

Linking Context Similar Information

Finding “more like this”



CINF General Session

ACS Chicago

August 30, 2001

Tim Hocter

MDL Information Systems, Inc

- ▶ Background
- ▶ Need
 - ◆ single query that then automatically generates multiple queries to retrieve a complete and relevant set of options and results, regardless of the data sources or data types (i.e. “more like this”).
- ▶ Meta layer solutions
- ▶ Four examples of context similar searching
 - ◆ text; structure; dynamic structure; text to structure
- ▶ Conclusion

Background: Evolution of Searching Expectations

- ▶ Print: manual searching of multiple print sources
- ▶ Databases: manual searching of databases
- ▶ Internet: has changed expectations; brought huge amount of diverse data into a simple single query-able environment
- ▶ Context similarity searching: metadata pre-processing and dynamic processing allows queries to be 'translated' to find relevant data.

Scientists want a single query that automatically generates multiple queries to retrieve a complete and relevant set of results, regardless of the data sources or data types (i.e. “more like this” or “read my mind”).

- ▶ Data models do not support single common interface query
- ▶ Different terminology
- ▶ Different query formulations
- ▶ Too many applications
- ▶ Multiple types of applications
- ▶ Proprietary and non proprietary
- ▶ Data fields not normalized or de-duplicated

- ▶ Create a meta data layer
- ▶ Transparently pass queries to the meta layer
- ▶ Transparently translate queries through metadata to different formats and different query types
- ▶ Transparently search multi-data sources with different query formats
- ▶ Transparently search multi-data sources with different query types
- ▶ Present results to the user

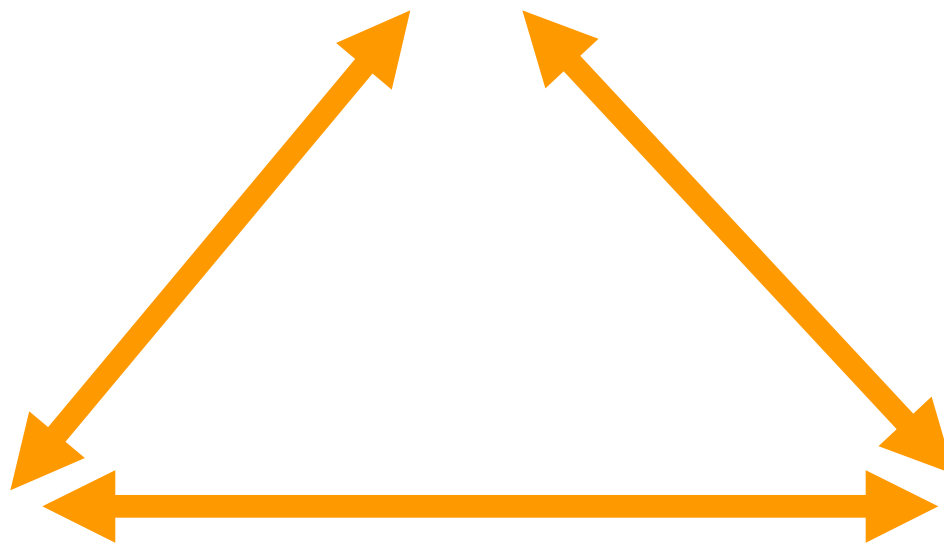
How can context similarity searching be implemented? 'Synaptic Linking'

- ▶ Embed in hyperlinks query formulations that emulate the thought process (read my mind)

What Type of Data Sources Need to be Included for Context Similarity Searching?

