PMI RESEARCH & DEVELOPMENT

Computational Platform for Compound Identification

Aurelien Monge, Stephane Cano, Stephane Albrecht, Pavel Pospisil, Elyette Martin**

Philip Morris International, R&D

19th June 2013

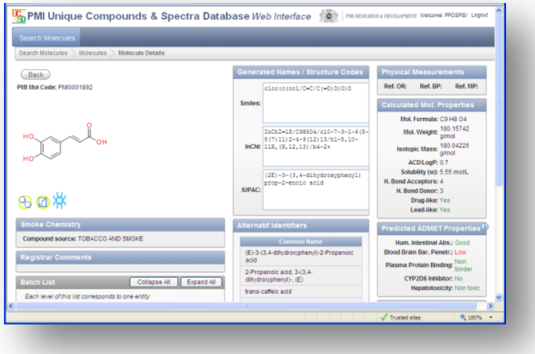
ACD/Labs Symposium on Laboratory Intelligence (EUM), Neuchatel Switzerland

Introduction

Chemoinformatics challenges at PMI:

- Compound identification from complex matrices
- Efficient and/or automatized compound registration (including registration of mixtures and stereochemical isomers)
- Managing spectral data
- Associating toxicity data to compounds and mixtures
- Inserting chemical data in corporate Scientific Data Warehouse
- Reporting chemical data in a relevant way
- Building R&D chemoinformatics platform using Rapid Development Tools

PMI Chemoinformatics platform



PMI Unique Compounds & Spectra Database *Web Interface*

PMI RESEARCH & DEVELOPMENT
Welcome: PPOSPIBI
Logout
Search Molecules
Search Molecules › Molecules › Molecule Details
Back

PMI Mol Code: PMI0001892

Smoke Chemistry

Compound source: TOBACCO AND SMOKE

Registrar Comments

Batch List

Each level of this list corresponds to one entity:

Generated Names / Structure Codes

Smiles: c1cc(c(cc1/C=C/C(=O)O)O)O

InChI: InChI=1S/C9H8O4/c10-7-3-1-6(5-8(7)11)2-4-9(12)13/h1-5,10-11H,(H,12,13)/b4-2+

IUPAC: (2E)-3-(3,4-dihydroxyphenyl)prop-2-enoic acid

Alternatif Identifiers

Common Name
(E)-3-(3,4-dihydroxyphenyl)-2-Propenoic acid
2-Propenoic acid, 3-(3,4-dihydroxyphenyl)-, (E)
trans-caffeic acid

Physical Measurements

Ref. OR: Ref. BP: Ref. MP:

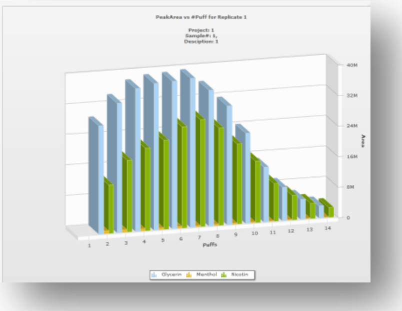
Calculated Mol. Properties

Mol. Formula: C9 H8 O4
Mol. Weight: 180.15742 g/mol
Isotopic Mass: 180.04226 g/mol
ACD/LogP: 0.7
Solubility (w): 5.55 mol/L
H. Bond Acceptors: 4
H. Bond Donor: 3
Drug-like: Yes
Lead-like: Yes

Predicted ADMET Properties

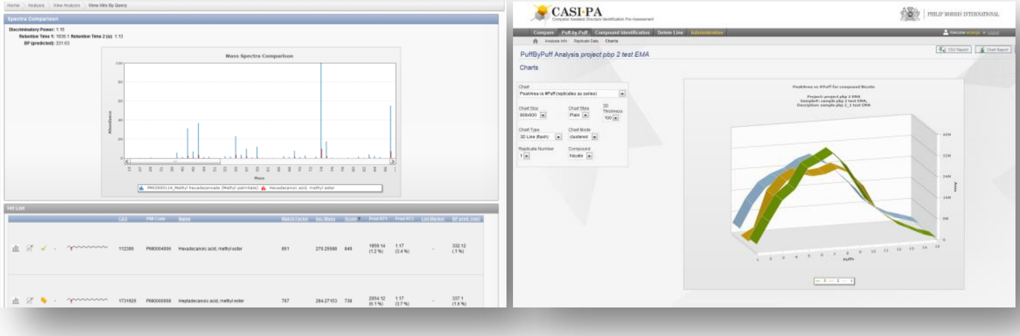
Hum. Intestinal Abs.: Good
Blood Brain Bar. Penetr.: Low
Plasma Protein Binding: Non binder
CYP2D6 Inhibitor: No
Hepatotoxicity: Non toxic

1. Unique Compounds & Spectra Database



3. Chemoinformatics Knowledge Base

Scientific Data Warehouse



Home › Analysis › View Analysis › View Hits By Query

Spectra Comparison

Discriminatory Power: 1.15
Retention Time 1: 1936.1 Retention Time 2 (s): 1.13
BP (predicted): 331.63

Mass Spectra Comparison
10K
8K
6K
4K
2K
0
Abundance
Mass
PMI0000114_Methyl hexadecanoate (Methyl palmitate)
Hexadecanoic acid, methyl ester

Hit List

CAS	PMI Code	Name	Match Factor	Iso. Mass	Score	Pred RT1	Pred RT2	List Marker	BP pred. (var)
112390	PMI0004006	Hexadecanoic acid, methyl ester	851	270.25588	849	1959.14 (1.2 %)	1.17 (3.4 %)	-	332.12 (.1 %)
1731926	PMI0000069	Heptadecanoic acid, methyl ester	767	284.27153	738	2054.12 (6.1 %)	1.17 (3.7 %)	-	337.1 (1.6 %)

CASI·PA
Computer Assisted Structure Identification Pre-Assessment

PuffByPuff Analysis *project pbp 2 test EMA*

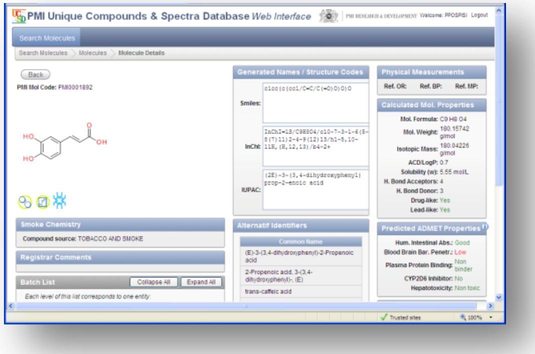
Charts

Chart: PeakArea vs #Puff (replicates as series)
Chart Size: 800x600
Chart Style: Plain
3D Thickness: 100
Chart Type: 3D Line (flash)
Chart Mode: clustered
Replicate Number: 1
Compound: Nicotin

PeakArea vs #Puff for compound Nicotin
Project: project pbp 2 EMA
Sample#: sample pbp 2 test EMA,
Description: sample pbp 2_1 test EMA
40M
32M
24M
16M
8M
0
Area
Puffs
3
2
1

2. Computer Assisted Structure Identification

1. Unique Compounds & Spectra Database



PMI Unique Compounds & Spectra Database *Web Interface*

PMI RESEARCH & DEVELOPMENT Welcome: PPOSPISI Logout

PMI Mol Code: PMI0001892

Generated Names / Structure Codes

Smiles: c1cc(c(cc1/C=C/C(=O)O)O)O

InChI: InChI=1S/C9H8O4/c10-7-3-1-6(5-8(7)11)2-4-9(12)13/h1-5,10-11H,(H,12,13)/b4-2+

IUPAC: (2E)-3-(3,4-dihydroxyphenyl)prop-2-enoic acid

Physical Measurements

Ref. OR: Ref. BP: Ref. MP:

Calculated Mol. Properties

Mol. Formula: C9 H8 O4
Mol. Weight: 180.15742 g/mol
Isotopic Mass: 180.04226 g/mol
ACD/LogP: 0.7
Solubility (w): 5.55 mol/L
H. Bond Acceptors: 4
H. Bond Donor: 3
Drug-like: Yes
Lead-like: Yes

Smoke Chemistry

Compound source: TOBACCO AND SMOKE

Registrar Comments

Batch List

Each level of this list corresponds to one entity:

Alternatif Identifiers

Common Name
(E)-3-(3,4-dihydroxyphenyl)-2-Propenoic acid
2-Propenoic acid, 3-(3,4-dihydroxyphenyl)-, (E)
trans-caffeic acid

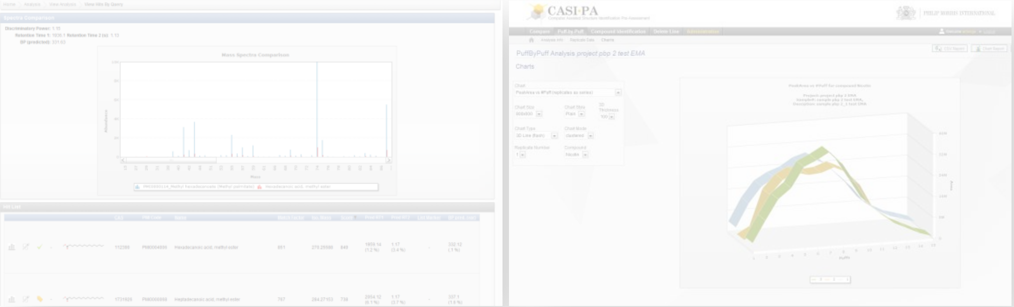
Predicted ADMET Properties

Hum. Intestinal Abs.: Good
Blood Brain Bar. Penetr.: Low
Plasma Protein Binding: Non binder
CYP2D6 Inhibitor: No
Hepatotoxicity: Non toxic

1. Unique Compounds & Spectra Database

3. Chemoinformatics Knowledge Base

Scientific Data Warehouse


Mass Spectra Comparison
CASI·PA
Computer Assisted Structure Identification Pre-Assessment
PHILIP MORRIS INTERNATIONAL
Compare
Puff-by-Puff
Compound Identification
Delete Line
Administration
Analysis Info
Replicate Data
Charts
CSV Report
Chart Report
PeakArea vs #Puff for compound Nicotin
Project: project pbp 2 EMA
Sample#: sample pbp 2 test EMA,
Description: sample pbp 2_3 test EMA

PuffByPuff Analysis *project pbp 2 test EMA*

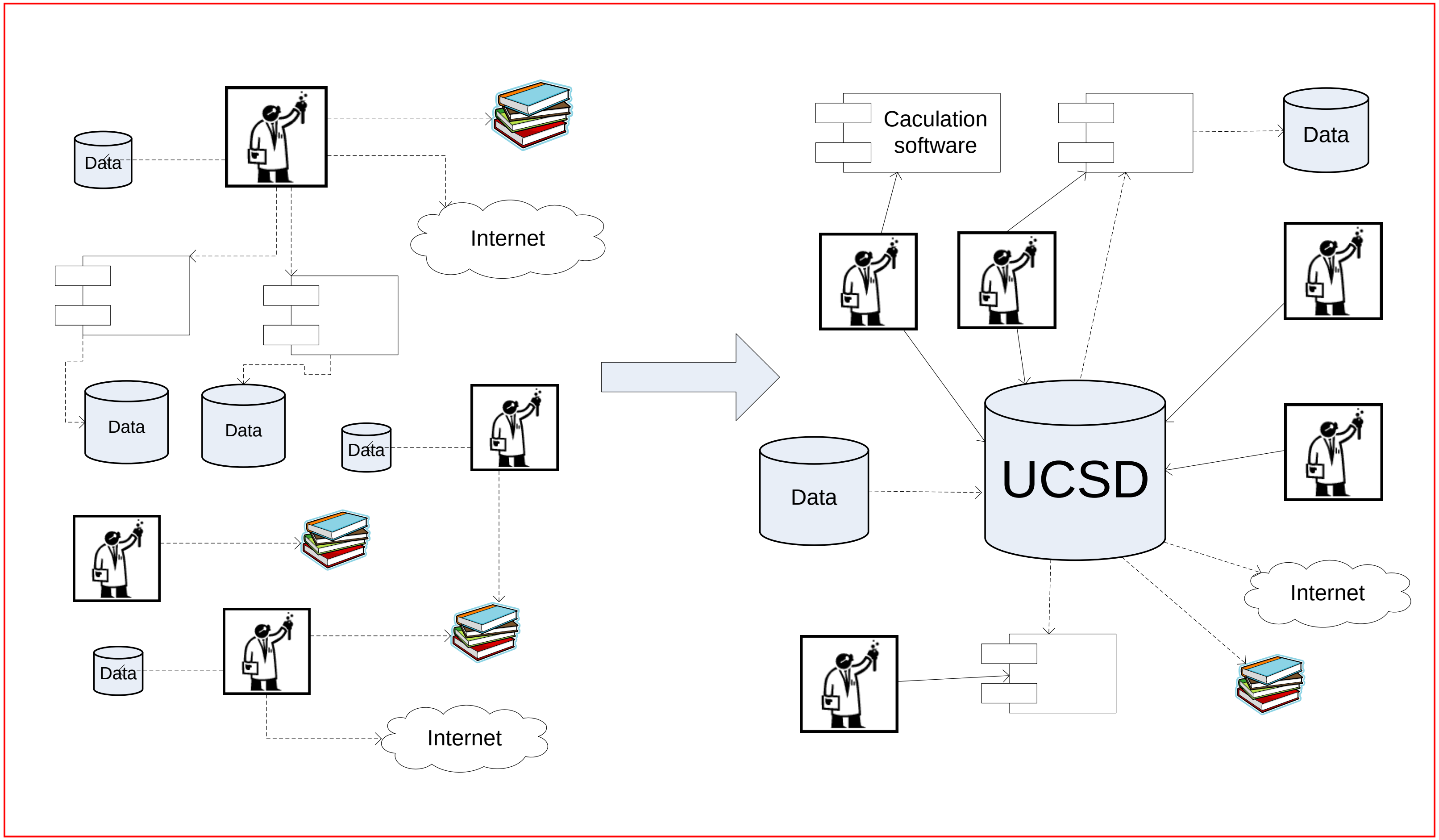
Charts

Chart: PeakArea vs #Puff (replicates as series)
Chart Size: 800x600
Chart Style: Plain
3D Thickness: 100
Chart Type: 3D Line (flash)
Chart Mode: clustered

