Retention Index Prediction Combined with *In Silico* Fragmentation Spectra Comparisons for Increasing Confidence in Structural Elucidation using Non-Targeted Gas Chromatography coupled with High Resolution Mass Spectrometry

"NonTarget 2016 Conference, Ascona, May, 30th 2016"

"P.A. Guy, E. Dossin, E. Martin, P. Diana, P. Pospisil, M. Bentley"

Philip Morris International R&D

Outline

- **Generation of aerosol sample / chemical complexity / GC-HR-MS analysis**
- **Building linear retention index (LRI) prediction models**
  - RapidMiner – Dragon software (RM)
  - ACD/ChromGenius software (CG)
  - LRI modeling assessment & usage to characterize aerosol constituents (library database)
- **Non-targeted screening workflow for aerosol characterization**
- **Case studies**
- **Conclusion and next steps**

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

PMI Science

- PMI is working on various Reduced Risk Products (RRP) delivering nicotine containing aerosols.

Heat-Not-Burn product

Tobacco Heating System (THS) 2.2

- In this context, it is important to fully characterize the chemical composition of RRP aerosols in comparison to smoke produced from cigarettes.
- For analytical method development purposes we use a reference cigarette (3R4F).

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Generation of Smoke Samples from a Reference Cigarette

2 cold impingers in series

- **Reference cigarette: 3R4F***
- **Smoking regimen: Health Canada**
  - 2 sticks accumulation
  - Puff volume: 55 mL
  - Puff duration: 2 sec
  - Frequency: 2 puffs / min
  - Puff count (butt length)

* University of Kentucky (Kentucky Tobacco R&D Center). http://www2.ca.uky.edu/refcig/

Cambridge filter is extracted

Gas Vapor Phase (GVP)

Total Particulate Matter (TPM) → **Whole smoke**

- **Cambridge filter is combined with the impingers** ⇨ **Whole smoke**
- **Addition of retention index chemical markers (n-alkanes) & isotopically labeled internal standards**

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Unique Compounds & Spectra Database (UCSD)

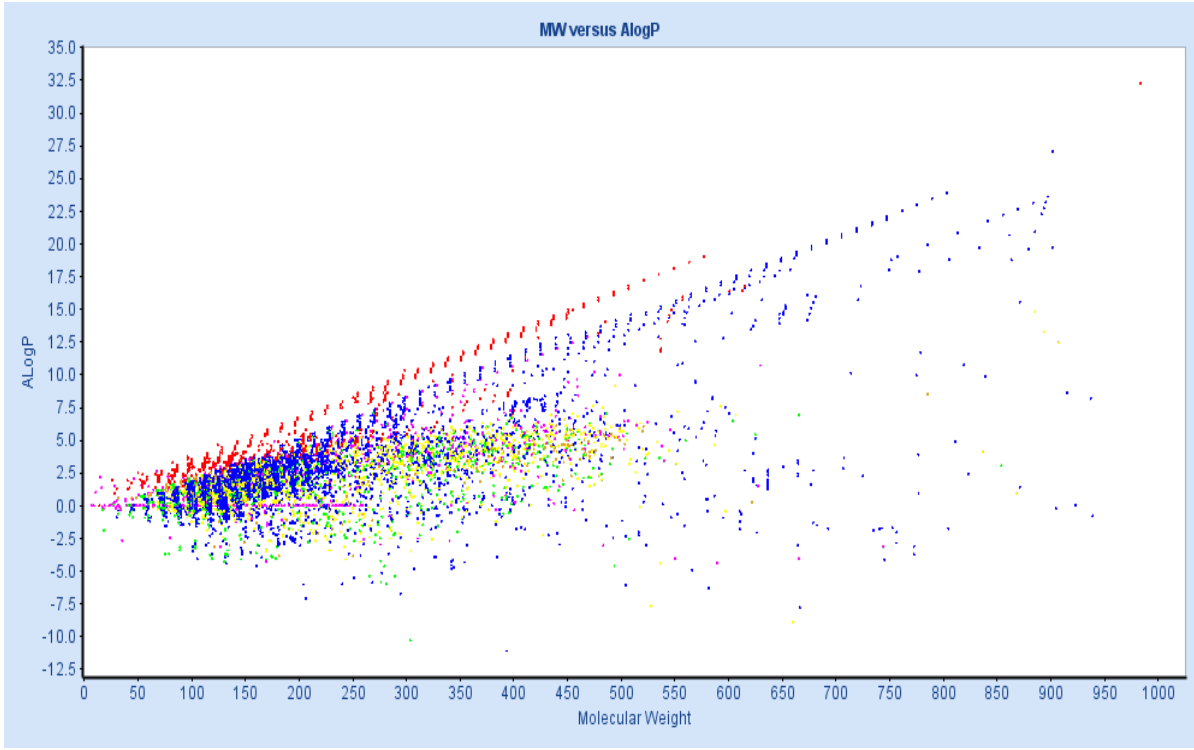
11,567 molecules are registered in our in-house database:

- ❑ Over 7,000 chemicals reported as present in tobacco and tobacco smoke[1]
- ❑ Over 3,000 molecules associated with flavor properties[2-3]

Martin, E. *et al*. **2012**. *J. Chemoinform.*, *4*, 1, 1-14.

1,013 (+EI) accurate mass spectra

- Hydrocarbon (n=1'081)
- Oxygen-containing functions (n=7'427)
- Nitrogen-containing functions (n=1'788)
- Nitrogen heterocyclic functions (n=2'715)
- Sulfur-containing functions (n=897)
- Miscellaneous functions (n=1195)



1 Rodgman, A.; Perfetti, T.A. The Chemical Components of Tobacco and Tobacco Smoke,, **2013**, *2nd Ed. CRC press*.
2 Leffingwell, J. C. *et al*. Tobacco flavoring for Smoking Products, R. J. Reynolds Tobacco Company, Winston-Salem, NC, **1972**.
3 EFSA flavoring substances database.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Analytical Technique: GC-High Resolution (GC-HR-MS)

GC-HR-MS_2
(7200B Agilent Q-TOF-MS)

Apolar and polar
From LRI of 1,000 to 3,000
(HP-5ms GC column)

GC-HR-MS_1
(7200A Agilent Q-TOF-MS)

Volatile and semi-volatiles
From LRI of 500 to 1,900
(**DB-624 GC column**)

Goal is to screen the broadest range of smoke constituents in a "non-targeted screening" approach.