Primary Information

Journals Proprietary Patents



Tertiary Information

Major Reference Works

Secondary Information

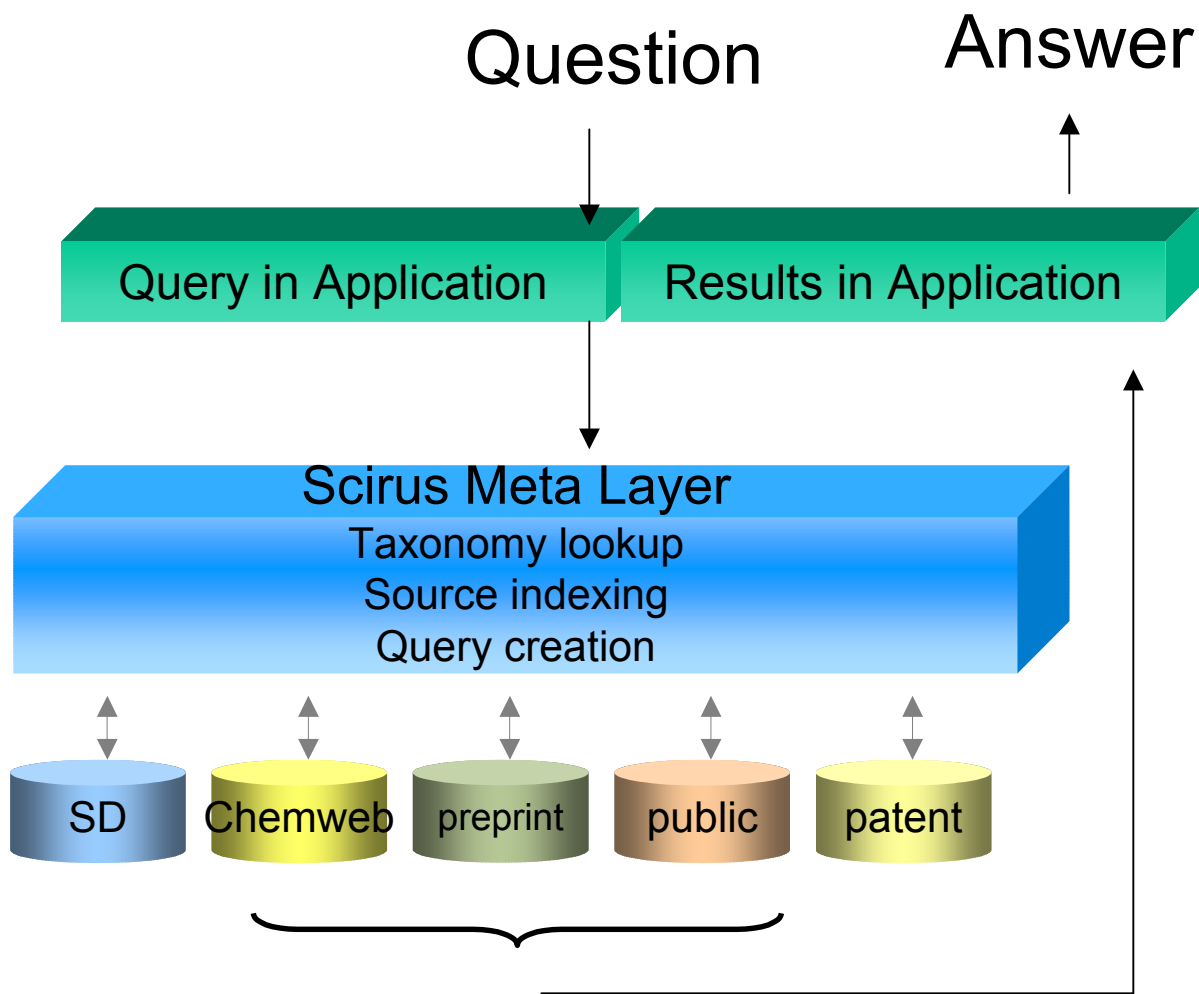
ADME/Tox Spectra
Reaction Details Synthesis
MSDS

Examples of Context Similarity Searching

- ▶ Text queries embodied as a hyperlinks created by the meta layer to the documents (not to an A&I layer)
 - ◆ Scirus
- ▶ Structure queries transparently translated by the meta layer to multiple applications
 - ◆ Compound Locator
- ▶ Structure queries embodied as a hyperlinks created by the meta layer
 - ◆ Dymond
- ▶ Transparently searching multi-data sources with different query types

- ▶ Build a lexicon (dictionary plus thesaurus)
 - ◆ Enter terms
 - ◆ Enter terminology
- ▶ Create taxonomy
 - ◆ Create subject index
 - ◆ Map terms to subjects
- ▶ Index available sources
 - ◆ Manually or programmatically

Text Meta Layer - Context Similarity Searching



Scirus- Text Meta Layer Context Similarity Searching

▶ Meta Layer-Indexing

- ◆ Non proprietary
- ◆ Programmatically generated index from full text (not abstract)
- ◆ What is missing? proprietary with non proprietary indexing

