



Version 2.6 – April 2009

***The powerful solution to
integrate storage and retrieval
of chemical structures and
reactions into a relational
Oracle® environment***

1 Introduction

1.1 Overview of Cartridge Features

InfoChem's Chemistry Cartridge *ICCARTRIDGE* is a powerful software module, designed to integrate storage and retrieval of chemical structures and reactions into a relational Oracle® environment. *ICCARTRIDGE* uses one of the fastest search and indexing algorithms for structures and reactions.

Using the InfoChem cartridge you can import, register, normalize, edit, index and search chemical structures and reactions. *ICCARTRIDGE* takes advantage of Oracle®'s basic technologies. Structural and factual data can be seamlessly integrated in an unrestricted, open Oracle® environment. Easy access is granted via various database connection standards like ADO, ODBC and JDBC.

Any existing Oracle® database with structure or reaction data in LOB columns and standard format can be indexed immediately – no data conversion is required. Unicode support is provided.

ICCARTRIDGE offers the following functionalities:

- New Oracle® indextypes and chemistry search operators for

Molecules:

- Various exact structure search operators (isotopes, tautomers or salts can be included in an exact search)
- Substructure and inverse substructure search
- Similarity search
- "All-in-one" search operator (hit list ordered by relevance)
- Sophisticated sum formula search

Reactions:

- Various exact reaction search operators (search for reaction subsets or stereo- and isotope derivatives can be included)
- Reaction Substructure Search
- Various Reaction Type Searches (based on InfoChem's ClassCodes)

- Comprehensive query feature support for substructure searches
- R-group search (exact and substructure)
- Chemistry toolbox
- Support of chemical standard formats like MOL, RXN, SMILE
(for further formats a conversion library is available)

1.2 New Features in Version 2.6

The new features in ICARTRIDGE version 2.6 make the well-proven InfoChem cartridge even better. In particular the enhancements in reaction handling in version 2.4 and the new features in this version show the leading role of InfoChem in chemical information management.

- Significantly enhanced performance for reaction queries (version 2.4 feature, see diagram in chapter 3)
- Visualization functions for reaction types
- Faster generation of reaction types: all three spheres returned with one function call
- Support of MOL format in query for role search operators (XRSRCT, XRSPRD, IXRSRCT, IXRSPRD)
- New molecule exact search operators XSSEZ, RXSS
- New “all-in-one” structure search operator (SSA)
The new operator retrieves results from exact, substructure and similarity search ordered by relevance performing only one query
- Calculation of InChI and InChI key
- Calculation of elemental composition
- Extended getHashCode function (optional EZ stereo chemistry included)
- Extended import functionality: - optional normalization and auto mapping included
- Oracle date type supported

1.3 New Features under Development

New developments prove InfoChem’s expertise in meeting the requirements of modern chemists in a more and more complex scientific world.

- Extended structure representation for database structures and reactions - not fully defined structures like variable points of attachment in rings or chains, undefined R-Groups, optional R-Groups – Markush support
- “All-in-one” search operator for reactions (RSA)
- Sophisticated handling of multi step reactions - automatic multi step reaction analysis for reaction records (scheme support)
- Recognition of Named Reactions
- Configuration tools for user defined standardization and reaction mapping rules
- Enhanced structure and reaction plausibility checking

2 ICARTRIDGE Components

ICARTRIDGE consists of new indextypes with associated search operators for chemical structures and reactions and a chemistry toolbox with various methods and functions.

2.1 MOL_INDEXTYPE

- Basic indextype for structures
- Storage of chemical structures in a CLOB- or BLOB-field in any table
- Access via standard SQL (see chapter 4.1 for examples)
- The following operators are provided for the MOL_INDEXTYPE:

SSA	“All-in-one” structure search
SSS	substructure search operator
ISSS	inverse substructure search operator
XSSEZ	exact structure search operator, including E/Z stereochemistry
XSS	exact structure search operator
RXSS	relative exact structure search operator
IXSS	operator for exact search, including stereo- and isotope derivatives
TXSS	operator for exact search, including tautomers
PXSS	operator for exact search of parents of salt molecules
CompXSS	operator for exact search of molecule components
FlexXSS	operator for most flexible exact search
MSIMFP	operator for similarity search
SFS	sum formula search operator
XSFS	exact sum formula search operator