2. Computer Assisted Structure Identification

PMI RESEARCH & DEVELOPMENT

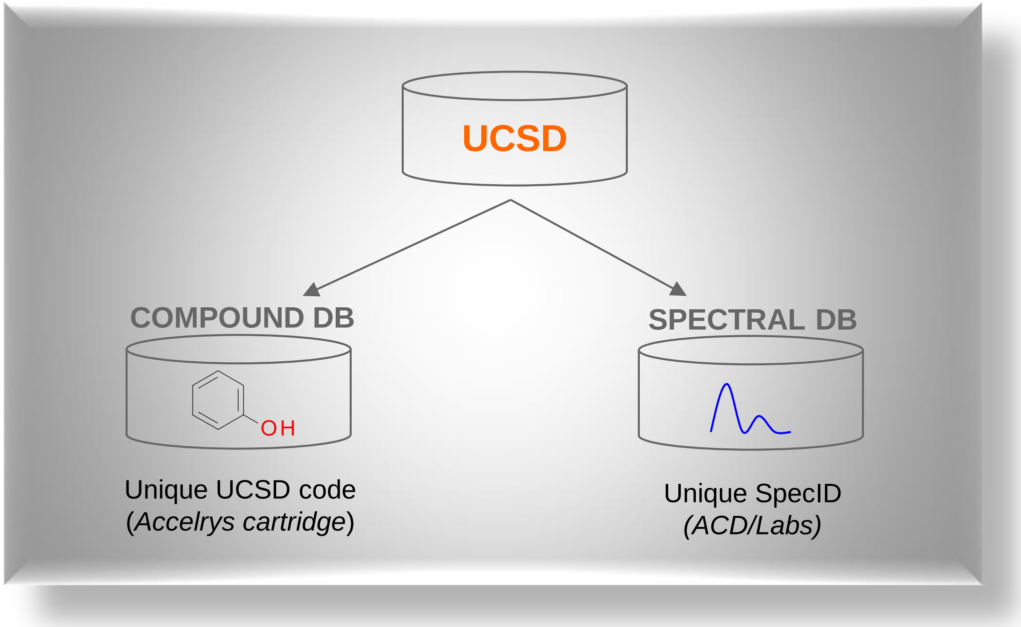
Concept of UCSD



Unique Compound and Spectra Database: To assemble R&D chemical information into a central repository of chemical substances and analytical spectra.

Concept of UCSD

- UCSD is an internal database with no external access
- Accelrys chemical cartridge with enhanced stereochemistry
- ACD/Labs for the analytical spectra part
- Web interface developed in Oracle Application Express and automatic processes in Accelrys Pipeline Pilot



UCSD Roles

1. Submitter

A chemist who can **search**, **view** data and also **submit** molecules and spectra.

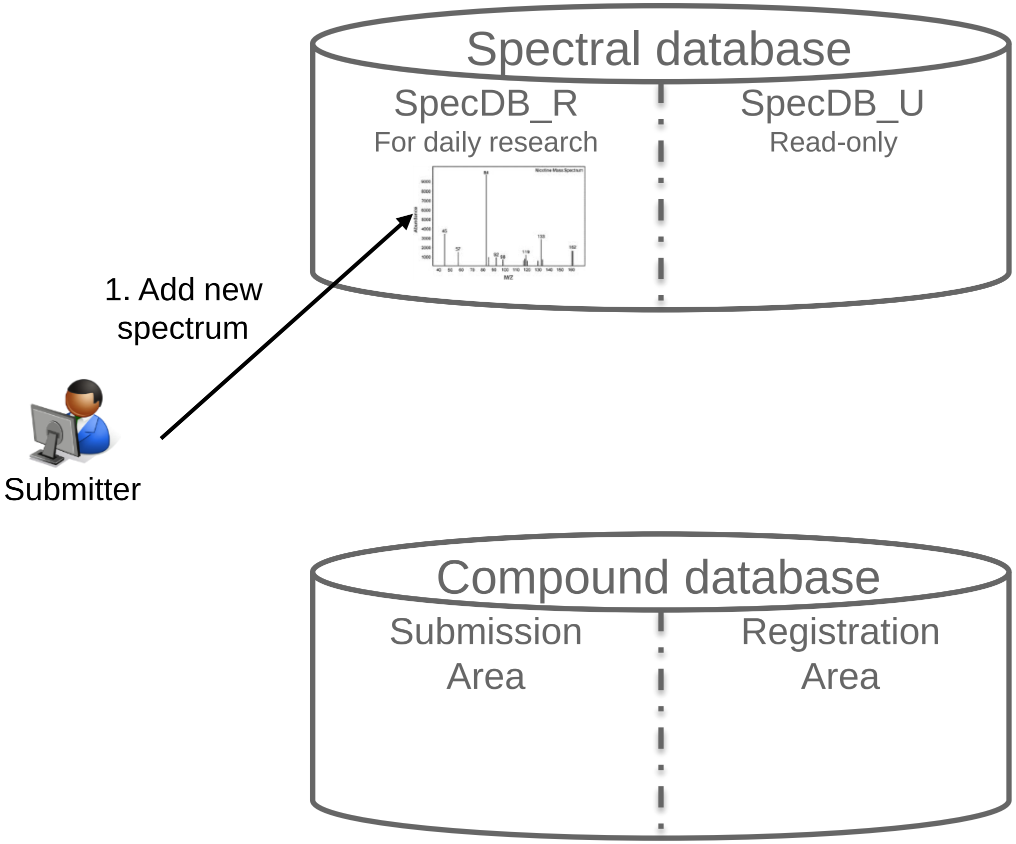
2. Registrar

A chemoinformatician that can **search, view, submit** and also check and **register** data.

Viewer

User can search and **view** registered data (print, export, list creation)

UCSD Workflow



Spectral database

SpecDB_R

For daily research

SpecDB_U

Read-only

1. Add new spectrum

Submitter

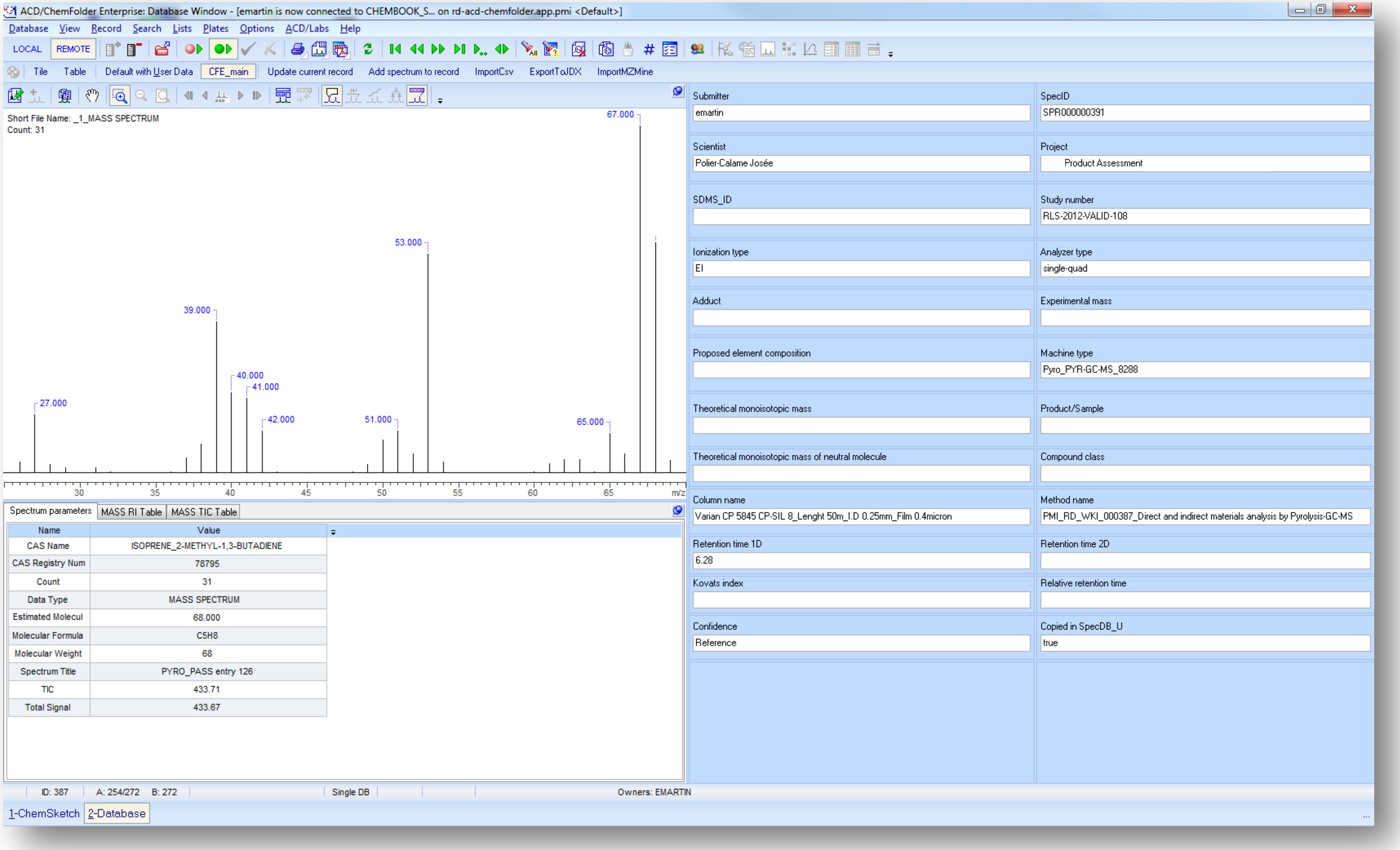
Compound database

Submission Area

Registration Area

SpecDB_R

ACD/ChemFolder Enterprise module



ACD/ChemFolder Enterprise: Database Window - [emartin is now connected to CHEMBOOK_S... on rd-acd-chemfolder.app.pmi <Default>]

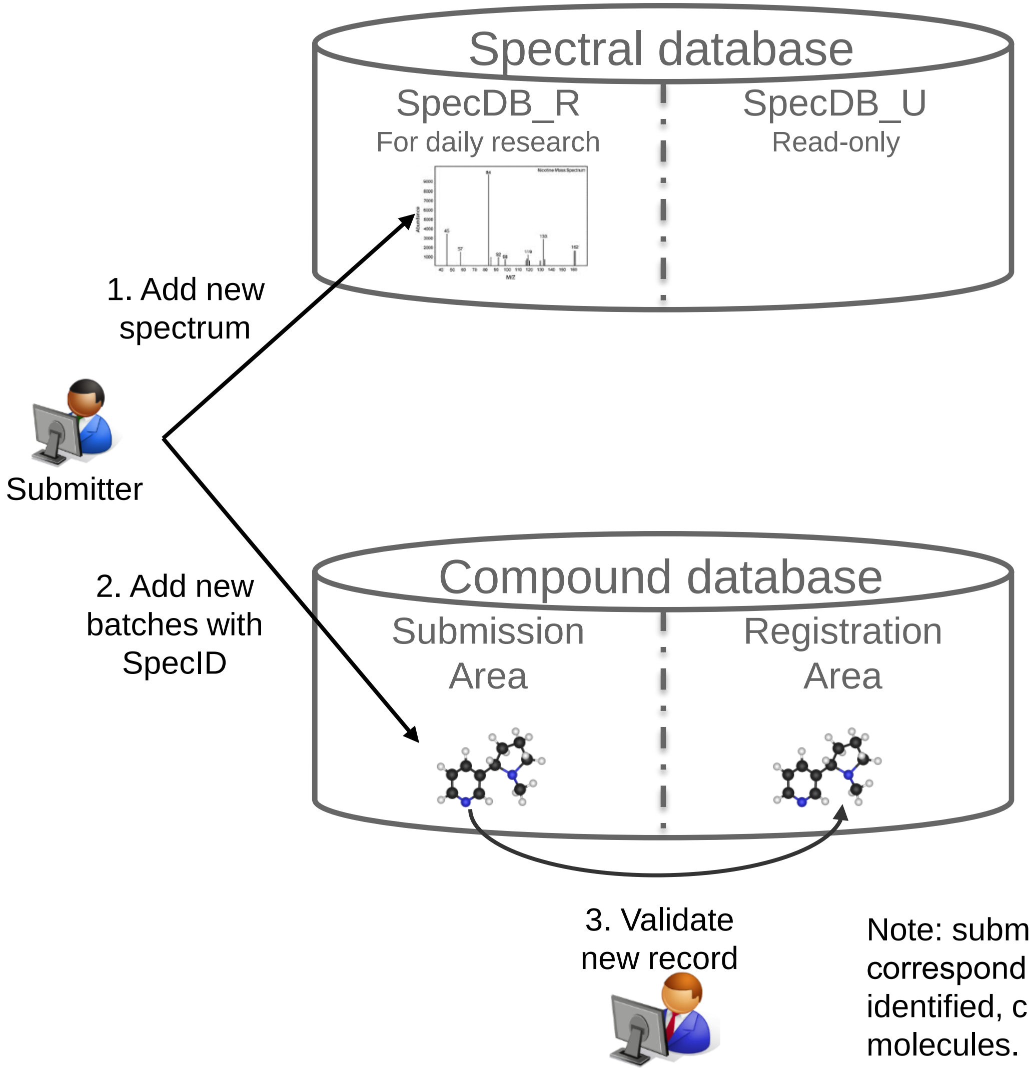
Short File Name: _1_MASS SPECTRUM
Count: 31