SCIRUS
for scientific information only

ChemWeb.com

Daily News Journals
Jobs Conferences
Databases Discussions

FIND

Customize Scirus About Scirus Submit URL Contact Scirus Help

Search Advanced Search

All of the words All Sources

Subject Areas

<input type="checkbox"/> All	<input type="checkbox"/> Earth and Planetary Sciences	<input type="checkbox"/> Mathematics
<input type="checkbox"/> Agricultural and Biological Sciences	<input type="checkbox"/> Economics, Business and Management	<input type="checkbox"/> Medicine
<input type="checkbox"/> Astronomy	<input type="checkbox"/> Engineering, Energy and Technology	<input type="checkbox"/> Neuroscience
<input type="checkbox"/> Biosciences	<input type="checkbox"/> Environmental Sciences	<input type="checkbox"/> Pharmacology
<input checked="" type="checkbox"/> Chemistry and Chemical Engineering	<input type="checkbox"/> Life Sciences	<input type="checkbox"/> Physics
<input type="checkbox"/> Computer Science	<input type="checkbox"/> Materials Science	<input type="checkbox"/> Social and Behavioral Sciences

Searched for all of the words **mutagenic, cancer**. Results 1 - 10 of **628** documents.

All Results Membership Results **i** Web Results **i**

1 [HERP \(Human Exposure/Rodent Potency\)](#)
HERP Index (Human Exposure/Rodent Potency) We have ranked possible carcinogenic hazards from a variety of common human exposures to rodent carcinogens, including naturally occurring chemicals in the diet. The ranking uses and index (HERP) that indicates for each exposure, the percentage of the rodent carcinogenic dose that a person receives. The higher the HERP value, the higher Last modified 07/31/01
<http://potency.berkeley.edu/herp.html>

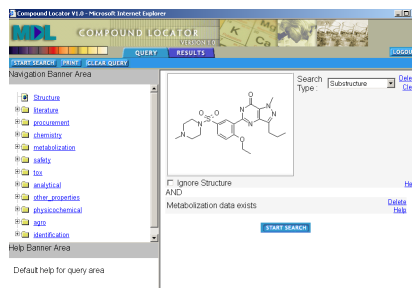
[More Like This Result](#) [Email This Result](#)

▶ Meta Layer-Context Similarity Searching

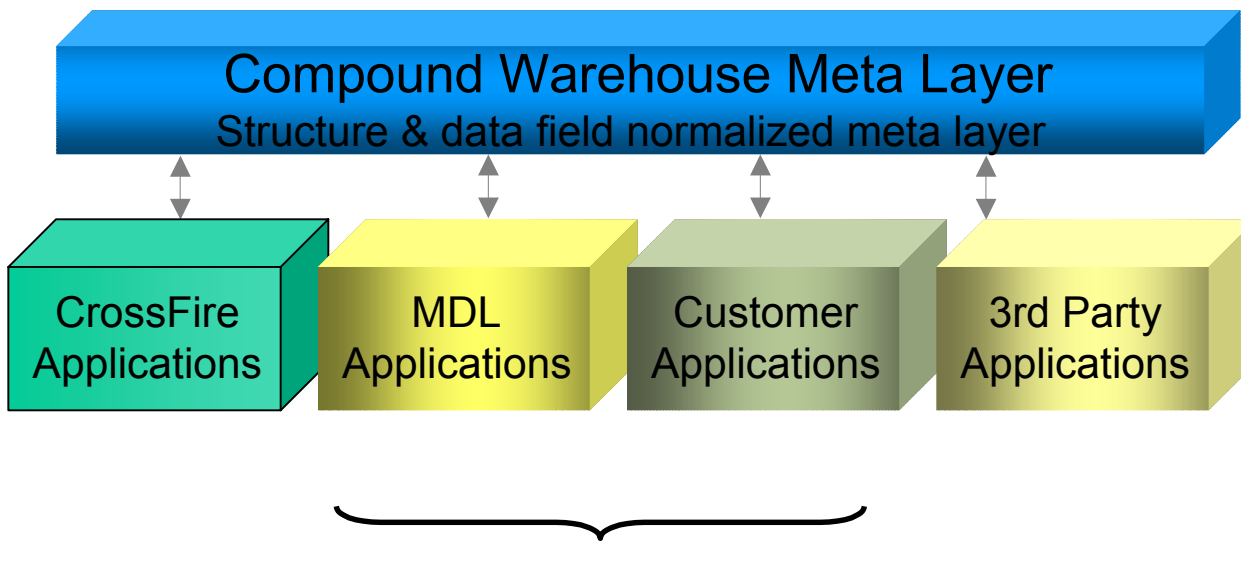
- ◆ Ability to dynamically create query from results (more like this searching)
- ◆ What is missing? “more like this” from proprietary sources

