Institute. Bethesda, MD. <http://dtp.nci.nih.gov/>.



WWW.TRIPOS.COM				CONTACT_US@TRIPOS.COM		
AUSTRALIA +61 (7) 5439 9775	CANADA +1 450 4334500	FRANCE +33 1 69 59 29 49	GERMANY +49 89 45 10 300	JAPAN +81 3 5166 1721	UNITED KINGDOM +44 1908 650000	UNITED STATES 800 323-2960 +1 314 647-1099
©Tripos, Inc. 2005 0075 Printed in USA All trademarks are the property of their respective owners.						