Name	Value
CAS Name	ISOPRENE_2-METHYL-1,3-BUTADIENE
CAS Registry Num	78795
Count	31
Data Type	MASS SPECTRUM
Estimated Molecul	68.000
Molecular Formula	C5H8
Molecular Weight	68
Spectrum Title	PYRO_PASS entry 126
TIC	433.71
Total Signal	433.67

Submitter	emartin	SpecID	SPR000000391
Scientist	Polier-Calame Josée	Project	Product Assessment
SDMS_ID		Study number	RLS-2012-VALID-108
Ionization type	EI	Analyzer type	single-quad
Adduct		Experimental mass	
Proposed element composition		Machine type	Pyro_PYR-GC-MS_8288
Theoretical monoisotopic mass		Product/Sample	
Theoretical monoisotopic mass of neutral molecule		Compound class	
Column name	Varian CP 5845 CP-SIL 8_Lenght 50m_I.D 0.25mm_Film 0.4micron	Method name	PMI_RD_WKI_000387_Direct and indirect materials analysis by Pyrolysis-GC-MS
Retention time 1D	6.28	Retention time 2D	
Kovats index		Relative retention time	
Confidence	Reference	Copied in SpecDB_U	true

ID: 387 A: 254/272 B: 272 Single DB Owners: EMARTIN

UCSD Workflow



Spectral database

SpecDB_R
For daily research

SpecDB_U
Read-only

1. Add new spectrum

Submitter

2. Add new batches with SpecID

Compound database

Submission Area

Registration Area

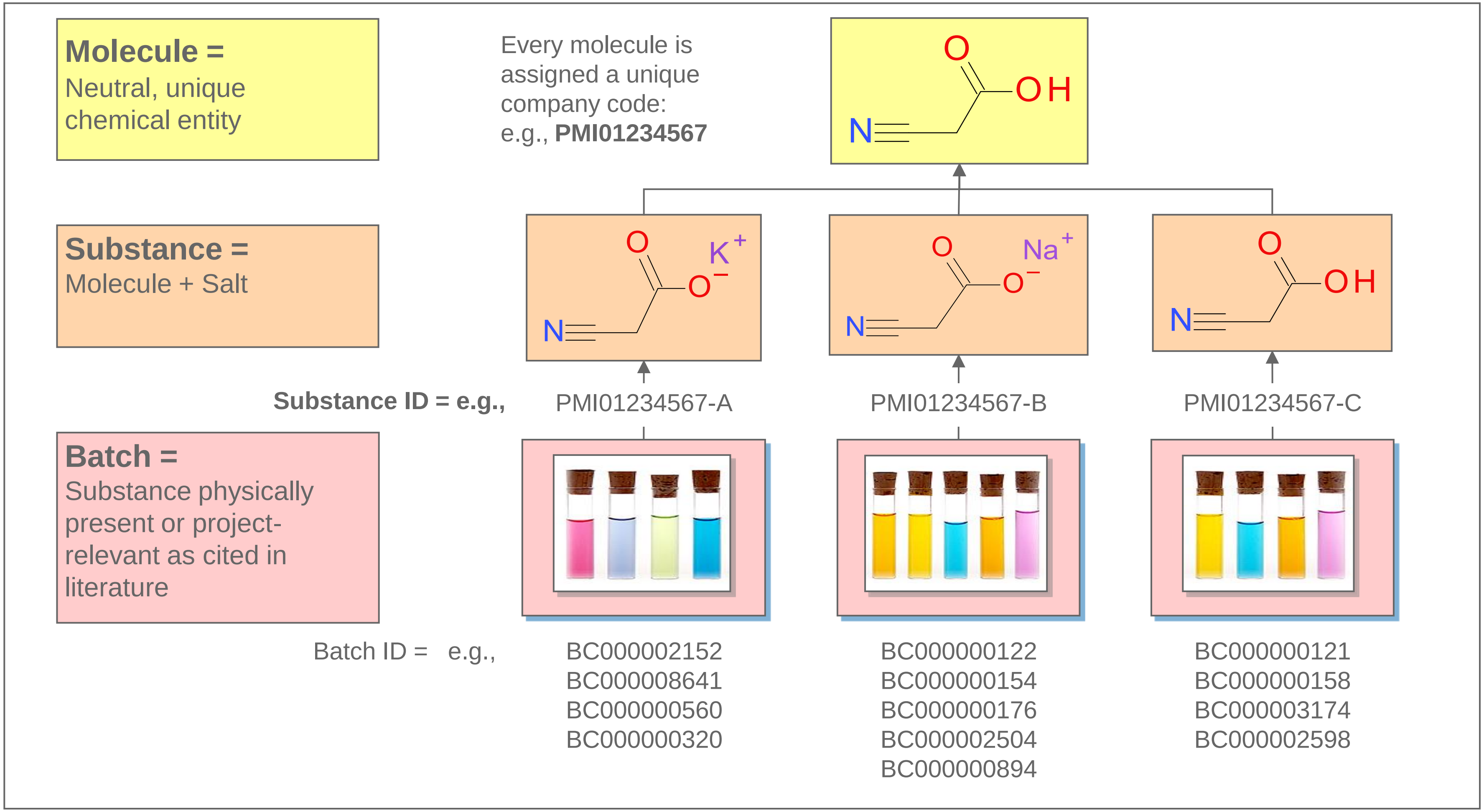
3. Validate new record

Note: submitted batches can also correspond to ‘known unknowns’ that once identified, can be later registered as known molecules.

PMI RESEARCH & DEVELOPMENT

Concept of Molecule, Substance and Batch Assignment

Unique means there is no redundancy, ensured by uniqueness of chemical structure and systematic and accurate registration process.



Molecule =
Neutral, unique chemical entity

Every molecule is assigned a unique company code:
e.g., **PMI01234567**

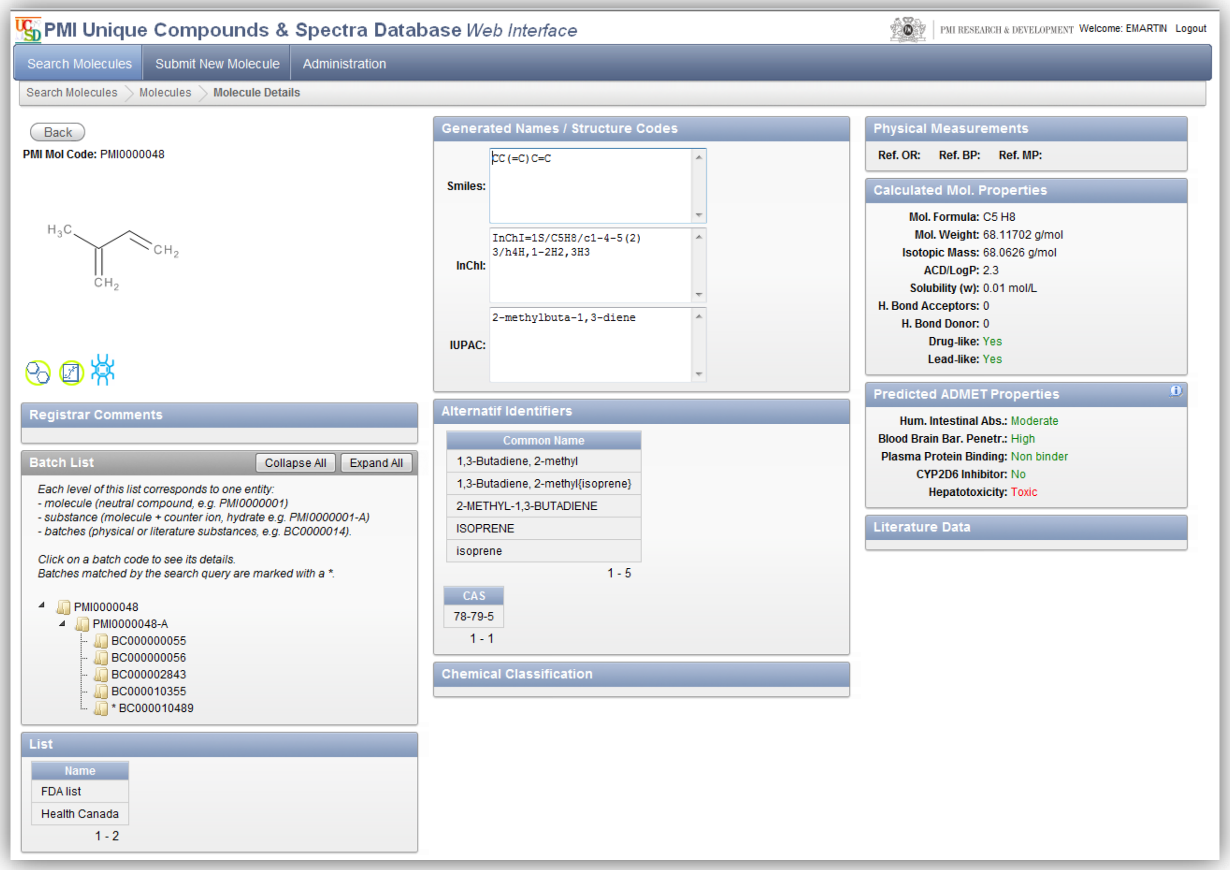
Substance =
Molecule + Salt

Substance ID = e.g., PMI01234567-A PMI01234567-B PMI01234567-C

Batch =
Substance physically present or project-relevant as cited in literature

Batch ID = e.g.,	PMI01234567-A	PMI01234567-B	PMI01234567-C
	BC000002152 BC000008641 BC000000560 BC000000320	BC000000122 BC000000154 BC000000176 BC000002504 BC000000894	BC000000121 BC000000158 BC000003174 BC000002598

UCSD Compound Part


UCSD PMI Unique Compounds & Spectra Database *Web Interface*
PMI RESEARCH & DEVELOPMENT Welcome: EMARTIN Logout
Search Molecules
Submit New Molecule
Administration
Search Molecules > Molecules > Molecule Details
Back
H_3C
CH_2
CH_2
Collapse All
Expand All

PMI Mol Code: PMI0000048

Generated Names / Structure Codes

Smiles: CC(=C)C=C

InChI: InChI=1S/C5H8/c1-4-5(2)3/h4H,1-2H2,3H3

IUPAC: 2-methylbuta-1,3-diene

Physical Measurements

Ref. OR: **Ref. BP:** **Ref. MP:**

Calculated Mol. Properties

Mol. Formula: C5 H8
Mol. Weight: 68.11702 g/mol
Isotopic Mass: 68.0626 g/mol
ACD/LogP: 2.3
Solubility (w): 0.01 mol/L
H. Bond Acceptors: 0
H. Bond Donor: 0
Drug-like: Yes
Lead-like: Yes

Registrar Comments

Batch List

Each level of this list corresponds to one entity:
- molecule (neutral compound, e.g. PMI0000001)
- substance (molecule + counter ion, hydrate e.g. PMI0000001-A)
- batches (physical or literature substances, e.g. BC0000014).

Click on a batch code to see its details.
*Batches matched by the search query are marked with a *.*

- PMI0000048
  - PMI0000048-A
    - BC000000055
    - BC000000056
    - BC000002843
    - BC000010355
    - * BC000010489

Alternatif Identifiers

Common Name
1,3-Butadiene, 2-methyl
1,3-Butadiene, 2-methyl{isoprene}
2-METHYL-1,3-BUTADIENE
ISOPRENE
isoprene

1 - 5

CAS
78-79-5

1 - 1

Chemical Classification

Predicted ADMET Properties

Hum. Intestinal Abs.: Moderate
Blood Brain Bar. Penetr.: High
Plasma Protein Binding: Non binder
CYP2D6 Inhibitor: No
Hepatotoxicity: Toxic

Literature Data

List

Name
FDA list
Health Canada

1 - 2

- To each newly registered compound we developed automatic processes that:

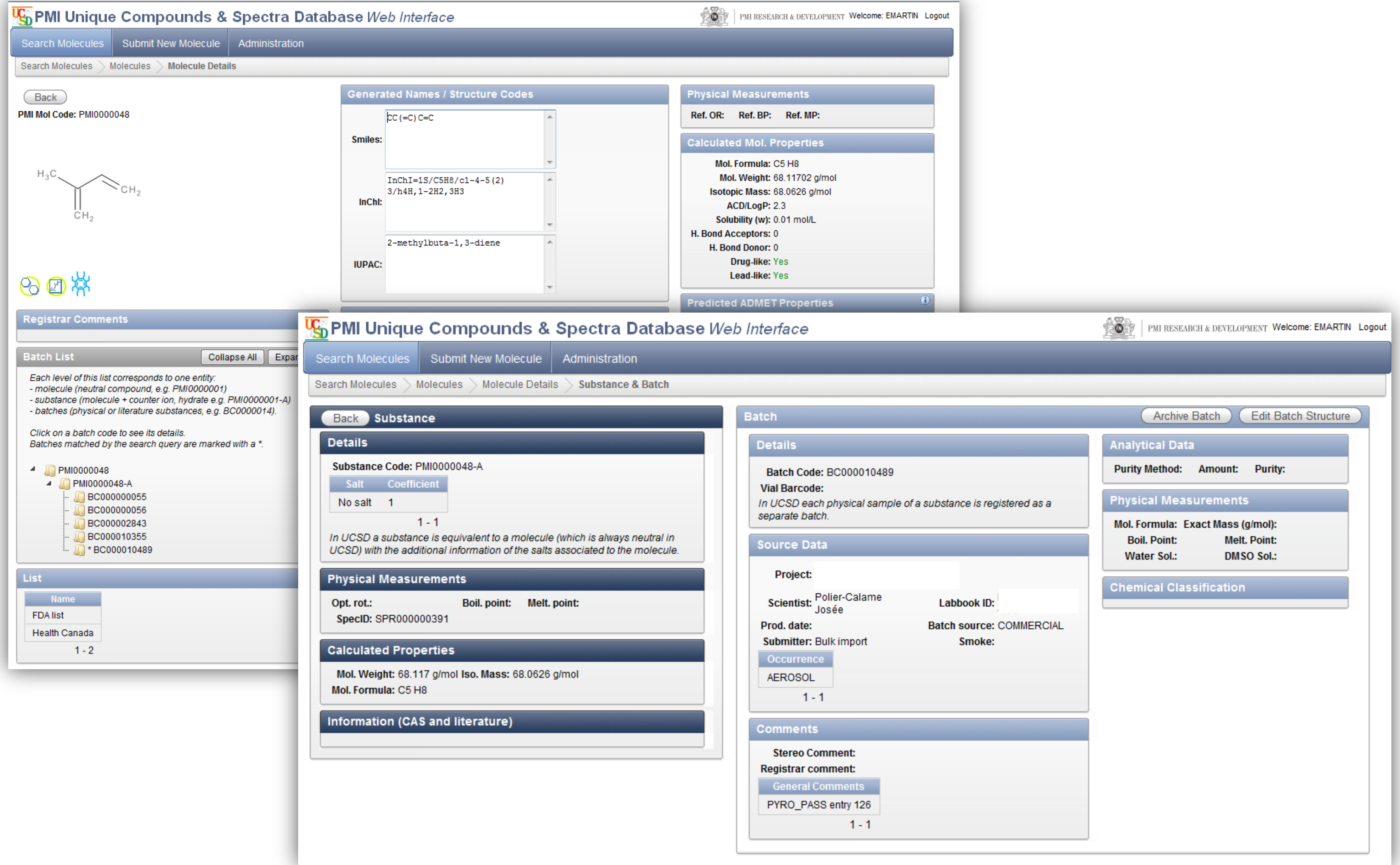
- attach SMILES strings, InChI codes and IUPAC names