Structure Meta Layer- Context Similarity Searching

Compound Locator: One query access to multiple sources



Results in Application



Compound Warehouse- Structure Meta Layer Context Similarity Searching

▶ Meta Layer-Context Similarity Searching

- ◆ Ability to dynamically create query from results (more like this searching)

CrossFire Web - Microsoft Internet Explorer

MDL Information Systems CrossFire Web VERSION 1.0

QUERIES RESULTS REPORTS RXN SCHEMES

COPY TO REPORT PAGE SETUP PRINT SAVE REFINER QUERY NEW QUERY HELP LOGOUT

Search results 1 to 1 of 1

Pages: 1

[View your results in Compound Locator](#)

BRN: 4393773

[Details](#) [Synthesize](#)

Compound Locator V1.8 - Microsoft Internet Explorer

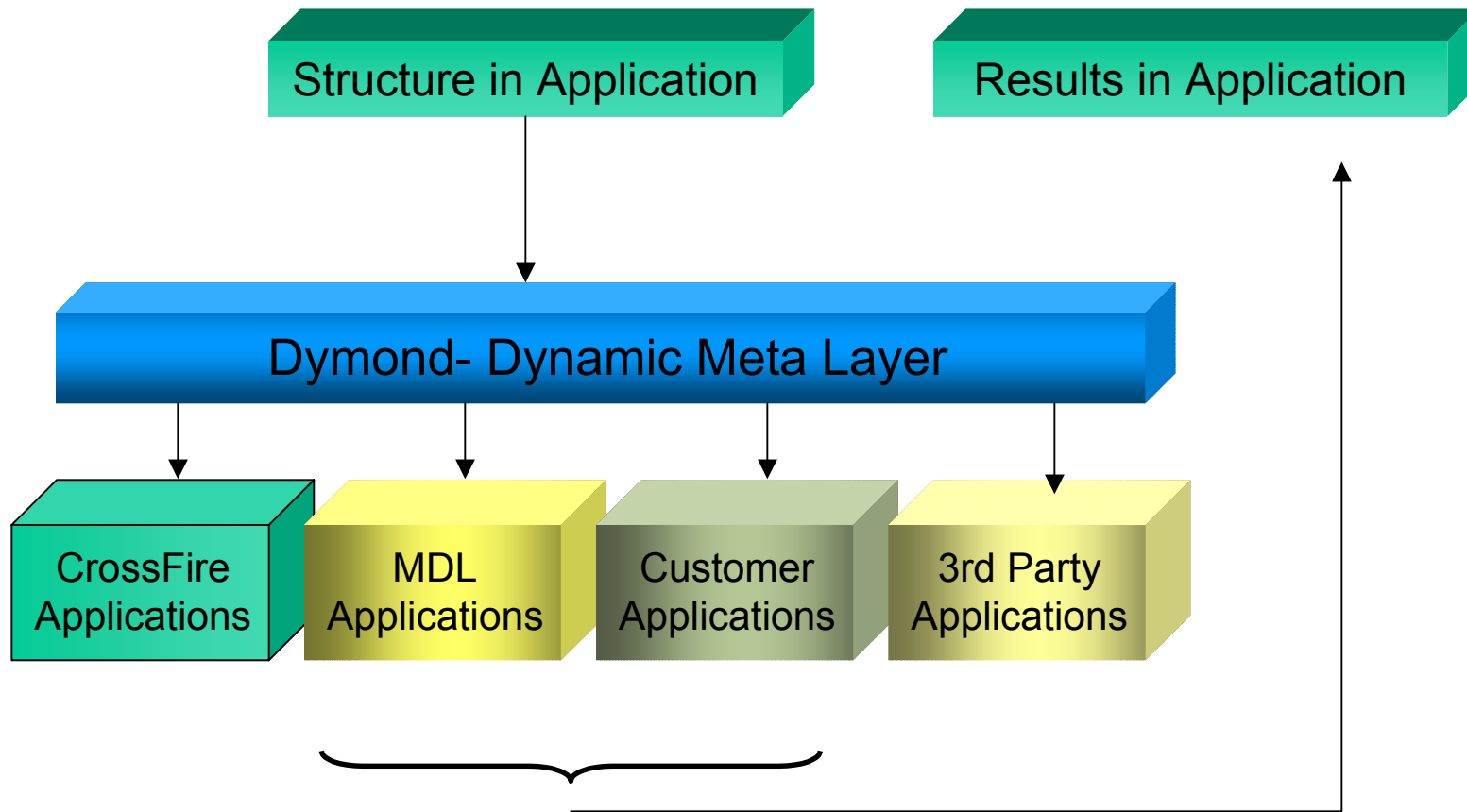
MDL COMPOUND LOCATOR VERSION 1.0

QUERY RESULTS

Database Filter 16 of 16 Show records: 10 1 to 10 of 17

#	Structure	availability	proc_preparation	toxicology	patent	safety	metabolization
1		ACD	Boitstein	Tox			Bio
2		ACD	Boitstein			OHS	
3		ACD	Boitstein	Tox		OHS	Bio

Structure Meta Layer- Dynamic Context Similarity Searching



Structure Meta Layer- Dynamic Context Similarity Searching

Present options from context similar searches

- ◆ Spectra details
- ◆ Reaction details
- ◆ Similar compound information
- ◆ Related bibliographic information

Keywords

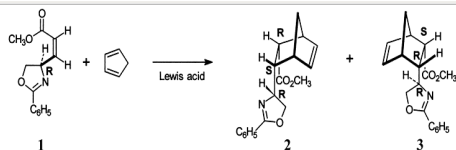
catalysis
complexation
diastereoselection
Diels-Alder reactions

Abstract