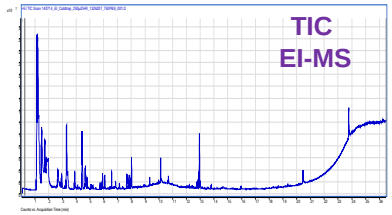
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Building Linear Retention Index Models using QSPR

Model 1
Structural descriptors

rapidminer

Experimental LRI data
(n=552 ref. stds)
DB-624 GC column



Training set
Reference chemicals
(n=401, 2/3)

Rapid Miner software

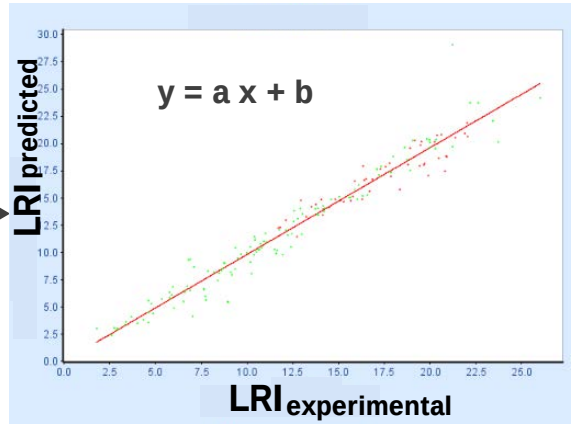
Model optimization

ACD/ChromGenius software

ACD/Labs

Model 2
Compound similarities

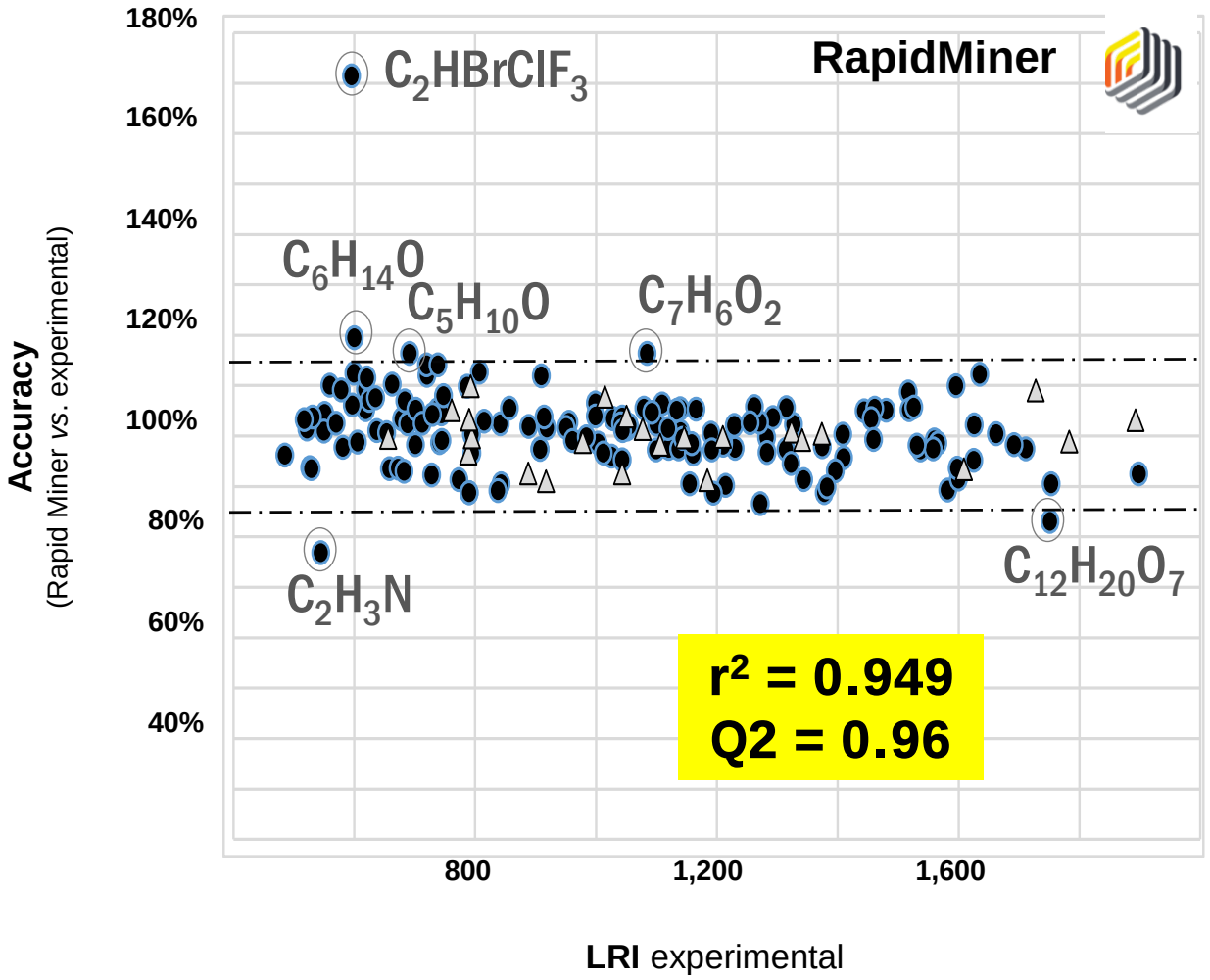
Test set
Reference chemicals
(n=151, 1/3)