- calculate structure-derived physico-chemical and ADMET properties

⇒ Naming and calculation of compounds after registration (uniqueness check) reduces further ambiguity

See more in Martin et al., *Journal of Cheminformatics* 2012, 4:11

UCSD Compound Part



PMI Unique Compounds & Spectra Database *Web Interface*
PMI RESEARCH & DEVELOPMENT
Welcome: EMARTIN
Logout
Search Molecules
Submit New Molecule
Administration
Search Molecules > Molecules > Molecule Details
Back
H_3C
CH_2
CH_2

PMI Mol Code: PMI0000048

Generated Names / Structure Codes

Smiles: CC(=C)C=C

InChI: InChI=1S/C5H8/c1-4-5(2)3/h4H,1-2H2,3H3

IUPAC: 2-methylbuta-1,3-diene

Physical Measurements

Ref. OR: **Ref. BP:** **Ref. MP:**

Calculated Mol. Properties

Mol. Formula: C5 H8
Mol. Weight: 68.11702 g/mol
Isotopic Mass: 68.0626 g/mol
ACD/LogP: 2.3
Solubility (w): 0.01 mol/L
H. Bond Acceptors: 0
H. Bond Donor: 0
Drug-like: Yes
Lead-like: Yes

Predicted ADMET Properties

Registrar Comments

Batch List

Each level of this list corresponds to one entity:
- molecule (neutral compound, e.g. PMI0000001)
- substance (molecule + counter ion, hydrate e.g. PMI0000001-A)
- batches (physical or literature substances, e.g. BC0000014).

Click on a batch code to see its details.
*Batches matched by the search query are marked with a *.*

- PMI0000048
  - PMI0000048-A
    - BC000000055
    - BC000000056
    - BC000002843
    - BC000010355
    - * BC000010489

List

Name
FDA list
Health Canada

1 - 2

PMI Unique Compounds & Spectra Database *Web Interface*
PMI RESEARCH & DEVELOPMENT
Welcome: EMARTIN
Logout
Search Molecules
Submit New Molecule
Administration
Search Molecules > Molecules > Molecule Details > Substance & Batch
Back
Archive Batch
Edit Batch Structure

Substance

Details

Substance Code: PMI0000048-A

Salt	Coefficient
No salt	1

1 - 1

In UCSD a substance is equivalent to a molecule (which is always neutral in UCSD) with the additional information of the salts associated to the molecule.

Physical Measurements

Opt. rot.: **Boil. point:** **Melt. point:**
SpecID: SPR000000391

Calculated Properties

Mol. Weight: 68.117 g/mol **Iso. Mass:** 68.0626 g/mol
Mol. Formula: C5 H8

Information (CAS and literature)

Batch

Details

Batch Code: BC000010489
Vial Barcode:

In UCSD each physical sample of a substance is registered as a separate batch.

Source Data

Project:
Scientist: Polier-Calame Josée **Labbook ID:**
Prod. date: **Batch source:** COMMERCIAL
Submitter: Bulk import **Smoke:**

Occurrence
AEROSOL

1 - 1

Comments

Stereo Comment:
Registrar comment:

General Comments
PYRO_PASS entry 126

1 - 1

Analytical Data

Purity Method: **Amount:** **Purity:**

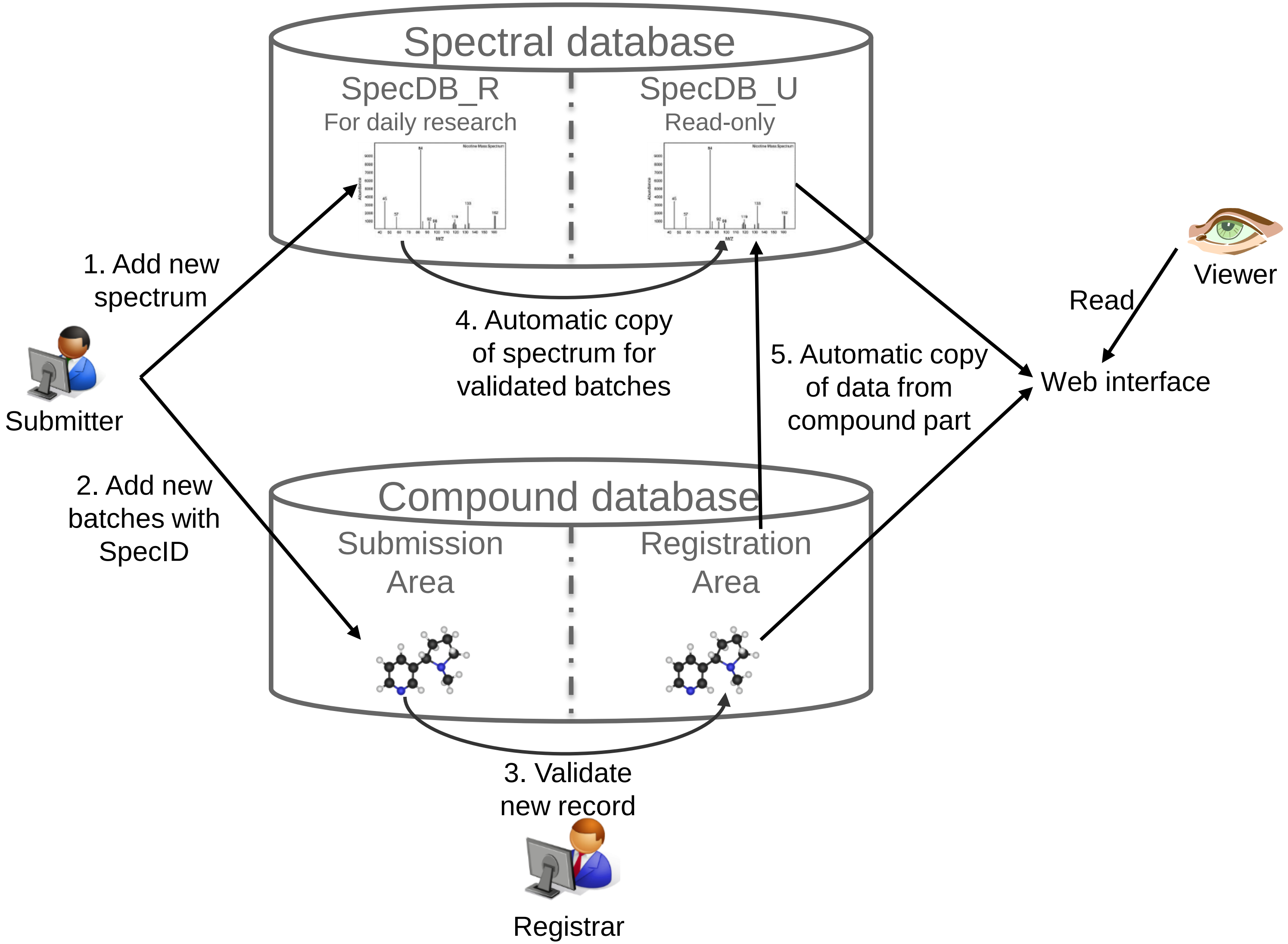
Physical Measurements

Mol. Formula: **Exact Mass (g/mol):**
Boil. Point: **Melt. Point:**
Water Sol.: **DMSO Sol.:**

Chemical Classification

See more in Martin et al., *Journal of Cheminformatics* 2012, 4:11

UCSD Workflow



Spectral database

SpecDB_R
For daily research

SpecDB_U
Read-only

1. Add new spectrum

Submitter

2. Add new batches with SpecID

4. Automatic copy of spectrum for validated batches

5. Automatic copy of data from compound part

Viewer

Read

Web interface

Compound database

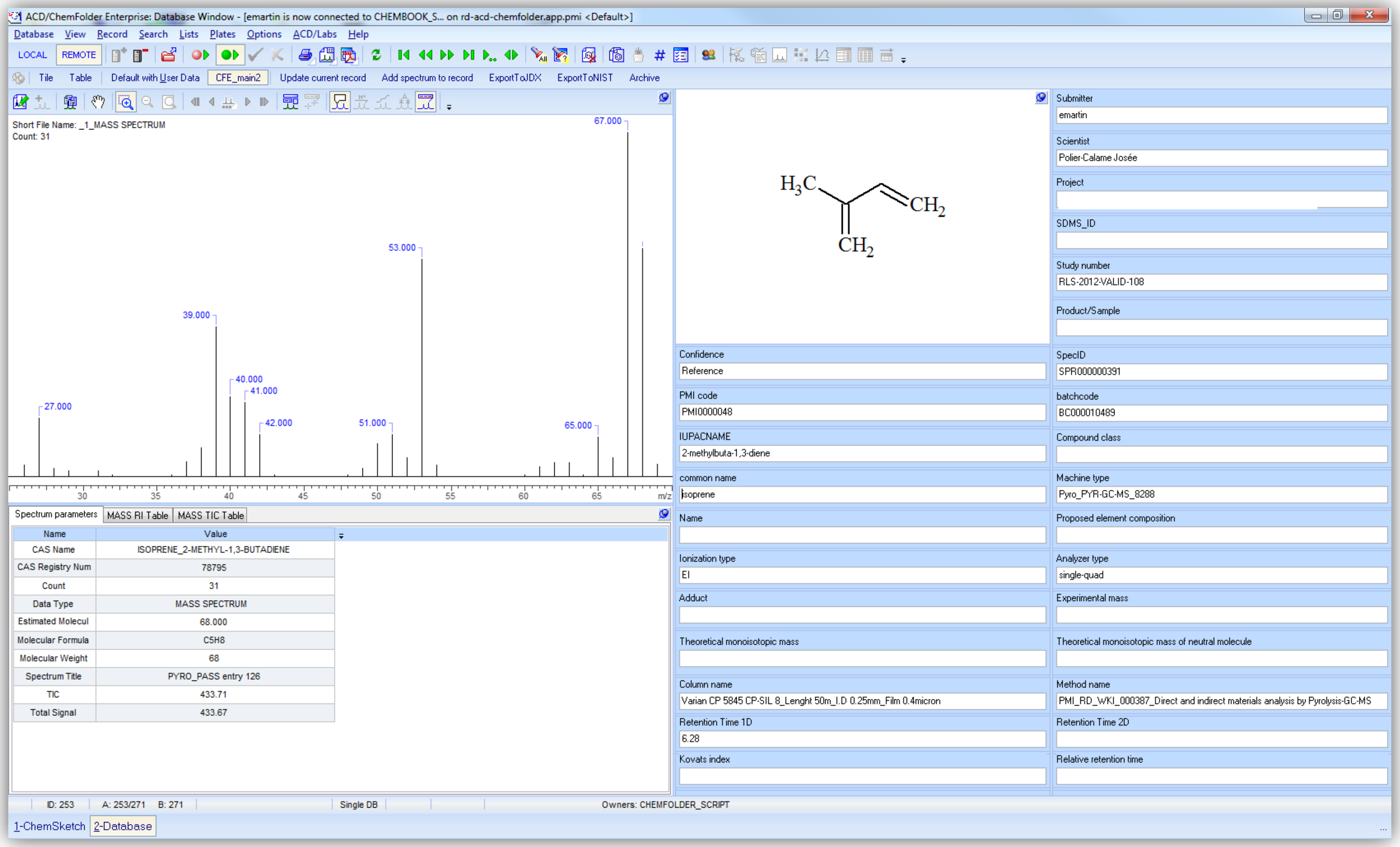
Submission Area

Registration Area

3. Validate new record

Registrar

SpecDB_U



Short File Name: _1_MASS SPECTRUM
Count: 31

Name	Value
CAS Name	ISOPRENE_2-METHYL-1,3-BUTADIENE
CAS Registry Num	78795
Count	31
Data Type	MASS SPECTRUM
Estimated Molecul	68.000
Molecular Formula	C5H8
Molecular Weight	68
Spectrum Title	PYRO_PASS entry 126
TIC	433.71
Total Signal	433.67

Confidence: Reference

PMI code: PMI0000048

IUPACNAME: 2-methylbuta-1,3-diene

common name: isoprene

Name:

Ionization type: EI

Adduct:

Theoretical monoisotopic mass:

Column name: Varian CP 5845 CP-SIL 8_Lenght 50m_I.D 0.25mm_Film 0.4micron

Retention Time 1D: 6.28

Kovats index:

Submitter: emartin

Scientist: Polier-Calame Josée

Project:

SDMS_ID:

Study number: RLS-2012-VALID-108

Product/Sample:

SpecID: SPR000000391

batchcode: BC000010489

Compound class:

Machine type: Pyro_PYR-GC-MS_8288

Proposed element composition:

Analyzer type: single-quad

Experimental mass:

Theoretical monoisotopic mass of neutral molecule:

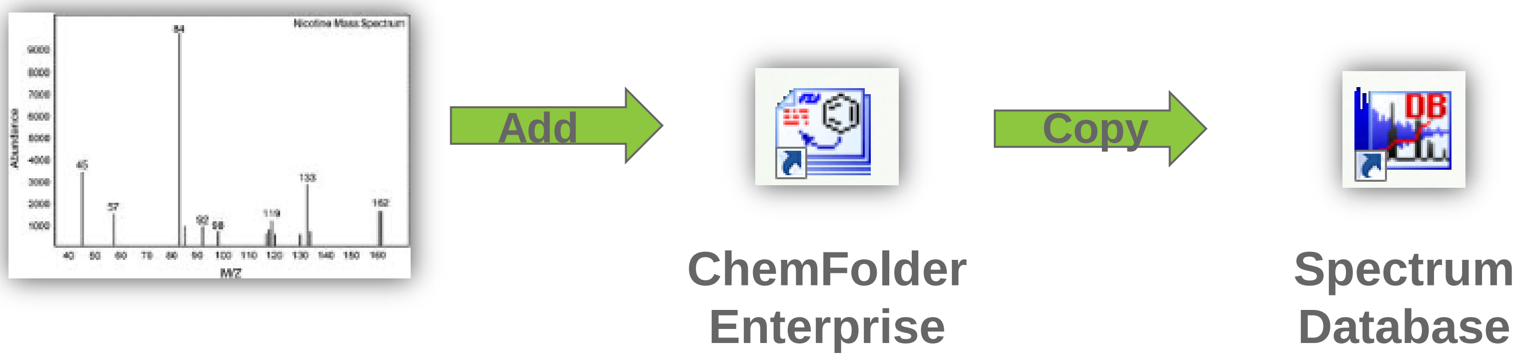
Method name: PMI_RD_WKI_000387_Direct and indirect materials analysis by Pyrolysis-GC-MS

Retention Time 2D:

Relative retention time:

In waiting for Spectrus

Search by spectra similarity is not possible in ChemFolder module ➔ implementation of a script that copies each entry in Spectrum Database module



Add

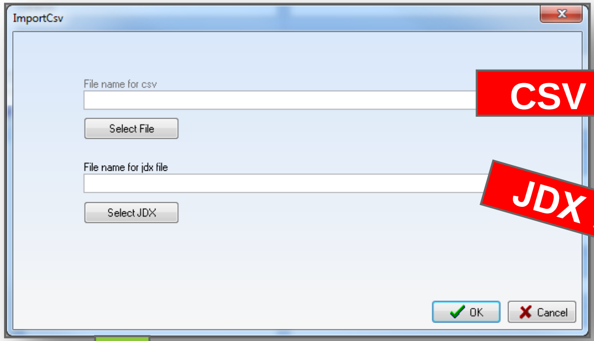
Copy

ChemFolder Enterprise

Spectrum Database

Process for bulk import

INPUT for SpecDB_R

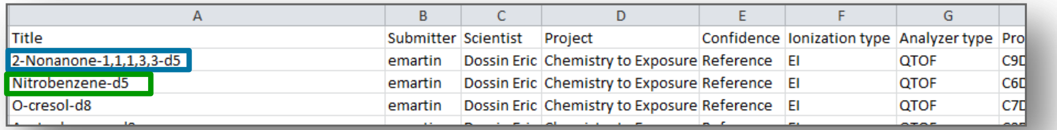


ImportCsv

File name for csv

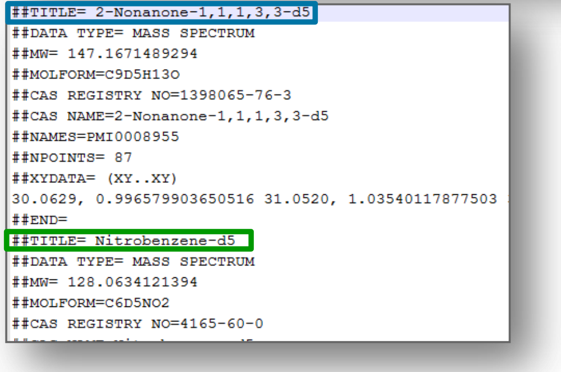
File name for jdx file

CSV file



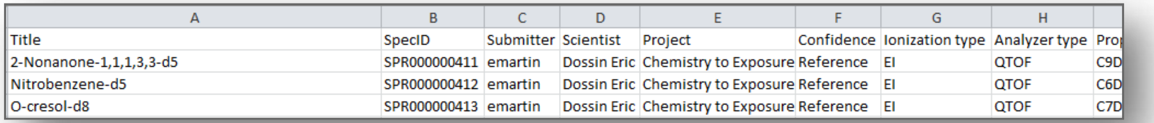
	A	B	C	D	E	F	G	
	Title	Submitter	Scientist	Project	Confidence	Ionization type	Analyzer type	Pro
	2-Nonanone-1,1,1,3,3-d5	emartin	Dossin Eric	Chemistry to Exposure	Reference	EI	QTOF	C9D
	Nitrobenzene-d5	emartin	Dossin Eric	Chemistry to Exposure	Reference	EI	QTOF	C6D
	O-cresol-d8	emartin	Dossin Eric	Chemistry to Exposure	Reference	EI	QTOF	C7D

JDX file

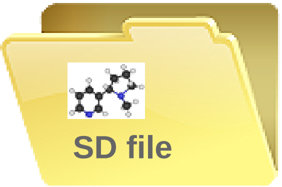


```
##TITLE= 2-Nonanone-1,1,1,3,3-d5
##DATA TYPE= MASS SPECTRUM
##MW= 147.1671489294
##MOLFORM=C9D5H13O
##CAS REGISTRY NO=1398065-76-3
##CAS NAME=2-Nonanone-1,1,1,3,3-d5
##NAMES=PMI0008955
##NPOINTS= 87
##XYDATA= (XY..XY)
30.0629, 0.996579903650516 31.0520, 1.03540117877503
##END=
##TITLE= Nitrobenzene-d5
##DATA TYPE= MASS SPECTRUM
##MW= 128.0634121394
##MOLFORM=C6D5NO2
##CAS REGISTRY NO=4165-60-0
```

OUTPUT

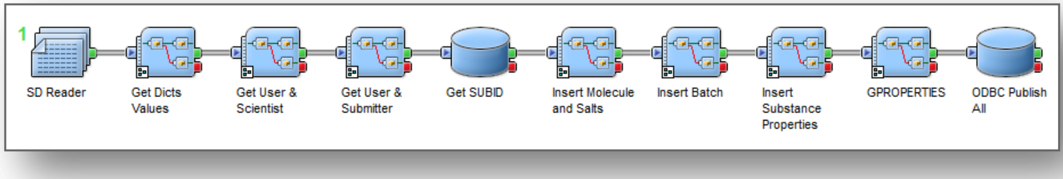


	A	B	C	D	E	F	G	H	
	Title	SpecID	Submitter	Scientist	Project	Confidence	Ionization type	Analyzer type	Pro
	2-Nonanone-1,1,1,3,3-d5	SPR000000411	emartin	Dossin Eric	Chemistry to Exposure	Reference	EI	QTOF	C9D
	Nitrobenzene-d5	SPR000000412	emartin	Dossin Eric	Chemistry to Exposure	Reference	EI	QTOF	C6D
	O-cresol-d8	SPR000000413	emartin	Dossin Eric	Chemistry to Exposure	Reference	EI	QTOF	C7D

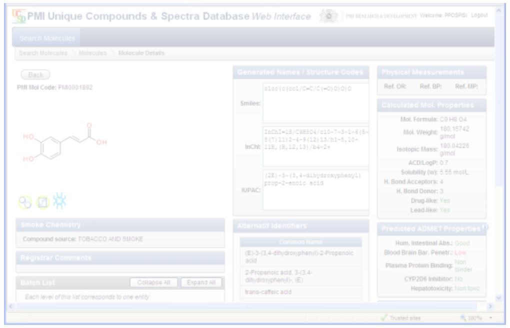


SD file

Merge and import in the compound part via Pipeline Pilot protocol



2. Computer Assisted Structure Identification



PMI Unique Compounds & Spectra Database *Web Interface*

PMI Mol Code: PMI0001892

Smoke Chemistry

Compound source: TOBACCO AND SMOKE

Registrar Comments

Batch List

Each level of this list corresponds to one entity

Generated Names / Structure Codes

Smiles: c1cc(c(cc1/C=C/C(=O)O)O)O

InChI: InChI=1S/C9H8O4/c10-7-3-1-6(5-8(7)11)2-4-9(12)13/h1-5,10-11H,(H,12,13)/b4-2+

IUPAC: (2E)-3-(3,4-dihydroxyphenyl)prop-2-enoic acid

Alternatif Identifiers

Common Name
(E)-3-(3,4-dihydroxyphenyl)-2-Propenoic acid
2-Propenoic acid, 3-(3,4-dihydroxyphenyl)-, (E)
trans-caffeic acid

Physical Measurements

Ref. OR: Ref. BP: Ref. MP:

Calculated Mol. Properties

Mol. Formula: C9 H8 O4
Mol. Weight: 180.15742 g/mol
Isotopic Mass: 180.04226 g/mol
ACD/LogP: 0.7
Solubility (w): 5.55 mol/L
H. Bond Acceptors: 4
H. Bond Donor: 3
Drug-like: Yes
Lead-like: Yes

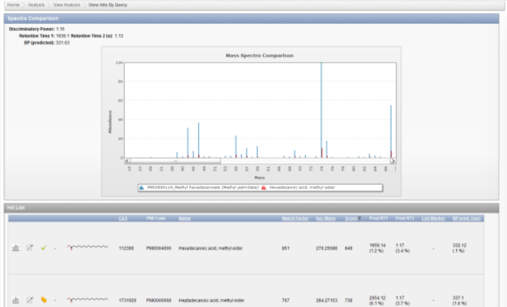
Predicted ADMET Properties

Hum. Intestinal Abs.: Good
Blood Brain Bar. Penetr.: Low
Plasma Protein Binding: Non binder
CYP2D6 Inhibitor: No
Hepatotoxicity: Non toxic

1. Unique Compounds & Spectra Database

3. Chemoinformatics Knowledge Base

Scientific Data Warehouse



Home > Analysis > View Analysis > View Hits By Query

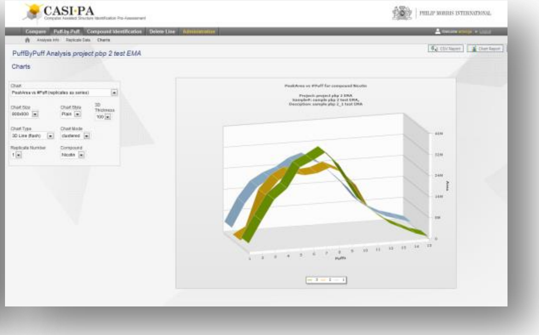
Spectra Comparison

Discriminatory Power: 1.15
Retention Time 1: 1936.1 Retention Time 2 (s): 1.13
BP (predicted): 331.63

Mass Spectra Comparison
Abundance
Mass
0
2K
4K
6K
8K
10K
PMI0000114_Methyl hexadecanoate (Methyl palmitate)
Hexadecanoic acid, methyl ester

Hit List

CAS	PMI Code	Name	Match Factor	Iso. Mass	Score	Pred RT1	Pred RT2	List Marker	BP pred. (rar)
112390	PMI0004006	Hexadecanoic acid, methyl ester	851	270.25588	849	1959.14 (1.2 %)	1.17 (3.4 %)	-	332.12 (.1 %)
1731926	PMI0000068	Heptadecanoic acid, methyl ester	767	284.27153	738	2054.12 (6.1 %)	1.17 (3.7 %)	-	337.1 (1.6 %)



CASI·PA

Computer Assisted Structure Identification Pre-Assessment

PuffByPuff Analysis *project pbp 2 test EMA*

Charts

Chart: PeakArea vs #Puff (replicates as series)
Chart Size: 800x800
Chart Style: Plain
3D Thickness: 100
Chart Type: 3D Line (flash)
Chart Mode: clustered
Replicate Number: 1
Compound: Nicotin

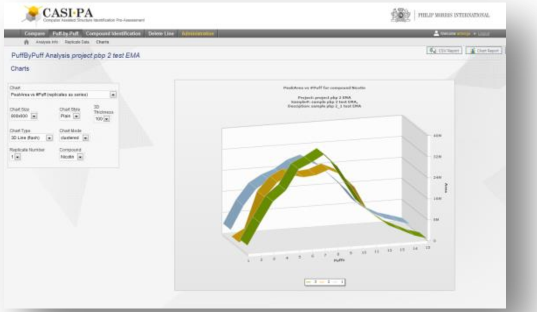
PeakArea vs #Puff for compound Nicotin
Project: project pbp 2 EMA
Sample#: sample pbp 2 test EMA,
Description: sample pbp 2_1 test EMA
Area
Puffs
0
8M
16M
24M
32M
40M
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

2. Computer Assisted Structure Identification

A. Monge & E. Martin, PMI R&D

Computer Assisted Structure Identification

Our in-house developed Computer Assisted Structure Identification platform is constituted of two software:



PuffByPuff Analysis project pbp 2 test EMA

Charts

Chart: PeakArea vs #Puff (replicates as series)

Chart Size: 800x600

Chart Style: Plain

3D Thickness: 100

Chart Type: 3D Line (flash)