2.2 CHEMSTOR_INDEXTYPE

The ChemStorage is a predefined data model for structures:

- Storage of chemical structures in a separate table in the cartridge schema
- Structure access is completely encapsulated by views and the ICARTRIDGE Interface (*iccart.method (...)*) (see chapter 4.3 for examples)
=> automatic registration and normalization possible
- Only a subset of the MOL_INDEXTYPE operators are provided
- The support of the ChemStorage may be discontinued completely in future versions

2.3 RXN_INDEXTYPE

- Indextype for reactions
- Atom-atom-mapping and reaction center information can be generated automatically during the index creation
- Storage of chemical reactions in a CLOB- or BLOB-field
- Access via standard SQL (see chapter 4.2 for examples)
- The following operators are provided for the RXN_INDEXTYPE:

RSS	reaction substructure search operator	
XRS	reaction exact match operator	
XRSSUB	reaction subset exact match operator	
XRSRCT	reaction reactant exact match operator	
XRSPRD	reaction product exact match operator	
IXRS	reaction exact match operator	(incl. stereo- and isotope derivates)
IXRSSUB	reaction subset exact match operator	(incl. stereo- and isotope derivates)
IXRSRCT	reaction reactant exact match operator	(incl. stereo- and isotope derivates)
IXRSPRD	reaction product exact match operator	(incl. stereo- and isotope derivates)

Available only with the optional module ICCART_RXNTYPE:

RTS_[SPHERE]	(Reaction Type Search Broad/Medium/Narrow)
RXCCLUSTER_[SPHERE]	(Reaction Center Cluster Search Broad/Medium/Narrow)
RSS_RT_[SPHERE]	(Combined RSS/RTS Search Broad/Medium/Narrow)

2.4 Chemistry Toolbox

ICCARTRIDGE provides advanced chemistry tools, which are accessible via standard SQL and PL/SQL interfaces. Thus these tools can be very easily integrated into your applications.

- Sophisticated import/export routines; simultaneous import of structures/reactions and associated data from MDL SDFiles and RDFiles into any relational data model is possible
- Calculation of InChI and InChI key
- Calculation of molecule properties like various molecular masses and chemical formulas
- Calculation of stereo descriptors
- Standardization of chemical structures and reactions (standardization rules can be customized)
- Calculation of unique structure identifiers
- Highlighting of query in structure and reaction records
- Functions providing access to single reaction member
- Enhanced atom-atom-mapping function (only in the optional module ICCART_RXNMAP)
- Reaction type classification functions (only in the optional module ICCART_RXNTYPE)
- Visualization functions for reaction type (only in the optional module ICCART_RXNTYPE)

3 Performance

System for reaction and structure search performance tests:

AMD Opteron processor 146, 2.01 GHz, 4.00 GB RAM

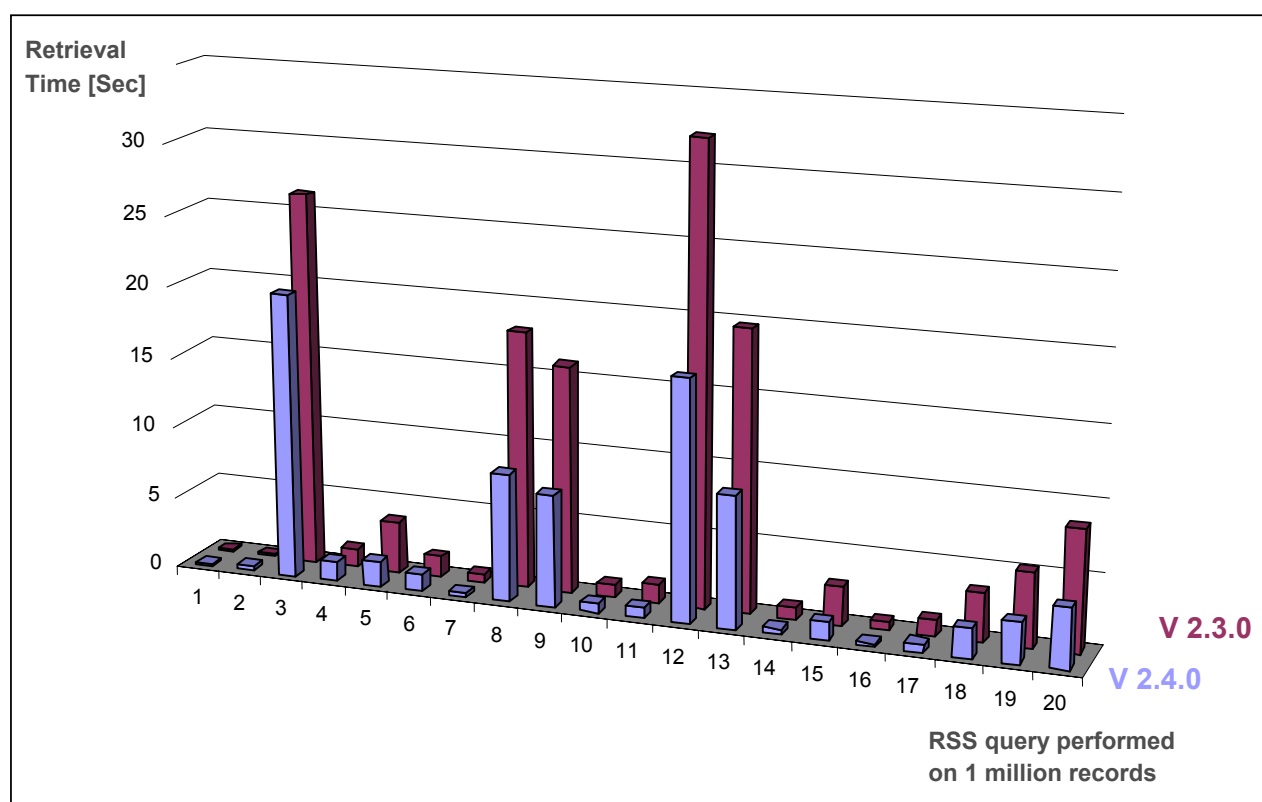
Windows 2003 Server

Oracle 10.1.0.4.0

3.1 Reactions

Data set: Table SPRESI_RXN with 1.061.207 records containing reactions

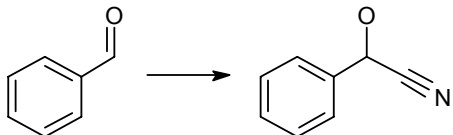
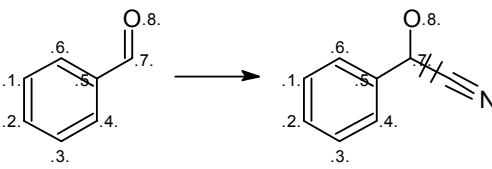
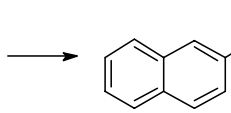
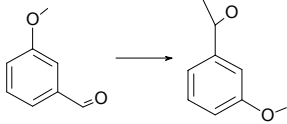
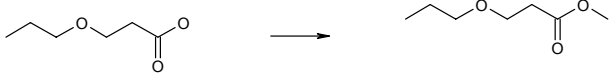
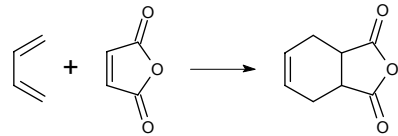
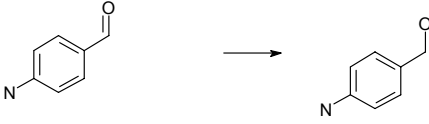
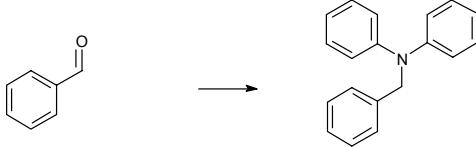
Index creation without additional parameters. The default parameter AUTOMAP=FORCE is used; for all reactions atom-atom-mapping and reaction center information is generated automatically and saved in the index, the records in the database remain unchanged.