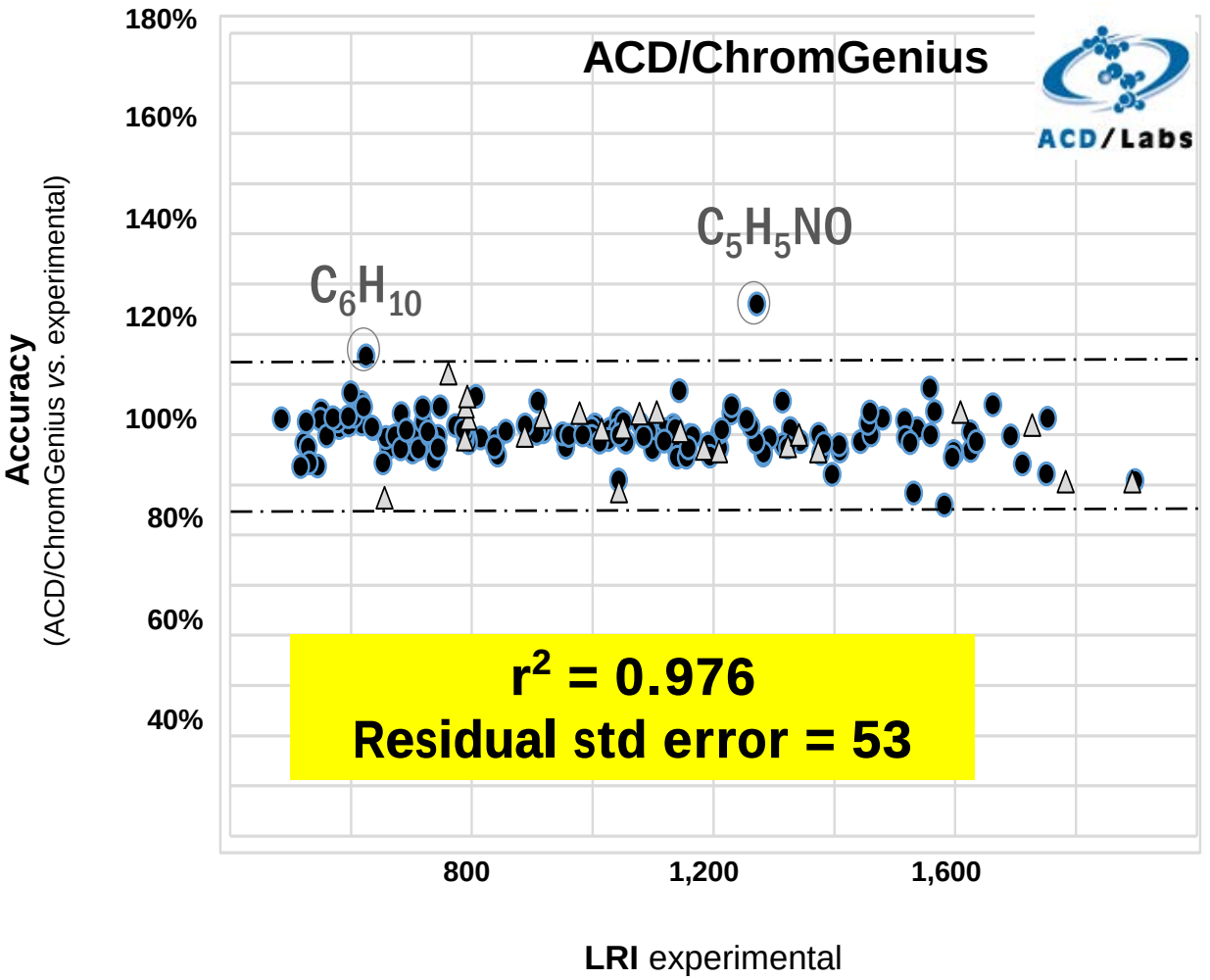


Model Assessment

Validation set
Reference chemicals (n=23)

Accuracy Data for Predicted versus Experimental LRI Values



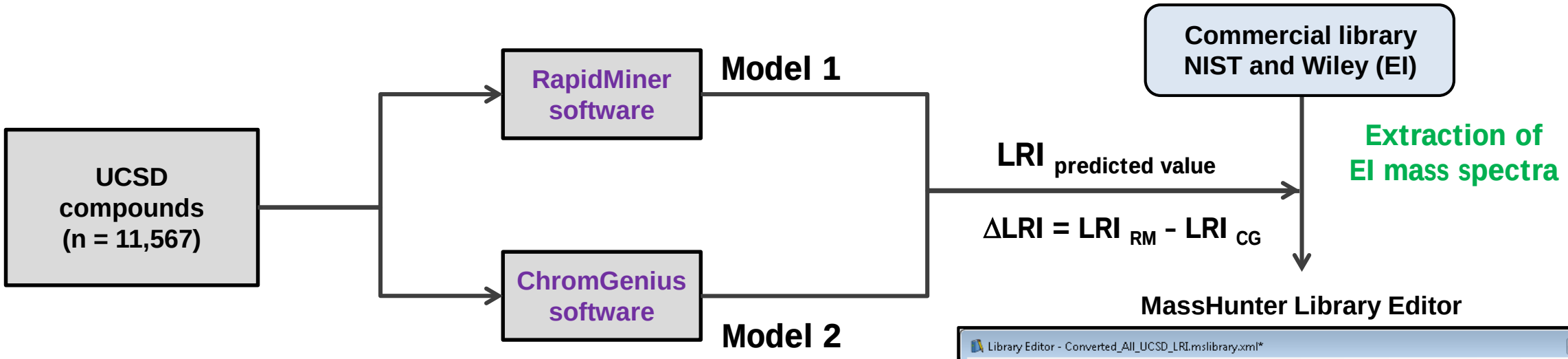


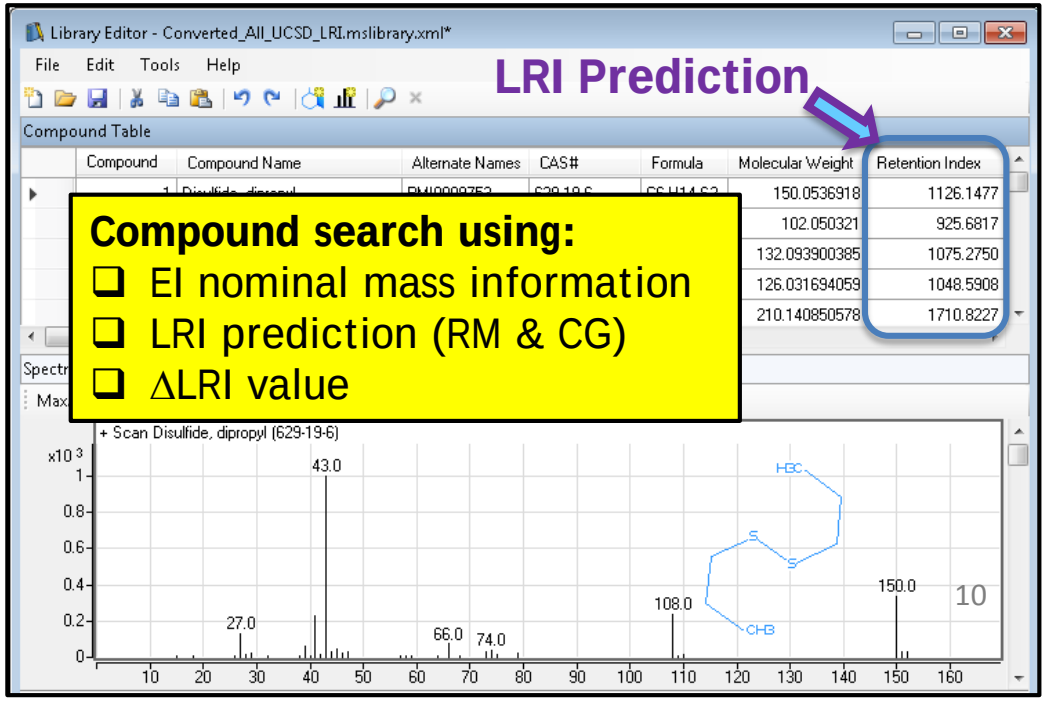
- n=151 reference standards (Test set)
- △ n=23 reference standards (Validation set)