Chart Mode: clustered

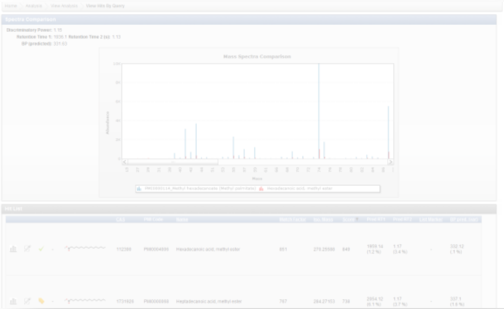
Replicate Number: 1

Compound: Nicotin

Project: project pbp 2 EMA

Sample#: sample pbp 2 test EMA,

Description: sample pbp 2_1 test EMA



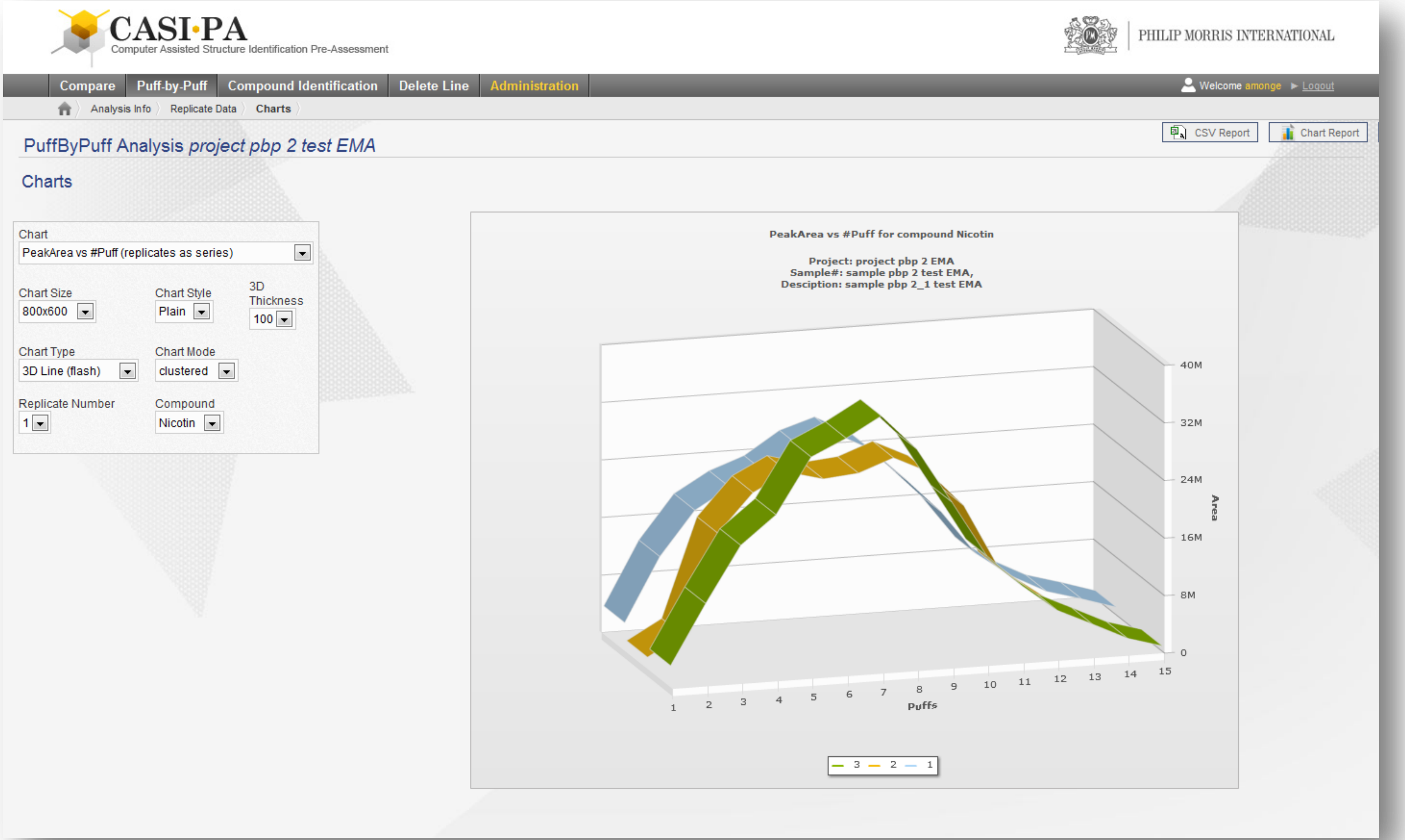
Software 1 “CASI-PA”: assists the processing of MS data for given analyses:

- Compound Identification
- Comparison of 2 samples
- Puff-by-Puff

Automated non-target processing of MS data.

Puff-by-Puff Analysis of Cigarette Aerosols

Analysis of the concentration of given chemical compounds per puff.


CASI·PA
Computer Assisted Structure Identification Pre-Assessment
PHILIP MORRIS INTERNATIONAL
Compare
Puff-by-Puff
Compound Identification
Delete Line
Administration
Welcome amonge
Logout
Analysis Info
Replicate Data
Charts
CSV Report
Chart Report

PuffByPuff Analysis *project pbp 2 test EMA*

Charts

Chart: PeakArea vs #Puff (replicates as series)

Chart Size: 800x600

Chart Style: Plain

3D Thickness: 100

Chart Type: 3D Line (flash)

Chart Mode: clustered

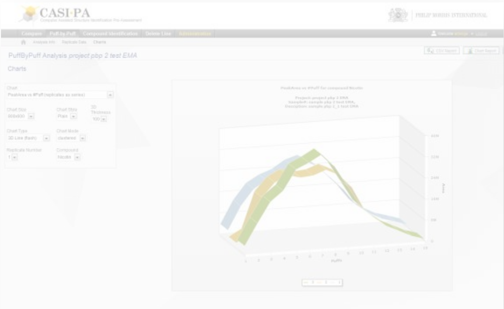
Replicate Number: 1

Compound: Nicotin

PeakArea vs #Puff for compound Nicotin
Project: project pbp 2 EMA
Sample#: sample pbp 2 test EMA,
Desciption: sample pbp 2_1 test EMA
40M
32M
24M
16M
8M
0
Area
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Puffs
3
2
1

Computer Assisted Structure Identification

Our Computer Assisted Structure Identification platform is constituted of two software:



CASI·PA
Computer Assisted Structure Identification Pre-Assessment

PuffByPuff Analysis project pbp 2 test EMA

Charts

Chart: PeakArea vs #Puff (replicates as series)

Chart Size: 800x600

Chart Style: Plain

3D Thickness: 100

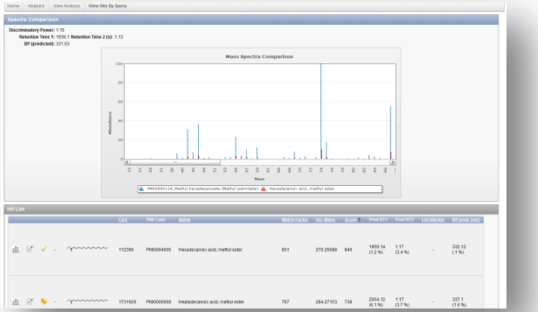
Chart Type: 3D Line (flash)

Chart Mode: clustered

Replicate Number: 1

Compound: Nicotin

PeakArea vs #Puff for compound Nicotin
Project: project pbp 2 EMA
Sample#: sample pbp 2 test EMA,
Description: sample pbp 2_1 test EMA
40M
32M
24M
16M
8M
0
Area
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Puff#



Spectra Comparison

Discriminatory Power: 1.15

Retention Time 1: 1936.1 Retention Time 2 (s): 1.13

BP (predicted): 331.63

Mass Spectra Comparison
10K
8K
6K
4K
2K
0
Abundance
Mass
PMI0000114_Methyl hexadecanoate (Methyl palmitate)
Hexadecanoic acid, methyl ester

Hit List

CAS	PMI Code	Name	Match Factor	Iso. Mass	Score	Pred RT1	Pred RT2	List Marker	BP pred. (rar)
112390	PMI0004806	Hexadecanoic acid, methyl ester	851	270.25588	849	1959.14 (1.2 %)	1.17 (3.4 %)	-	332.12 (.1 %)
1731926	PMI0000868	Heptadecanoic acid, methyl ester	767	284.27153	738	2054.12 (6.1 %)	1.17 (3.7 %)	-	337.1 (1.6 %)

Help for the processing of MS data for given analyses:

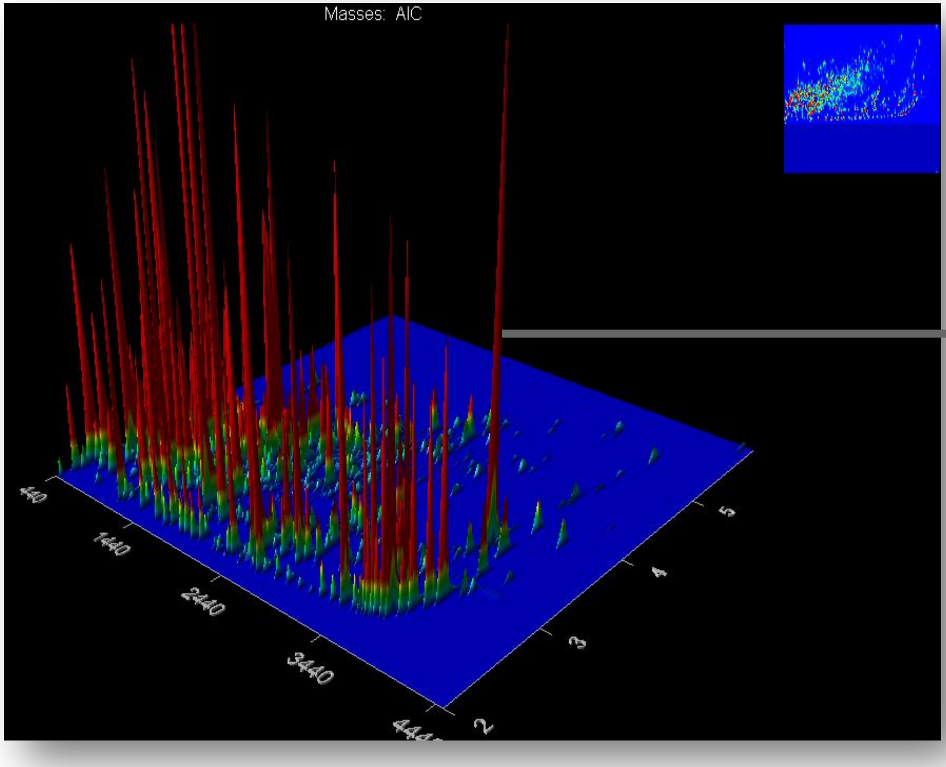
- Compound Identification
- Puff-by-Puff
- Comparison of two products

Software 2: Automated non-target processing of MS data.

Computer Assisted Structure Identification

Automated platform to accelerate and standardize identification of chemical structures of aerosol constituents with high confidence power.

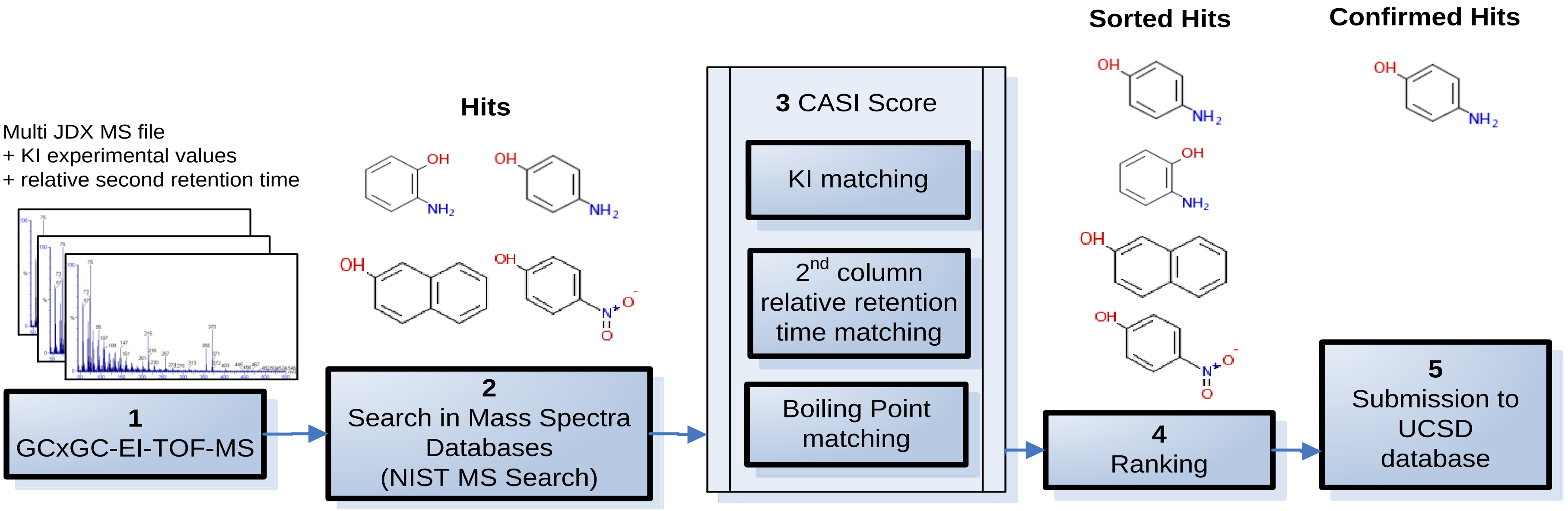
Smoke of a conventional cigarette, measured by GCxGC-EI-TOF-MS



Compound?