The Diels-Alder reaction between (R) -methyl (2*S*-(4*S*-hydroxy-2-phenyl-4-oxobutyl)-2-propenoate (1) and cyclopentadiene in the presence of one equivalent of Et_3AlCl gave stereoisomeric results opposite to those obtained with one equivalent of EtAlCl_2 . Energy minimizations of proposed complexes of these Lewis acids with the chiral dienophile at the B3E9-2/10 level suggest that the aluminum is tetrahedrally coordinated with Et_3AlCl , but bonded in a trigonal bipyramidal with EtAlCl_2 . These complexes expose the diastereotopic faces of the dienophile to reaction with diene.

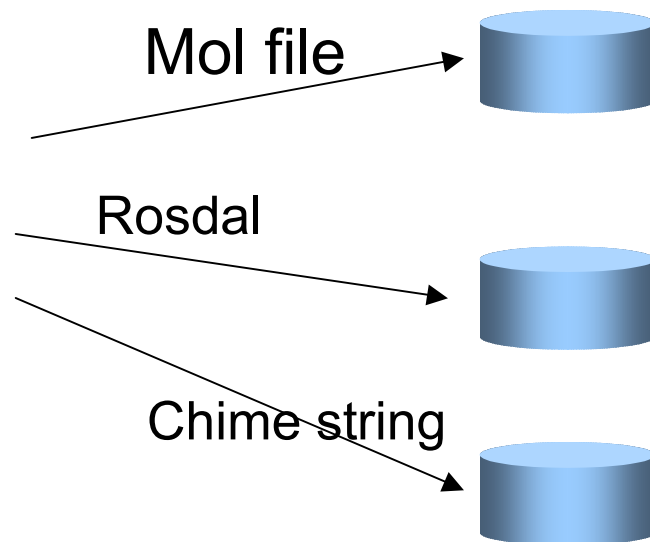
The utility of Lewis acid catalyzed Diels-Alder reactions for the synthesis of natural products, many of which are chiral, has been repeatedly demonstrated.¹ The clear value of this synthetic tool has prompted studies of asymmetric induction using chiral dienes²⁻⁴ as well as Lewis acids complexed with chiral auxiliaries⁵ or with chiral dienophiles.⁶ However it does not appear to have been explicitly recognized that the role of the auxiliary Lewis acid may specifically dictate the stereochemical outcome of the reaction in which it is employed.

The chiral oxazoline (R) -methyl (2*S*-(4*S*-hydroxy-2-phenyl-4-oxobutyl)-2-propenoate (1) is readily prepared from (2*S*)-serine⁷ and has been employed in reaction sequences leading to polyhydroxylated pyrrolidines (asaragins). We examined the reaction of 1 as a chiral dienophile with cyclopentadiene using both diethylaluminum chloride (Et_2AlCl) and ethylaluminum dichloride (EtAlCl_2); these Lewis acids produced dramatically different and opposite diastereomeric ratios of the endo-adducts 2 and 3 (Scheme 1).