ICCARTIDGE performance enhancements in version 2.4.0

SQL-statement for query :

SELECT COUNT (*)**FROM SPRESI_RXN a****WHERE RSS(a.REACTION, iccart.ReadRxnFile(...), 'RXN') = 1;**

No	query	hits	time [sec]
5		360	1.7
6		223	1.2
9		4334	7.8
11		889	0.8
14		656	0.4
16		374	0.2
18		4436	2.1
20		187	4.3

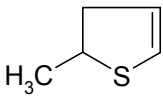
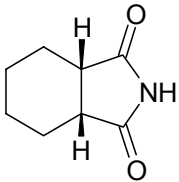
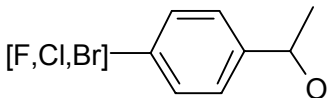
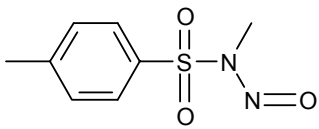
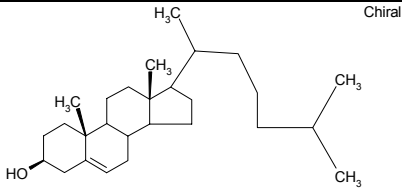
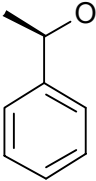
3.2 Molecules

Data Set: Table SPRESI_MOL with 1.057.341 records containing molecules

Index creation without additional parameters. The default parameter NORMALIZE=NO is used.

SQL-statement for query :

```
SELECT COUNT (*)
  FROM SPRESI_MOL a
 WHERE SSS(a.MOL, iccart.ReadMolFile(...), 'MOL') = 1;
```

query	hits	time [sec]
	211	0.23
	247	0.1
	4736	3.66
	22	0.1
	565	0.4
	12846	4.6

4 Syntax-Examples

4.1 Molecule Indextype

1. Fill table(s) with structures (CLOB-field) and data
2. Create index on the structure field

```
CREATE INDEX IDX_STRUCTURE ON TEST_TABLE (STRUCT) INDEXTYPE IS  
C$ICHMMOL1.MOL_INDEXTYPE;
```

3. Search using an InfoChem structure operator (SSS, XSS, IXSS, TXSS, PXSS, FlexXSS) and retrieve structure or other chemical information:

```
SELECT iccart.getFormula(a.STRUCT)  
FROM TEST_TABLE a  
WHERE SSS (a.STRUCT, iccart.readMolFile(...), 'MOL') = 1
```

4.2 Reaction Indextype

1. Fill table(s) with reactions and data (e. g. via RDFFile import)
 1. Fill control table with import rules
 2. Import reactions and data

```
BEGIN  
  ICCART.ImportRDFFile('RXN_TEST', 'C:\temp', 'reaction.rdf');  
END;
```

2. Create index on the reaction field

```
CREATE INDEX IDX_RXN ON RXN_TEST (RXN) INDEXTYPE IS  
C$ICHMMOL1.RXN_INDEXTYPE;
```

3. Search the reaction substructure via standard SQL

e.g. a reaction substructure in one specific solvent

```
SELECT a.RIREG, a.RXN  
FROM RXN_TEST a, SOLVENT b  
WHERE RSS (a.RXN, iccart.readRxnFile(...), 'RXN') = 1  
AND XSS(b.STRUCTURE, iccart.readMolFile( ), 'MOL') = 1  
AND a.RIREG = b.RIREG
```

5 System Requirements

Hardware:

Windows/Linux: Pentium III or higher; > 1 GHz recommended

SUN Solaris: SPARC > 700 MHz recommended

up to one million structures: 512 MB additional RAM

for every further million structures: 512 MB additional RAM

up to one million reactions: 800 MB additional RAM

for every further million reactions: 800 MB additional RAM

up to one million structures/reactions: 4 GB free disk space

for every further million structures/reactions: 4 GB additional disk space

Software:

Operating system: Windows 2000

Windows XP

Windows Server 2003

SUN Solaris 8 or higher, 32 / 64 Bit Version

Linux RHEL

Database: Oracle 9.2

Oracle 9.2 (64 bit) for SUN Solaris 64 bit

Oracle 10.1

Oracle 10.1 (64 bit) for SUN Solaris 64 bit

Oracle 10.1 UNICODE character set (verified for Windows)

Oracle 10.2 (verified for Windows)

Oracle XE (verified for Windows)

Oracle 11.1. (verified for Windows)

Oracle 11.2. (verified for Linux (RHEL))