Submitted in peer-reviewed Journal

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

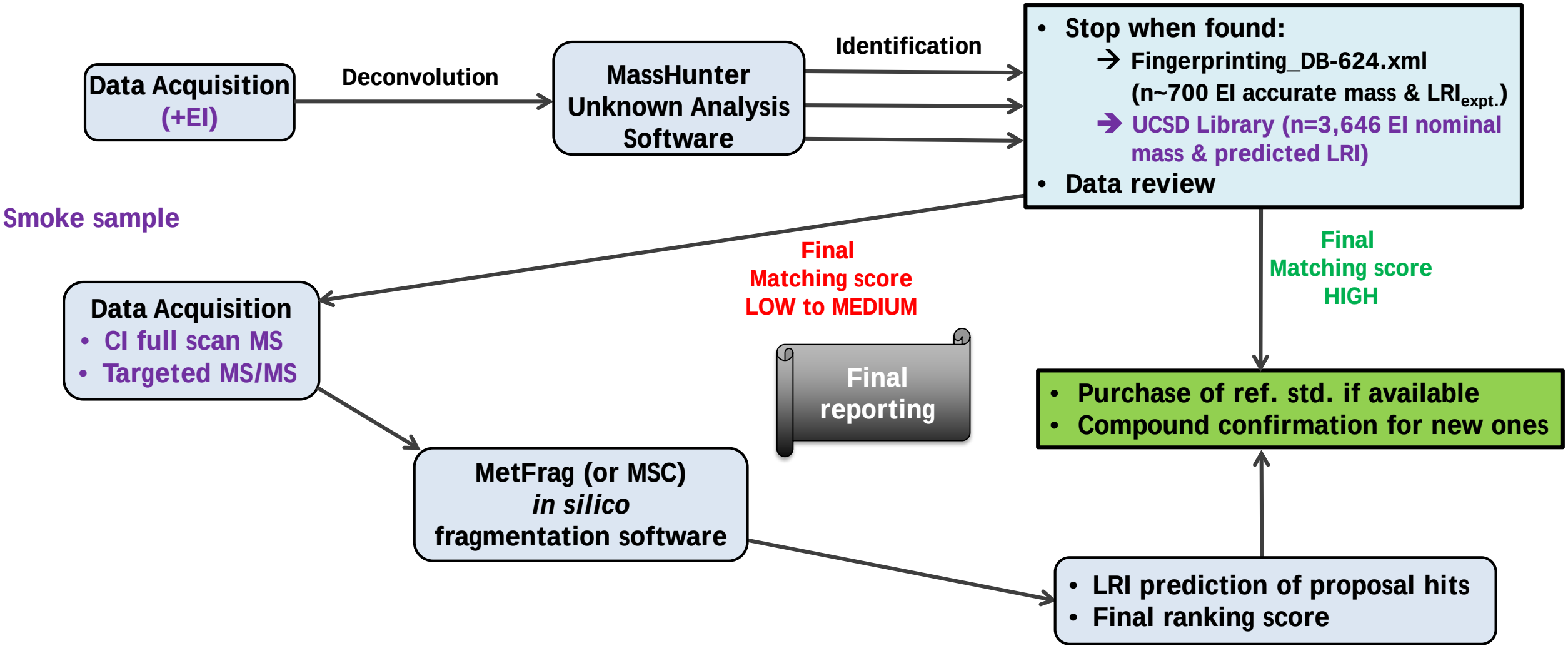
LRI Prediction for the Complete UCSD Compound Library



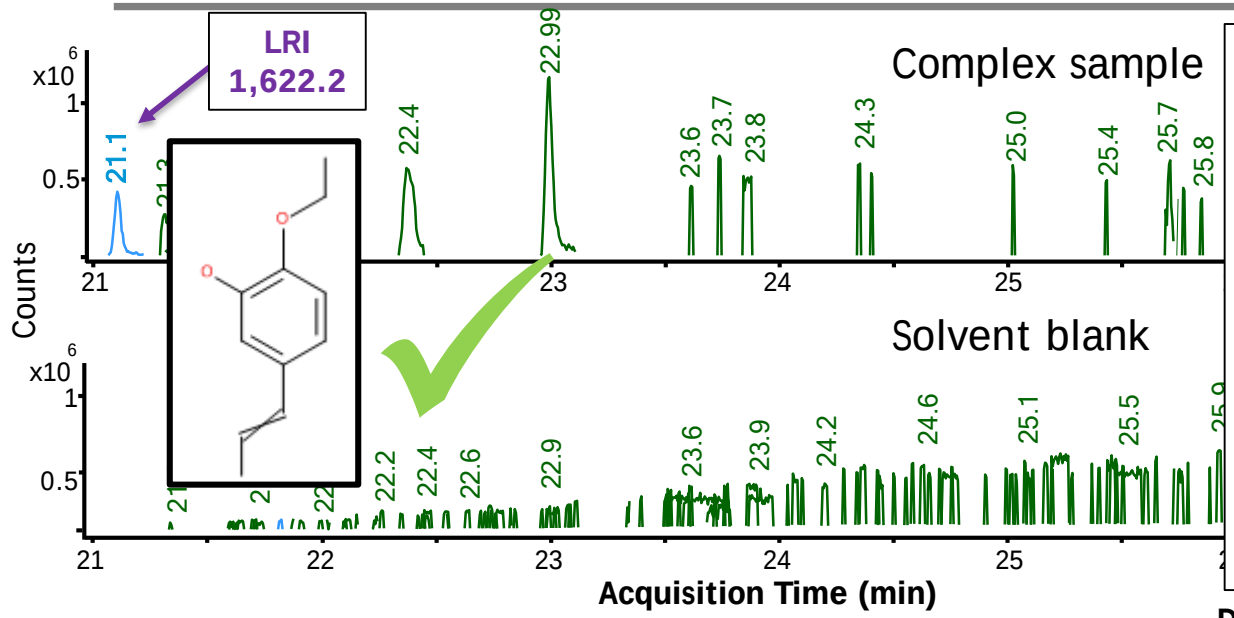


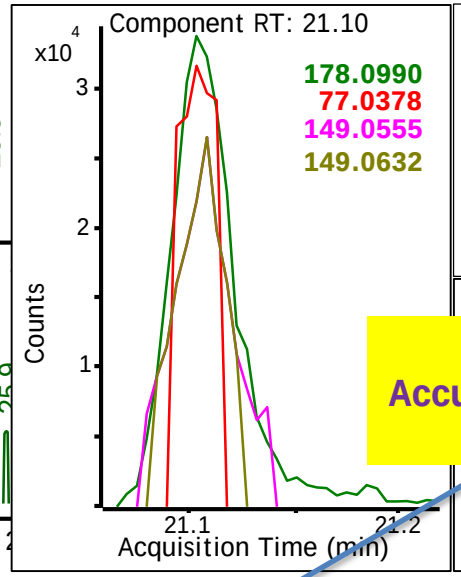
- 6,053 molecules were predicted with LRI values between 500 - 1,900 (targeted for DB-624 GC column)
- 3,646 molecules (60%) have an EI Mass Spectra (NIST or Wiley)
- LRI values can be predicted from any compound databases