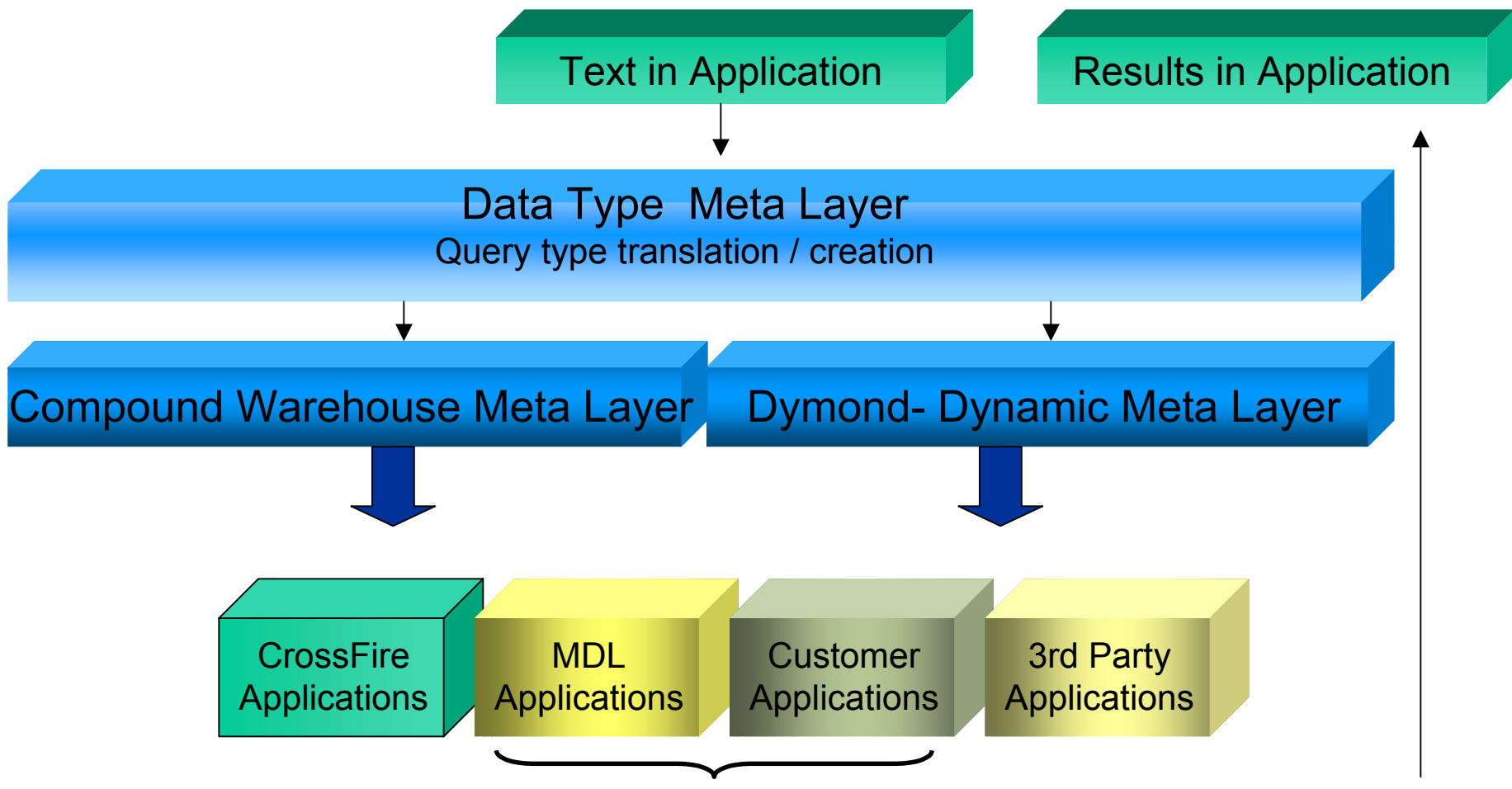


Reaction of 1 with one equivalent of diethylaluminum chloride in dichloromethane and an excess of cyclopentadiene over a 24 h period (78°C \rightarrow room temperature) produced a 88:12 mixture of 2:3 in 38% yield (77% conversion) while, under the same conditions, the use of ethylaluminum dichloride yielded a 2:98 mixture in 30% yield (63% conversion).¹ Compounds 2 and 3 were the only detectable Diels-Alder products of these reactions; exo-products were not observed. Compound 2 was an oil, but 3 was crystalline and its structure was determined as shown below (Fig. 1).