A. Knorr et al., Anal. Chem. 2013, submitted for publication

CASI Concept



Multi JDX MS file
+ KI experimental values
+ relative second retention time

1
GCxGC-EI-TOF-MS

Hits

2
Search in Mass Spectra Databases
(NIST MS Search)

3 CASI Score

KI matching

2nd column relative retention time matching

Boiling Point matching

Sorted Hits

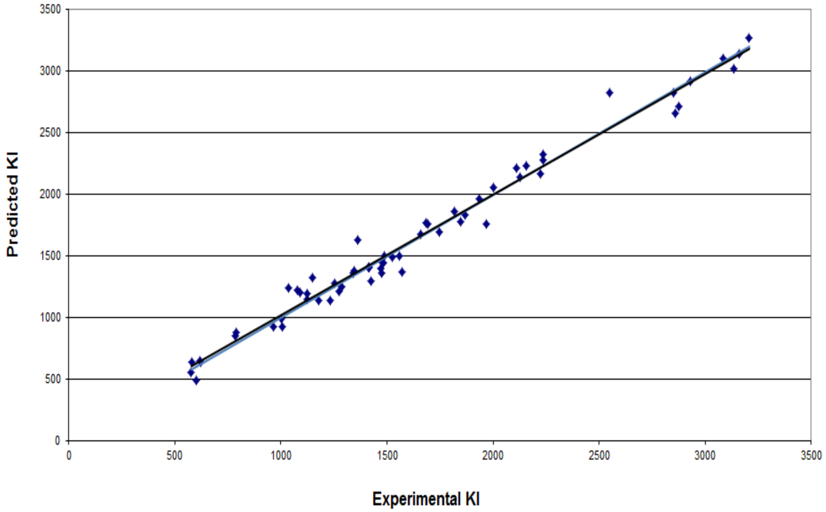
4
Ranking

Confirmed Hits

5
Submission to UCSD database

Predictive QSPR Models for KI and 2DrelRT

Kovats Index

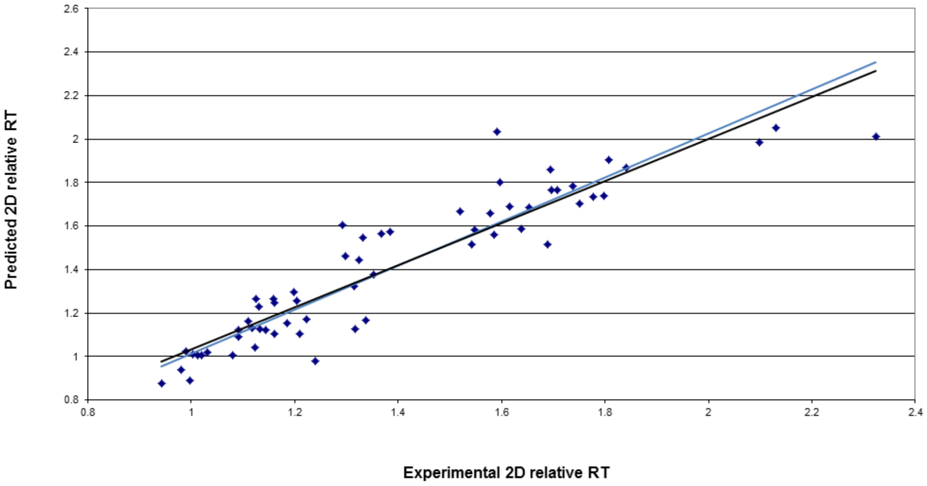


Validation n=60

$r^2 = 0.981$

GA – linear regression, 7 Molecular Descriptors

2DrelRT

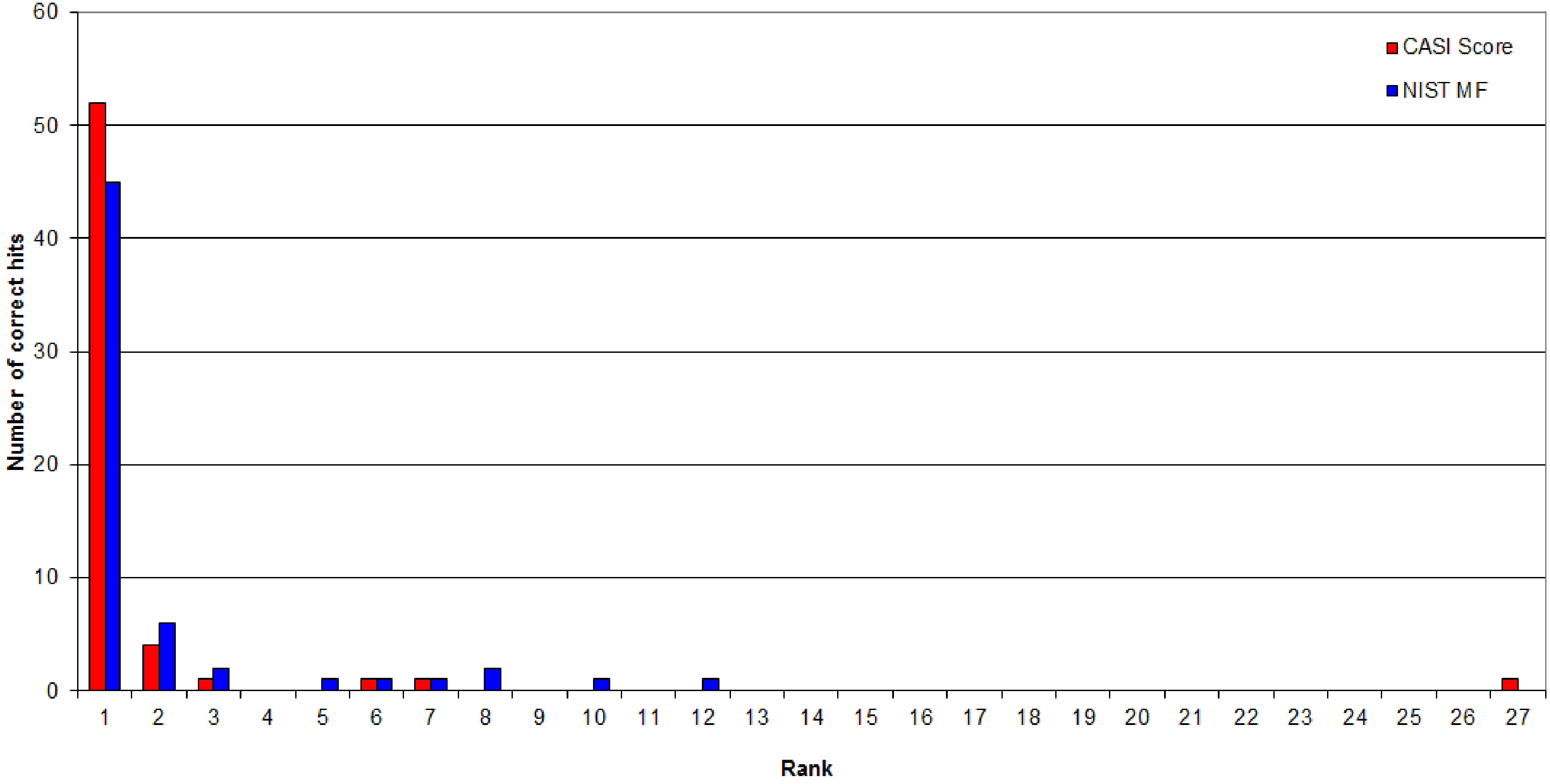


$r^2 = 0.855$

GA – Support Vector Regression, 12 molecular descriptors

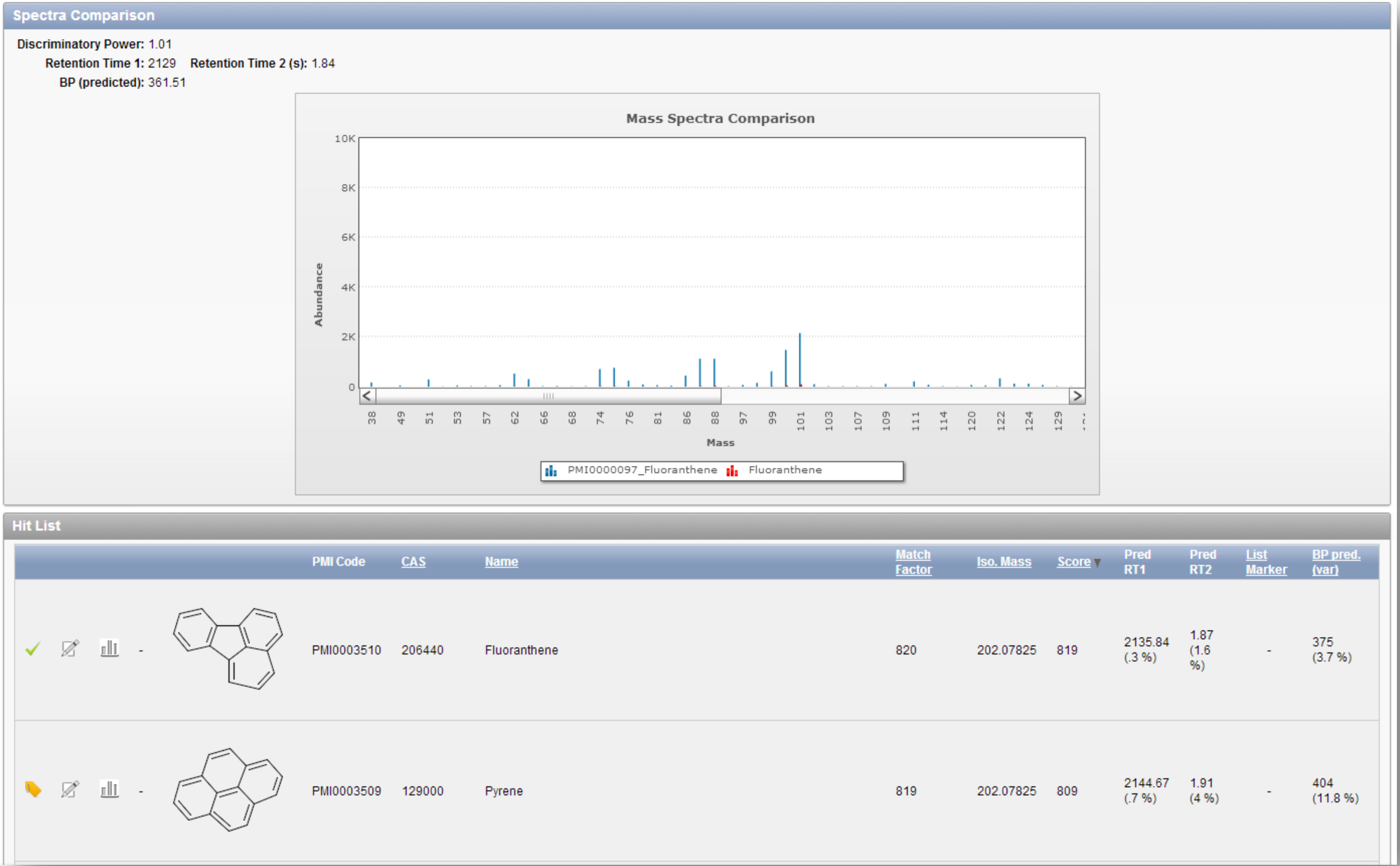
A. Monge & E. Martin, PMI R&D

Ranking of hits using CASI Score vs. NIST MS Match Factor



CASI Score improves NIST MS Search ranking.

CASI 2 Software



Spectra Comparison

Discriminatory Power: 1.01

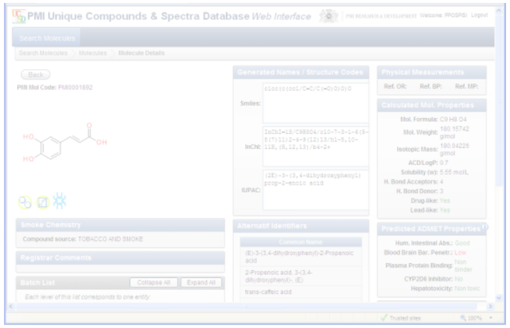
Retention Time 1: 2129 **Retention Time 2 (s):** 1.84

BP (predicted): 361.51

Hit List

					PMI Code	CAS	Name	Match Factor	Iso. Mass	Score ▼	Pred RT1	Pred RT2	List Marker	BP pred. (var)
✓			-		PMI0003510	206440	Fluoranthene	820	202.07825	819	2135.84 (.3 %)	1.87 (1.6 %)	-	375 (3.7 %)
			-		PMI0003509	129000	Pyrene	819	202.07825	809	2144.67 (.7 %)	1.91 (4 %)	-	404 (11.8 %)

3. Chemoinformatics Knowledge Base



PMI Unique Compounds & Spectra Database *Web Interface*

PMI RESEARCH & DEVELOPMENT Welcome: PPOSPISI Logout

Search Molecules

Search Molecules > Molecules > Molecule Details

Back

PMI Mol Code: PMI0001892

Generated Names / Structure Codes

Smiles: c1cc(c(cc1/C=C/C(=O)O)O)O

InChI: InChI=1S/C9H8O4/c10-7-3-1-6(5-8(7)11)2-4-9(12)13/h1-5,10-11H,(H,12,13)/b4-2+

IUPAC: (2E)-3-(3,4-dihydroxyphenyl)prop-2-enoic acid

Physical Measurements

Ref. OR: Ref. BP: Ref. MP:

Calculated Mol. Properties

Mol. Formula: C9 H8 O4

Mol. Weight: 180.15742 g/mol

Isotopic Mass: 180.04226 g/mol

ACD/LogP: 0.7

Solubility (w): 5.55 mol/L

H. Bond Acceptors: 4

H. Bond Donor: 3

Drug-like: Yes

Lead-like: Yes

Smoke Chemistry

Compound source: TOBACCO AND SMOKE

Registrar Comments

Batch List Collapse All Expand All

Each level of this list corresponds to one entity:

Alternatif Identifiers

Common Name
(E)-3-(3,4-dihydroxyphenyl)-2-Propenoic acid
2-Propenoic acid, 3-(3,4-dihydroxyphenyl)-, (E)
trans-caffeic acid

Predicted ADMET Properties

Hum. Intestinal Abs.: Good

Blood Brain Bar. Penetr.: Low

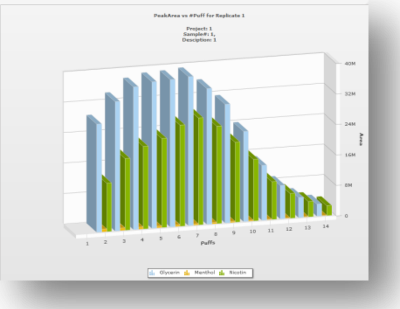
Plasma Protein Binding: Non binder

CYP2D6 Inhibitor: No

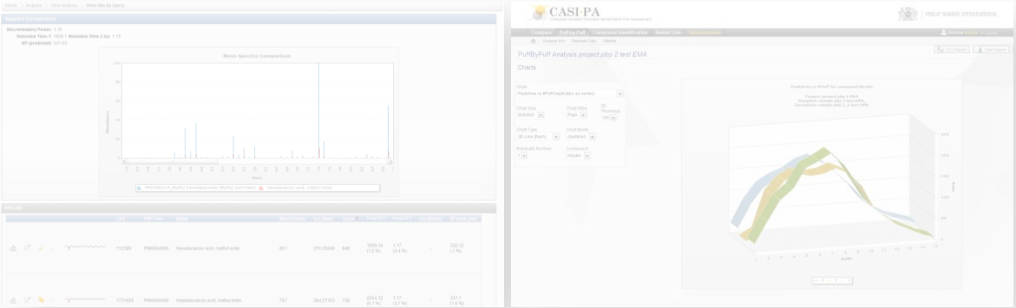
Hepatotoxicity: Non toxic

Trusted sites 100%

1. Unique Compounds & Spectra Database



3. Chemoinformatics Knowledge Base



2. Computer Assisted Structure Identification

Scientific Data Warehouse

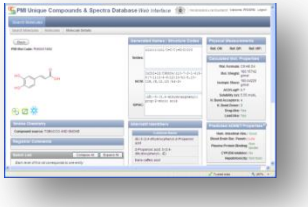
CIKB Scope

Central platform to **report** on the presence and concentration of constituents in smoke aerosols and their known or measured toxicities to support the toxicological assessment.

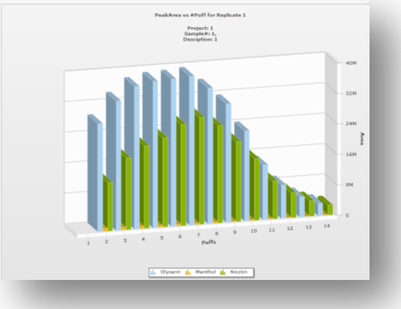
Information of interest to be combined with chemical – analytical information of UCSD and CASI platforms:

- **Product Lifecycle:** Complete specification of developed and tested products.
- **LIMS:** Quantitative smoke characterization.
- ***In-vitro* Assays:** Standardized *in vitro* assays to assess toxicity.
- **Toxicities:** Known toxicological information from literature.
- **Metabolites:** Known metabolites from literature.

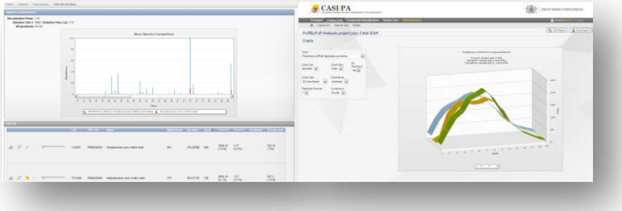
Chemistry Based Reporting



1. Unique Compounds & Spectra Database



3. Chemoinformatics Knowledge Base



2. Computer Assisted Structure Identification

Scientific Data Warehouse

Product Lifecycle

LIMS

Assays

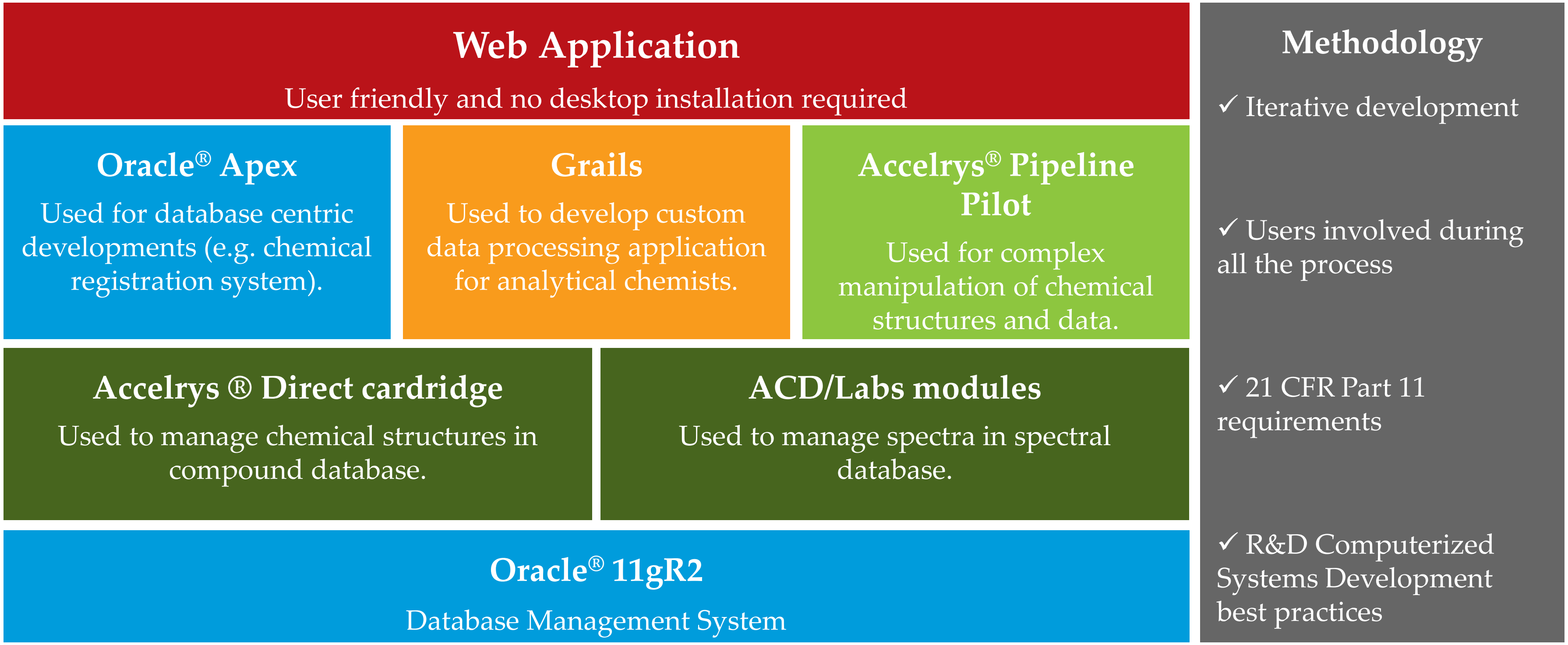
Toxicities

Metabolites

PMI RESEARCH & DEVELOPMENT

Used tools

- Rapid Application Development tools were used to develop software to support the processing of analytical chemistry data (mass spectrometry) and thus to answer quickly to business needs.



Web Application

User friendly and no desktop installation required

Oracle® Apex

Used for database centric developments (e.g. chemical registration system).

Grails

Used to develop custom data processing application for analytical chemists.

Accelrys® Pipeline Pilot

Used for complex manipulation of chemical structures and data.

Accelrys ® Direct cardridge

Used to manage chemical structures in compound database.

ACD/Labs modules

Used to manage spectra in spectral database.

Oracle® 11gR2

Database Management System

Methodology

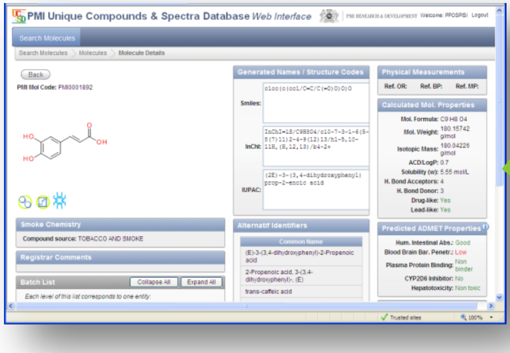
✓ Iterative development

✓ Users involved during all the process

✓ 21 CFR Part 11 requirements

✓ R&D Computerized Systems Development best practices

Conclusion



PMI Unique Compounds & Spectra Database *Web Interface*

PMI RESEARCH & DEVELOPMENT
Welcome: PPOSPISI
Logout
Search Molecules
Search Molecules › Molecules › Molecule Details
Back
HO
HO
O
OH
Trusted sites
100%

PMI Mol Code: PMI0001892

Generated Names / Structure Codes

Smiles: c1cc(c(cc1/C=C/C(=O)O)O)O

InChI: InChI=1S/C9H8O4/c10-7-3-1-6(5-8(7)11)2-4-9(12)13/h1-5,10-11H,(H,12,13)/b4-2+

IUPAC: (2E)-3-(3,4-dihydroxyphenyl)prop-2-enoic acid

Physical Measurements

Ref. OR: Ref. BP: Ref. MP:

Calculated Mol. Properties

Mol. Formula: C9 H8 O4
Mol. Weight: 180.15742 g/mol
Isotopic Mass: 180.04226 g/mol
ACD/LogP: 0.7
Solubility (w): 5.55 mol/L
H. Bond Acceptors: 4
H. Bond Donor: 3
Drug-like: Yes
Lead-like: Yes

Smoke Chemistry

Compound source: TOBACCO AND SMOKE

Registrar Comments

Batch List Collapse All Expand All

Each level of this list corresponds to one entity.

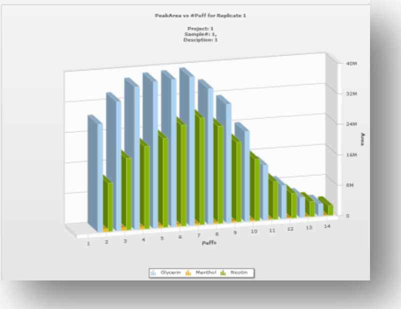
Alternatif Identifiers

Common Name
(E)-3-(3,4-dihydroxyphenyl)-2-Propenoic acid
2-Propenoic acid, 3-(3,4-dihydroxyphenyl)-, (E)
trans-caffeic acid

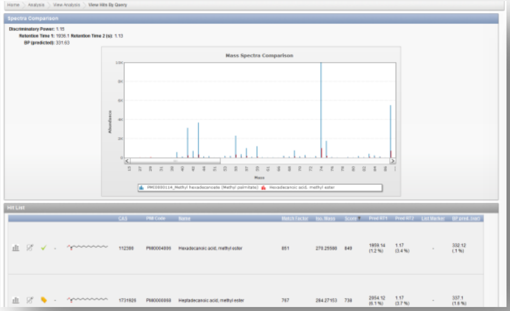
Predicted ADMET Properties

Hum. Intestinal Abs.: Good
Blood Brain Bar. Penetr.: Low
Plasma Protein Binding: Non binder
CYP2D6 Inhibitor: No
Hepatotoxicity: Non toxic

UCSD developed to facilitate and automatize processes of the unambiguous registration of compounds and their analytical spectra



CIKB in development to report on product chemical constituents and the associated toxicity and metabolic data



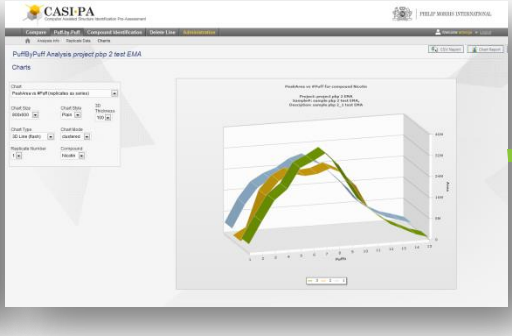
Home › Analysis › View Analysis › View Hits By Query
Mass Spectra Comparison
10K
8K
6K
4K
2K
0
Abundance
Mass
PMI0000114_Methyl hexadecanoate (Methyl palmitate)
Hexadecanoic acid, methyl ester

Spectra Comparison

Discriminatory Power: 1.15
Retention Time 1: 1936.1 Retention Time 2 (s): 1.13
BP (predicted): 331.63

Hit List

CAS	PMI Code	Name	Match Factor	Iso. Mass	Score	Pred RT1	Pred RT2	List Marker	BP pred. (var)
112390	PMI0004095	Hexadecanoic acid, methyl ester	851	270.25588	849	1959.14 (1.2 %)	1.17 (3.4 %)	-	332.12 (.1 %)
1731926	PMI0000068	Heptadecanoic acid, methyl ester	767	284.27153	738	2054.12 (6.1 %)	1.17 (3.7 %)	-	337.1 (1.6 %)



CASI·PA

Computer Assisted Structure Identification Pre-Assessment

PuffByPuff Analysis *project pbp 2 test EMA*

Charts

Chart: PeakArea vs #Puff (replicates as series)
Chart Size: 800x800
Chart Style: Plain
3D Thickness: 100
Chart Type: 3D Line (flash)
Chart Mode: clustered
Replicate Number: 1
Compound: Nicotin

CASI platforms developed to increase the confidence in the identification of compounds using GC-MS methods and facilitates to users the comparison of analyses

Scientific Data Warehouse

Acknowledgment

- *Stephane Cano, Stephane Albrecht, Pavel Pospisil*
- *PMI R&D chemistry analytics teams*
- *PMI R&D HPC team*
- *PMI R&D SIS team*
- *Accelrys Symyx consultants*
- *ACD/Labs consultants*

Thank you for your attention.