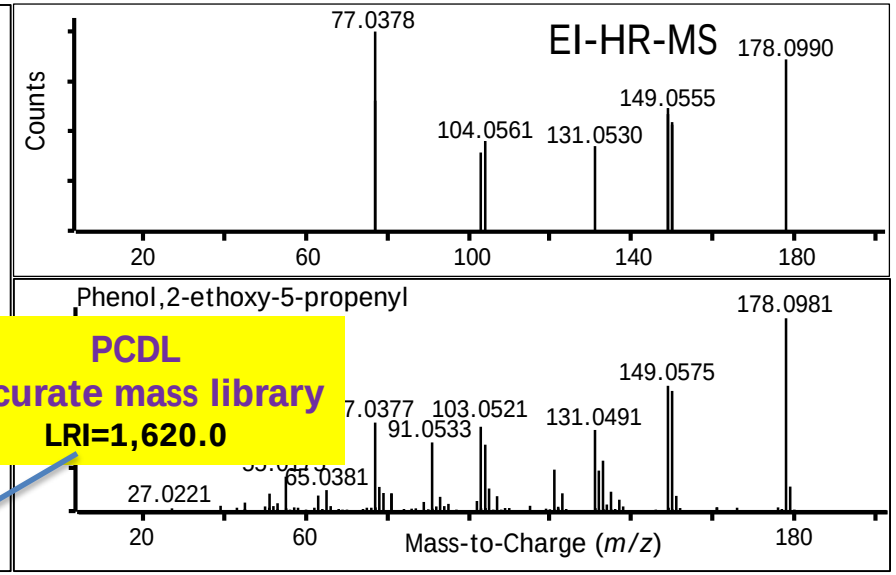
Non-targeted Screening Workflow for Aerosol Characterization



Case Study 1:
Compound Identification with Accurate Mass Library







PCDL
Accurate mass library
LRI=1,620.0

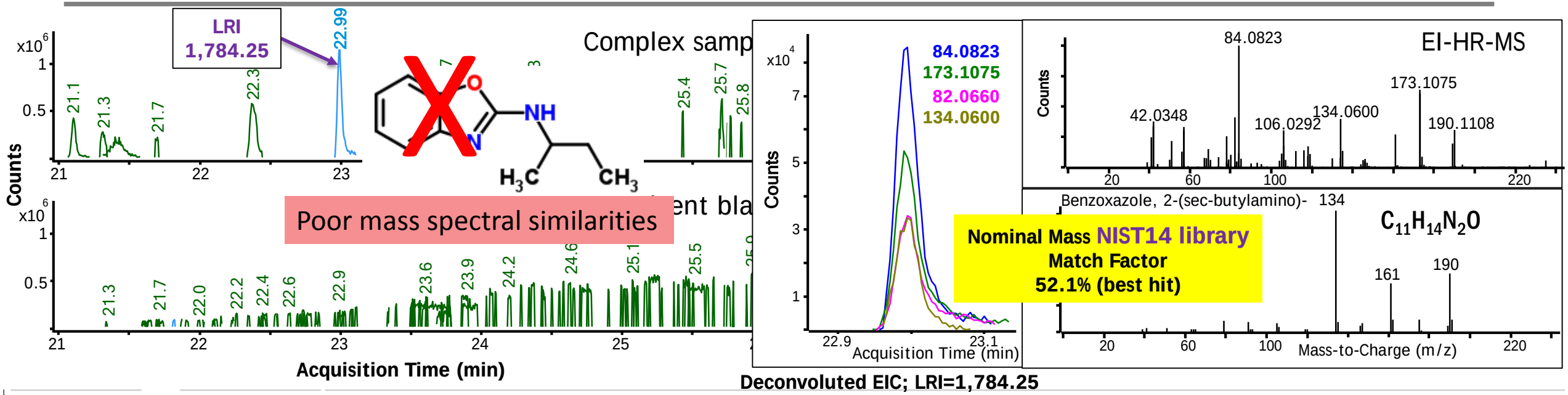
Deconvoluted EIC: LRI=1,622.2

	A	B	C	D	E	F	G	H	I	J	K	L
1	Compound Name	Formula	CAS#	Component RI	Library RI	Delta RI	Match Factor	Library File	Component Area	Component Height	Library RT	Delta RT
109	Phenol,2-ethoxy-5-propenyl	C11H14O2	94-86-0	1622.23	1620.00	-2.23	67.1	Fingerprinting_DB624.xml	568891	33972	21.08	-0.03
110	Isothiocyanic acid, s-phenenyl ester	C9H3N3S3	101670-67-1	1686.47			63.9	NIST14.L	15200	27006		
111	1,3-Dimethyl adamantane-1,3-dicarboxylate	C14H20O4	1000411-47-1	1663.89			54.6	NIST14.L	97339	29649		
112	Dimethylmalonyl chloride	C5H6Cl2O2	5659-93-8	1698.49			64.2	NIST14.L	6736	4121		
113	Piperonal	C8H6O3	120-57-0	1728.77			79.9	NIST14.L	302825	52401		
114	Benzoxazole, 2-(sec-butylamino)-	C11H14N2O	28291-82-9	1784.25			52.1	NIST14.L	1299540	97476		

Easy compound confirmation if reference standard is already present within our Personal Compound Database accurate mass Library (PCDL, n~700)

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Case Study 2:
Problematic Hit Proposals