<p>Chiral</p> <p>1</p> <p>Close Window</p>	<p>Find this compound in Beilstein CrossFire Go Help</p> <p>Note: users will need to have Commander 2000 installed and have access to CrossFire for this search to work</p>
	<p>Find related compounds in Combined Chemical Dictionary Go Help</p> <p><input type="radio"/> Exact <input checked="" type="radio"/> 60% similar <input type="radio"/> Substructure</p>
	<p>Generate IUPAC Name for this compound using AutoNom Standard Go Help</p>
	<p>Search for NMR spectral details in the ACD/Labs databases Go Help</p> <p><input type="radio"/> Exact <input checked="" type="radio"/> 60% similar <input type="radio"/> Substructure</p> <p><input checked="" type="radio"/> H NMR <input type="radio"/> C NMR <input type="radio"/> F NMR <input type="radio"/> P NMR</p>
	<p>Search for properties in the ACD/Labs databases Go Help</p> <p><input type="radio"/> Exact <input checked="" type="radio"/> 60% similar <input type="radio"/> Substructure</p> <p><input checked="" type="radio"/> logP <input type="radio"/> pKa</p>



Structure Meta Layer- Example of a Dynamic Meta Layer



Transparently searching multi-data sources with different query types

- ▶ Passing the following text to meta layer
(R)-(-)-methyl (Z)-3-(4,5-dihydro-2-phenyl-4-oxazolyl)-2-propenoate
- ▶ Which translates to a structure query type for multiple applications

Tetrahedron Letters
Volume 41, Issue 15
8 April 2000
Pages 2519-2522

[SummaryPlus](#)
[Article](#)
[Journal Format-PDF \(122 K\)](#)
[Cited By](#) | [Save as Citation Alert](#)
[Export Citation](#)

PH: S0040-4039(00)00241-0
Copyright © 2000 Elsevier Science Ltd. All rights reserved.

Diastereoselective Diels-Alder reactions. The role of the catalyst

Yifang Huang^a, Philip E. Sonnet^a, Patrick J. Carroll^{†,b} and David R. Dalton^{†,b,*,a}

^a The Department of Chemistry, Temple University, Philadelphia, PA 19122, USA
^b The Department of Chemistry, The University of Pennsylvania, Philadelphia, PA 19104, USA

Received 6 January 2000; revised 7 February 2000; accepted 8 February 2000. Available online 23 June 2000.

Abstract

The Diels-Alder reaction between (R)-(-)-methyl (Z)-3-(4,5-dihydro-2-phenyl-4-oxazolyl)-2-propenoate and cyclopentadiene in the presence of one equivalent of Et₂AlCl gave stereochemical results opposite to those obtained with one equivalent of EtAlCl₂. Energy minimizations of proposed complexes of these Lewis acids with the chiral dienophile at the RHF/3-21G level suggest that the aluminum is tetrahedrally complexed with Et₂AlCl, but bound in a trigonal bipyramid with EtAlCl₂. These complexes expose the diastereotopic faces of the dienophile to reaction with diene.

Author Keywords: catalysis; complexation; diastereoselection; Diels-Alder reactions

Chemical database search results showing a grid of chemical structures.

Chemical structure viewer showing the structure of 1-(2-ethoxypropenyl)pyridine.

ACD/HNMR DB Molecule information

Formula: C₁₀H₁₁NO₂
Molecule Name: 1-(2-ethoxypropenyl)pyridine
Molecule Weight: 185.26

Chemical Shifts:

Atom No.	Chem. Shift (ppm)	Ref.
4	7.53	L
8	7.18	L
6	7.14	L
9	4.23	L
10	3.18	L
11	3.18	L
12	1.87	L

More Information

Taxonomy indexing and mapping:

www.semio.com

www.autonomy.com

Metadata- index text searching:

www.scirus.com

Metadata- structure searching

www.dymondlinking.com

Metadata- usage – application to multiple data sources:

Timh@mdli.com

- ▶ Emerging metadata technology enables linking of information that is contextually similar from disparate sources without the user having to form multiple queries.
 - ◆ Text query to documents: Scirus
 - ◆ Structure to multiple applications: Compound Warehouse; Dymond