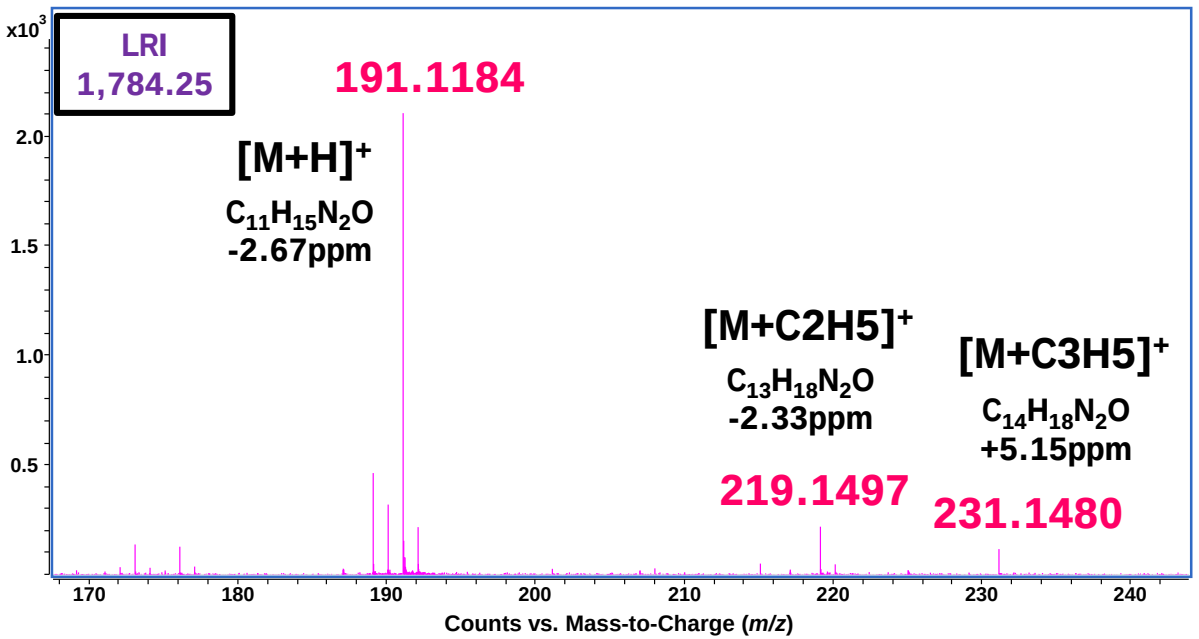
	A	B	C	D	E	F	G	H	I	J	K	L
1	Compound Name	Formula	CAS#	Component RI	Library RI	Delta RI	Match Factor	Library File	Component Area	Component Height	Library RT	Delta RT
109	Phenol,2-ethoxy-5-propenyl	C11H14O2	94-86-0	1622.23	1620.00	-2.23	67.1	Fingerprinting_DB624.xml	568891	33972	21.08	-0.03
110	Isothiocyanic acid, s-phenenyl ester	C9H3N3S3	101670-67-1	1686.47			63.9	NIST14.L	15200	27006		
111	1,3-Dimethyl adamantane-1,3-dicarboxylate	C14H20O4	1000411-47-1	1663.89			54.6	NIST14.L	97339	29649		
112	Dimethylmalonyl chloride	C5H6Cl2O2	5659-93-8	1698.49			64.2	NIST14.L	6736	4121		
113	Piperonal	C8H6O3	120-57-0	1728.77			79.9	NIST14.L	302825	52401		
114	Benzoxazole, 2-(sec-butylamino)-	C11H14N2O	28291-82-9	1784.25			52.1	NIST14.L	1299540	97476		

There is a need to develop alternative approaches when compounds are not registered in existing MS libraries

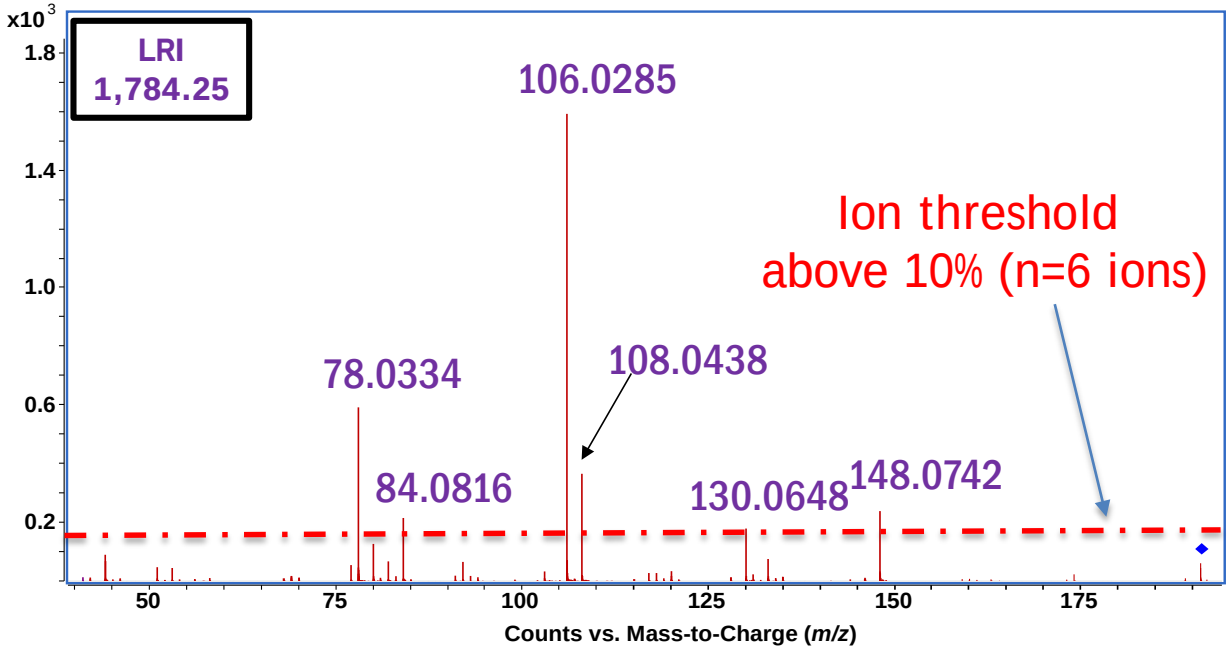
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Case Study 2:
GC-HR-MS in Chemical Ionization Mode & MS/MS

GC-HR-MS (Full Scan MS)
Positive Chemical Ionization (PCI)



GC-HR-MS (Full Scan MS/MS)
PCI data acquisition CID of 191.1184



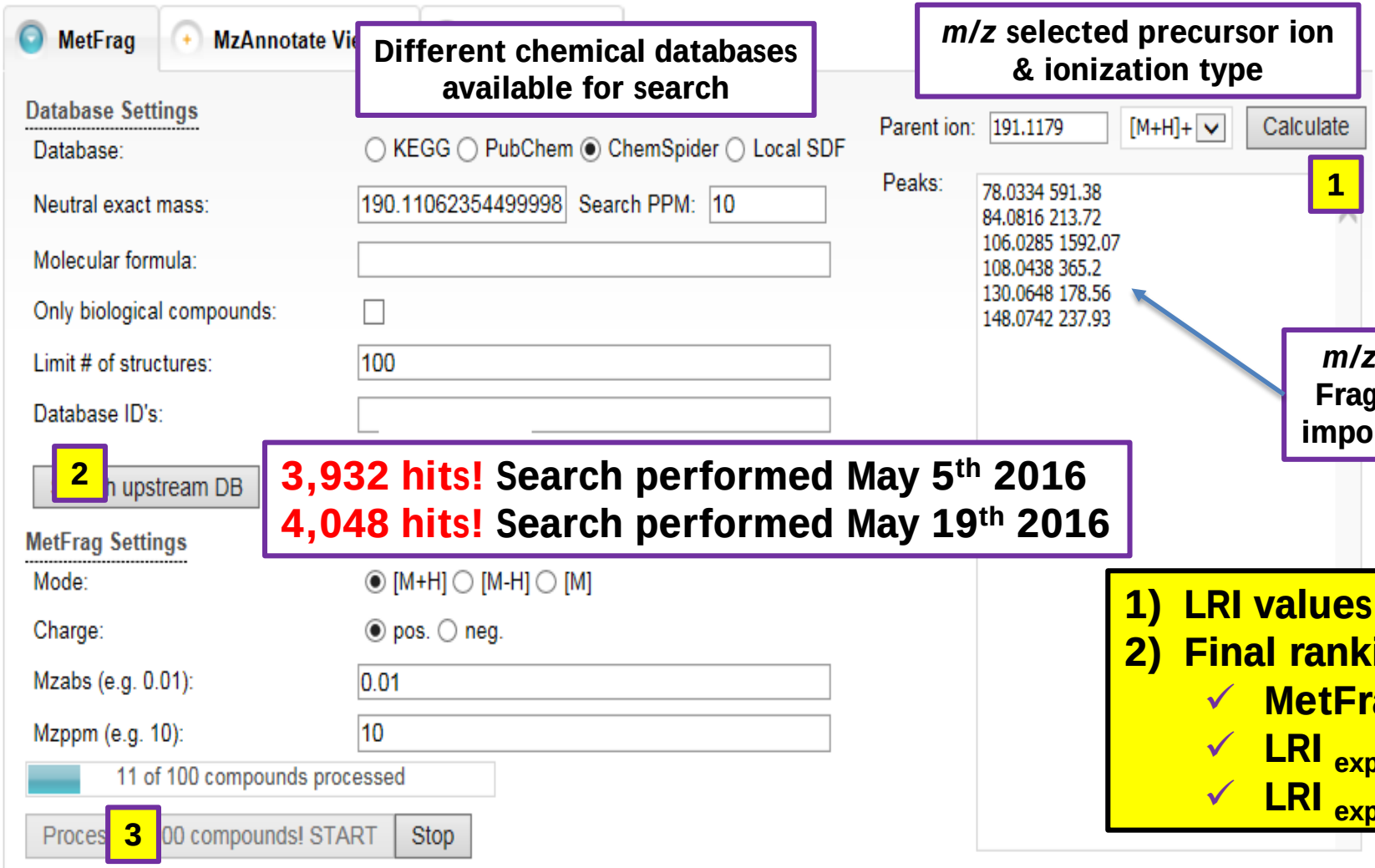
Determination of elemental formula (adduct ion species)

M: $C_{11}H_{14}N_2O$

MS/MS data processed using a larger chemical database with *in silico* predicted fragmentation software

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

In Silico Theoretical Fragmentation Software Evaluation: MetFrag



Different chemical databases available for search

***m/z* selected precursor ion & ionization type**

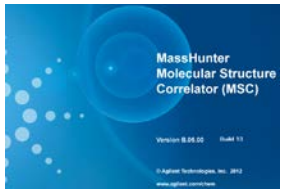
***m/z* & counts Fragment ions imported values**

3,932 hits! Search performed May 5th 2016
4,048 hits! Search performed May 19th 2016

1) **LRI values were predicted for all 100 proposals**
2) **Final ranking SCORE was calculated using:**
   - ✓ **MetFrag Score**
   - ✓ **$LRI_{expt.}$ Against LRI_{RM}**
   - ✓ **$LRI_{expt.}$ Against LRI_{CG} …**

Wolf, S. *et al*. *BMC Bioinformatics* **2010**, *11*, 148. http://msbi.ipb-halle.de/MetFrag/

In Silico Theoretical Fragmentation Software Evaluation: Molecular Structure Correlator (MSC)



1) MS/MS accurate mass spectra exported as .cef files
2) Open in MSC software
3) Several databases are available

- Elucidation of Product Ion Connectivity (EPIC) based-approach
- Systematic bond cleavages with a score penalty function

Different chemical databases available for search

Calculated Elemental formula

Fragment ions interpretation

***m/z* fragment ions & intensities imported values**

List of putative compounds

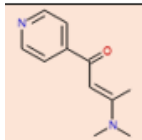
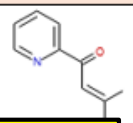
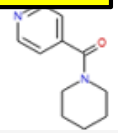
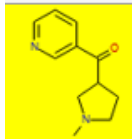
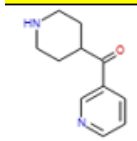
True compound was ranked in 43rd position

Hill, A.W. & Mortishire-Smith, R.J. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3111.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

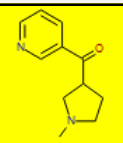
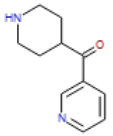
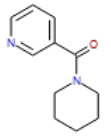
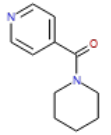
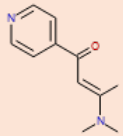
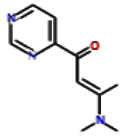
Assessment for MetFrag *In Silico* Fragmentation

Alone

PNG_Image	Comment	ChemSpider ID	Mass	MetFrag Score	Rank
	unspecified stereochem.	1221410 (2z & 2E) 1221411 (2Z form) 4603758 (2E form)	190.1106	1.0000	1st
	unspecified stereochem.	2045246	190.1106	1.0000	2nd
		1259330	190.1106	0.9860	3rd
		1256481	190.1106	0.9860	4th
		3716473	190.1106	0.9840	5th
		963178	190.1106	0.9840	6th

5th proposal confirmed (ref. standard)

+ LRI Prediction RM & CG

PNG_Image	LRI_pred CG	LRI_pred RM	LRI_exp	MetFrag & LRI_pred. SCORE	Rank
	1701.97 ΔLRI=-82.3	1763.39 ΔLRI=-20.9	1'784	0.930	1st
	1793.5298 ΔLRI=+9.3	1898.80 ΔLRI=+114.55	1'784	0.920	2nd
	1811.87 ΔLRI=+27.6	1891.95 ΔLRI=+107.7	1'784	0.916	3rd
	1820.33 ΔLRI=+36.1	1893.98 ΔLRI=+109.7	1'784	0.910	4
	1637.96 ΔLRI=-146.3	1702.82 ΔLRI=-81.4	1'784	0.884	5
	1634.80 ΔLRI=-149.5	1699.52 ΔLRI=-84.7	1'784	0.881	6

1st proposal confirmed (ref. standard)

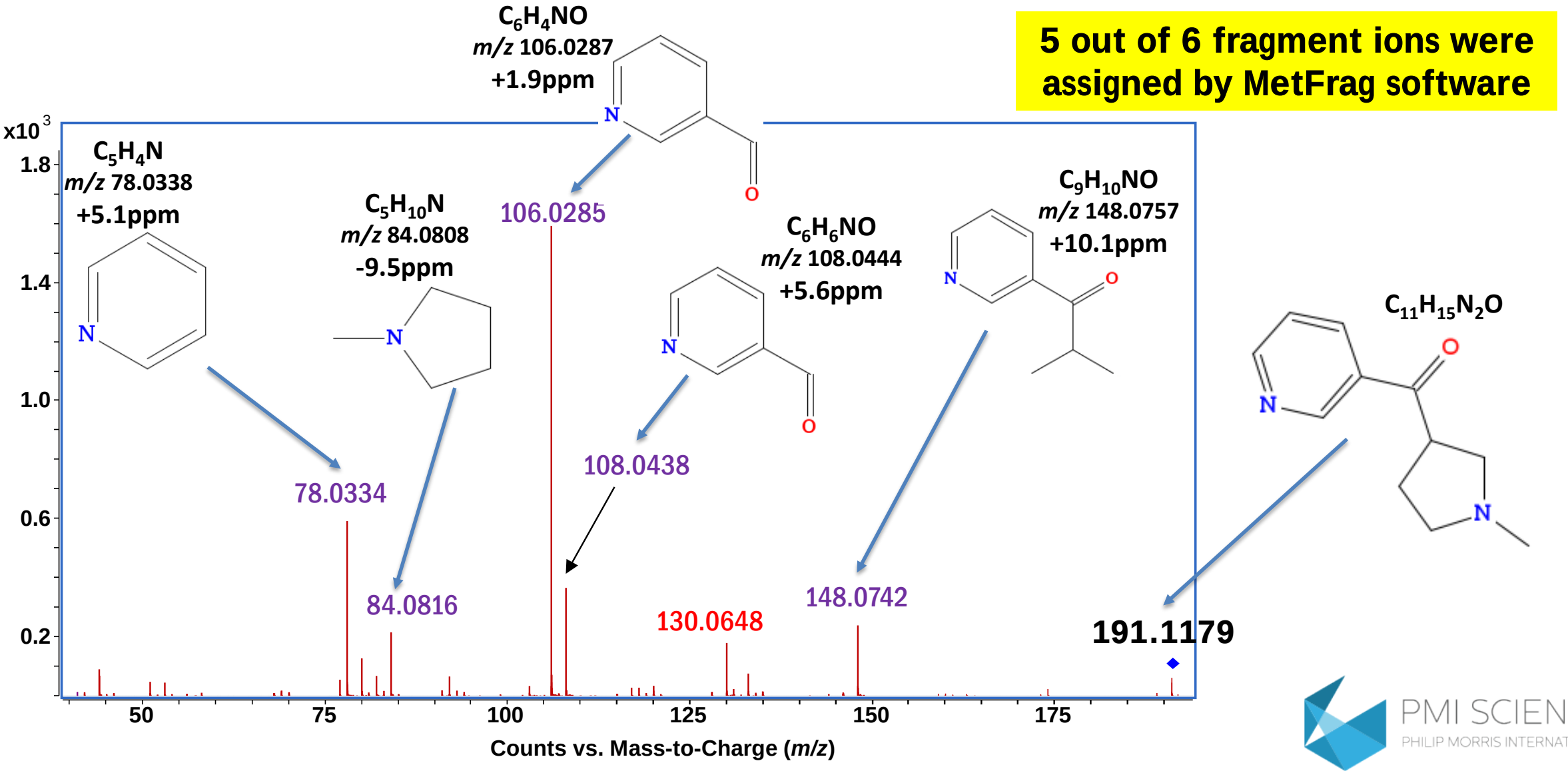
Better discriminatory power

Usefulness to combine LRI prediction with MetFrag score

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Interpretation of (1-Methyl-3-pyrrolidinyl)(3-pyridinyl)methanone MS/MS Spectrum Using MetFrag Software

5 out of 6 fragment ions were assigned by MetFrag software



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

MetFrag *vs*. Molecular Structure Correlator Software

TRUE COMPOUND	(R,S)-1-methyl-3-nicotinoylpyrrolidine	2,3-pentanedione	2-pentanone	3-penten-2-one
Formula	$C_{11}H_{14}N_2O$	$C_5H_8O_2$	$C_5H_{10}O$	C_5H_8O
RANKING NIST14 nominal classical search	not registered	Not present in hit list	**1st**	Not present in hit list
RANKING NIST14 with formula constraint	-	2nd	**1st**	Not present in hit list
# Cpds NIST14	**38**	**50**	**55**	**34**
# Cpds ChemSpider	**3,651**	**243**	**125**	**120**
# of Fragment ions (above 10%)	6	3	4	7
RANKING MetFrag	5th ranking	15th ranking	17th ranking	12th ranking
RANKING MSC	43th ranking	34th ranking	6th ranking	15th ranking
LRI expt	1'783	738	730	792
LRI (RM)	1763 (ΔLRI=-20)	842 (ΔLRI=+104)	714 (ΔLRI=-16)	746 (ΔLRI=-46)
LRI (CG)	1702 (ΔLRI=-81)	771 (ΔLRI=+33)	732 (ΔLRI=+2)	770 (ΔLRI=-22)
RANKING MetFrag & LRI pred.	**1st**	**7th**	**3rd**	**4th**

+EI

CI
Full scan
MS/MS

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Conclusions & Next Steps

- **Advantageous to combine state-of-the-art instrumentation with advanced chemoinformatic tools**
  - LRI prediction models using both RM & CG software (algorithms) showed great results
  - Low differences between the two LRI models enhanced the confidence level for compound identification
- **Existing MS libraries are not exhaustive and additional strategies need to be developed**
- **Targeted MS/MS combined with software to predict *in silico* fragmentation is mature**
  - MetFrag software seems to be more reliable than Molecular Structure Correlator
  - Addition of LRI prediction values demonstrated a greater potential to correctly rank putative hits than *in silico* fragmentation alone

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Conclusions & Next Steps (continued)

- **This combined approach significantly reduces the amount of compounds purchased for absolute confirmation**
  - Reducing the overall time for compound identification
  - Reducing the cost for purchasing chemicals
  - Minimizing the rate of false positive compound identification
- **Complete automated data-processing has to be developed and validated in order to reduce the workload for Non-Targeted Screening applications**
  - Final Ranking SCORE to be calculated on the fly (accurate mass results – LRI predictions)
  - Data fusion across volatile – semi-volatile & polar – apolar methods

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Acknowledgments

Agilent Technologies

- Joerg Riener
- Tomi Hamalainen

Philip Morris International R&D*

- Complex Matrix Analysis (M. Bentley)
- Fingerprinting & Special Analysis Team
  - E. Dossin
  - P. Diana
- Computational Chemistry Team (P. Pospisil)
  - E. Martin
  - A. Castellon
- Aerosol generation staff (R. Reis Pires)

* Philip Morris Products S.A. (part of Philip Morris International group of companies)

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL