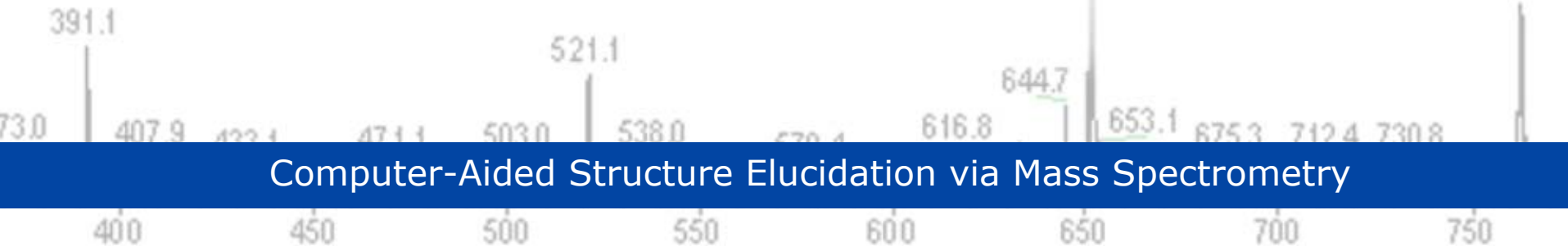
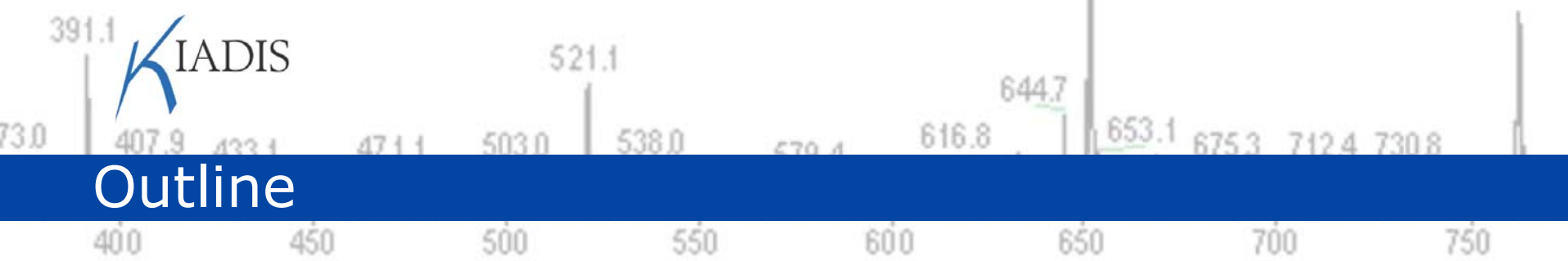
Markus Meringer

Generation of Molecular Structures and Structure Ranking by Mass Spectra



Computer-Aided Structure Elucidation via Mass Spectrometry

MATH/CHEM/COMP 2005
Dubrovnik
June 22, 2005



Outline

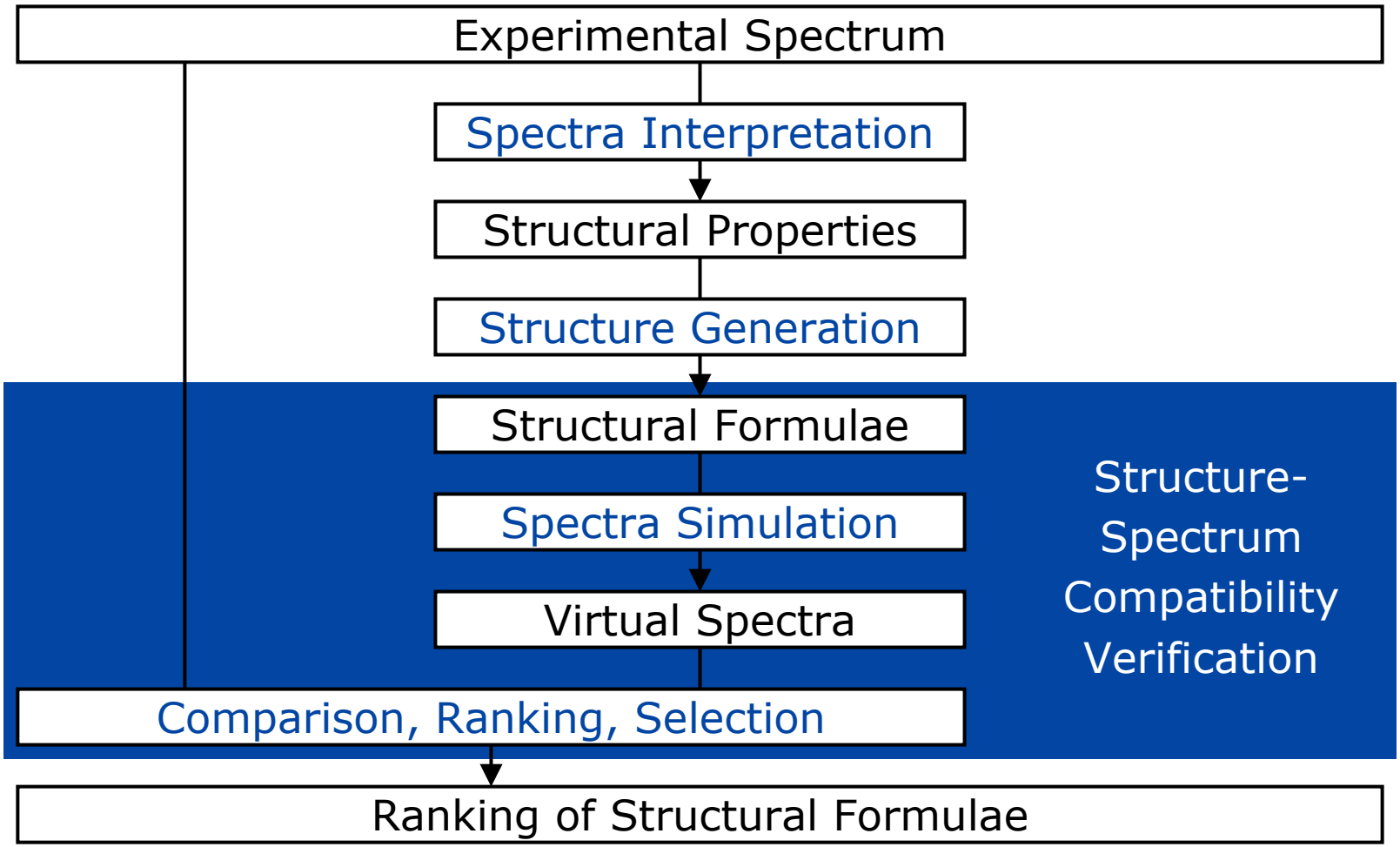
- Introduction
- Methods
  - Exhaustive structure generation
  - Virtual MS fragmentation
  - Structure-spectrum compatibility matchvalue
- Computational experiment
  - Random sample of 100 structure spectra pairs
- Outlook

Introduction: Aspects of CASE

- Various spectroscopic methods
  - NMR, IR, UV, AAS, AES, MS
  - This study: Low resolution electron impact MS
- Two approaches towards CASE
  - Database searching
  - De-novo structure elucidation
- Three steps in de-novo structure elucidation
  - Spectra interpretation
  - Structure generation
  - Structure-spectum compatibility verification

KIADIS

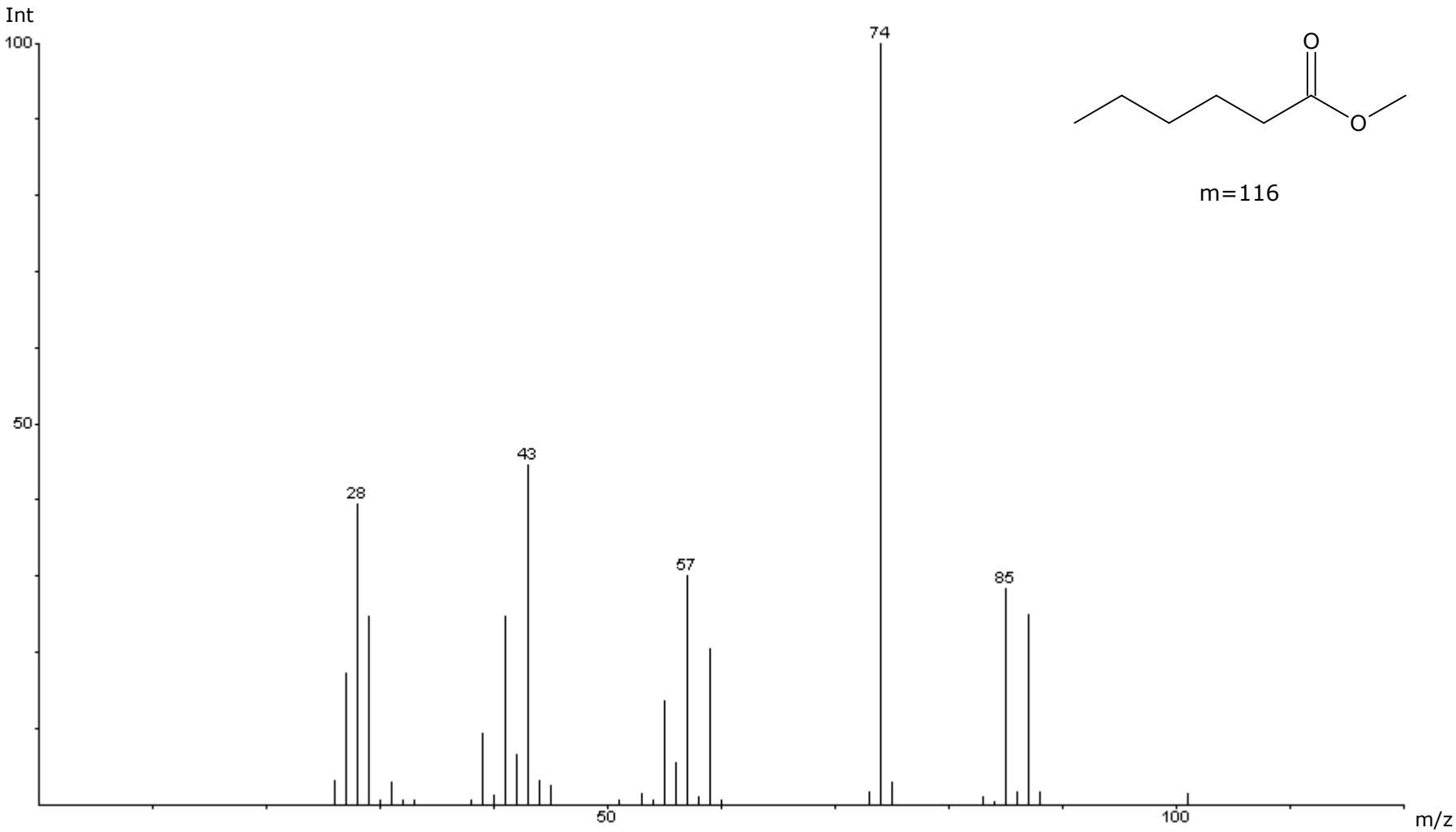
Workflow of De-Novo Structure Elucidation



DeNovo Structure Elucidation via MS

- Why MS?
  - High sensibility
  - Good selectivity
- Allows determination of
  - Molecular mass (soft ionization MS)
  - Molecular formula (high resolution MS)
  - Structural formula (characteristic fragmentation)
- Difficulty in structure determination by MS
  - Not the unknown compound itself is measured, but rather a mixture of fragment ions
  - Concentrations of fragment ions and intensities of peaks are almost uncomputable

Example: EI-MS of an ‚Unknown' $C_6H_{12}O_2$ Isomer



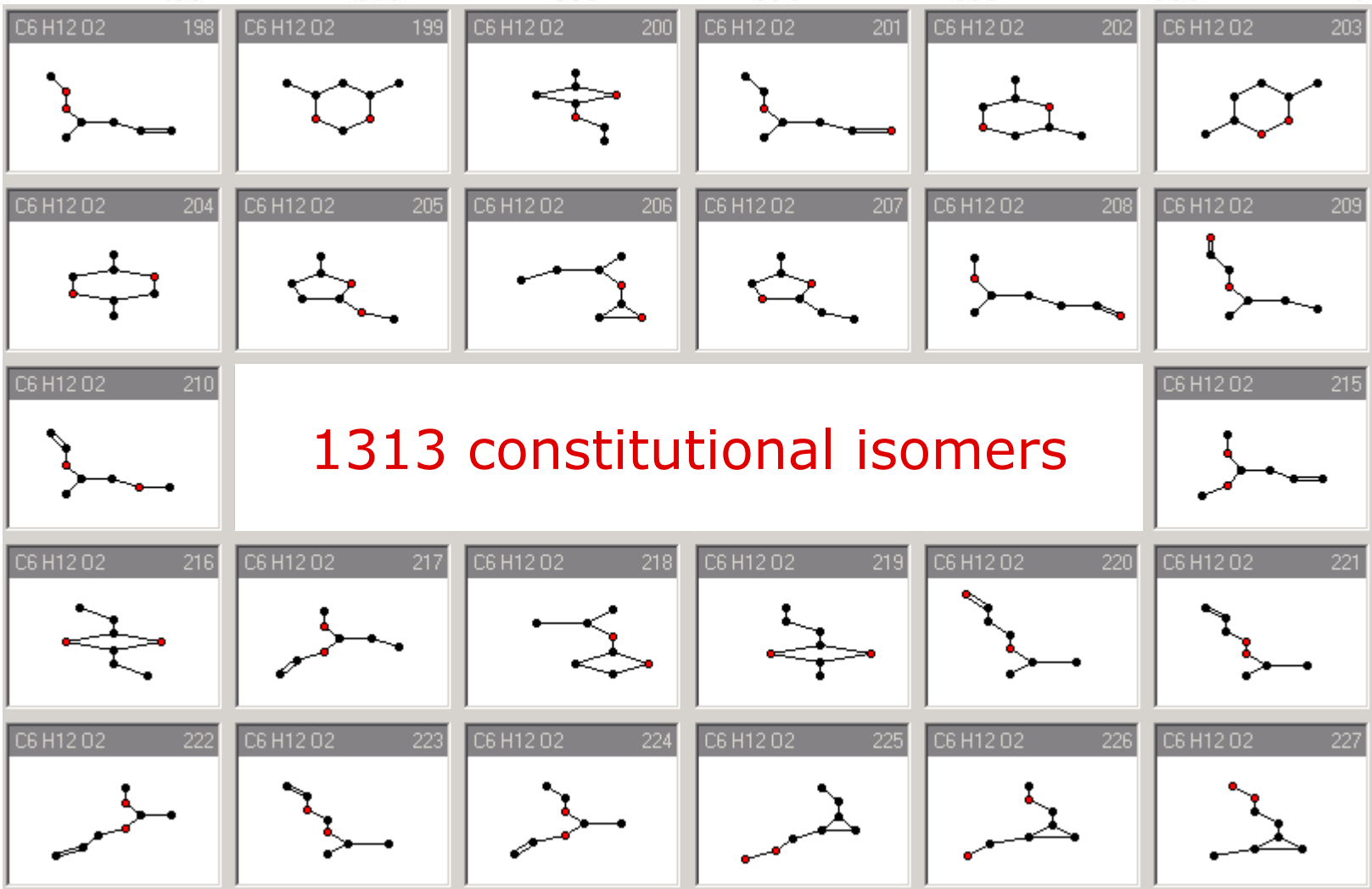
m=116

Exhaustive Structure Generation

- Purpose: Generation of candidate structures
- Input: Molecular formula
- Method:
  - Based on molecular graphs
  - Orderly generation
  - Fast isomorphism tests
- Software: Structure generator MOLGEN
- Output: List of constitutional isomers
  - Complete
  - Redundancy free

KIADIS

Example: Generate all Isomers of $C_6H_{12}O_2$



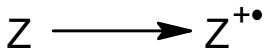
Virtual Fragmentation

- First step: Apply ionization reactions
- Next steps: Apply fragmentation reactions (recursively)
  - Cleavages
  - Rearrangements
- After each step
  - Remove neutral losses
  - Filter isomorphic structures
- Recursion terminates, when no more fragmentation reactions can be applied
- Output: Set of possible fragment ions

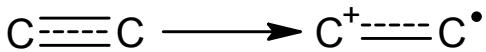
Ionization and Fragmentation Reactions

Ionization reactions

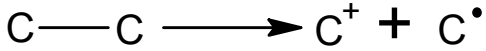
- *n* - ionization



- π - ionization

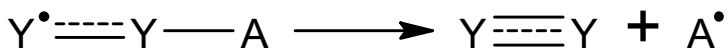


- σ - ionization

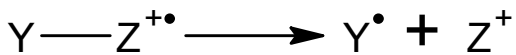


Cleavage reactions

- α - cleavage

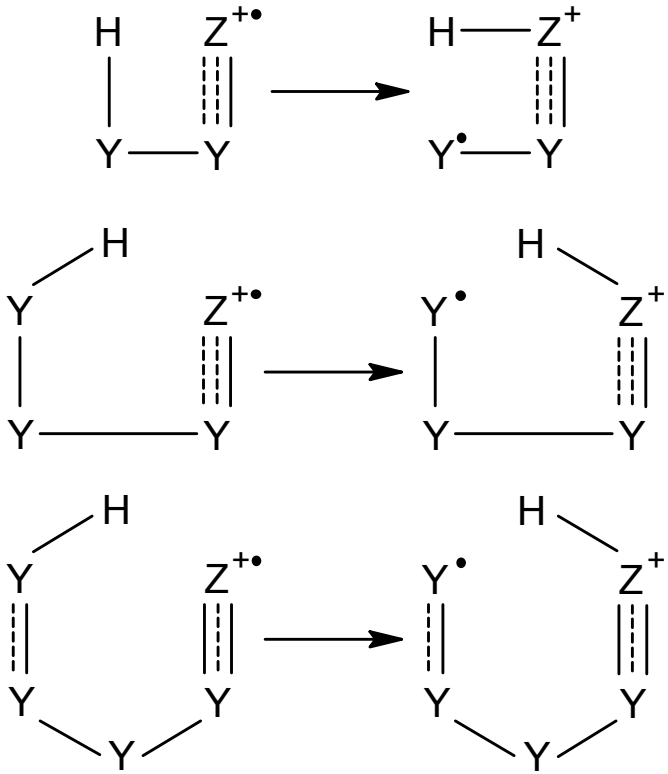


- σ - cleavage



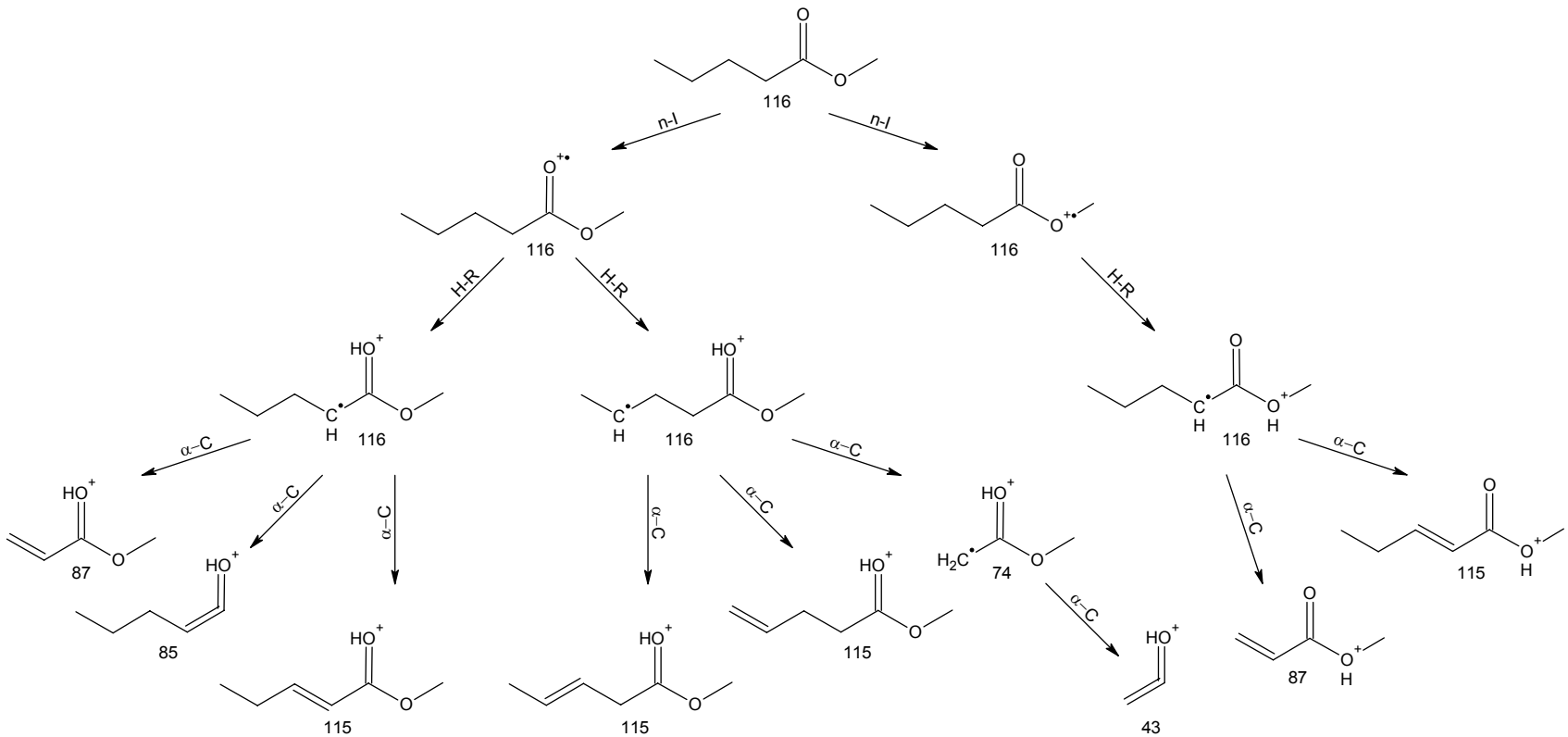
Rearrangement reactions

- H - rearrangements



A: any element; Y: any heavy element; Z: any heavy element bearing a free electron pair

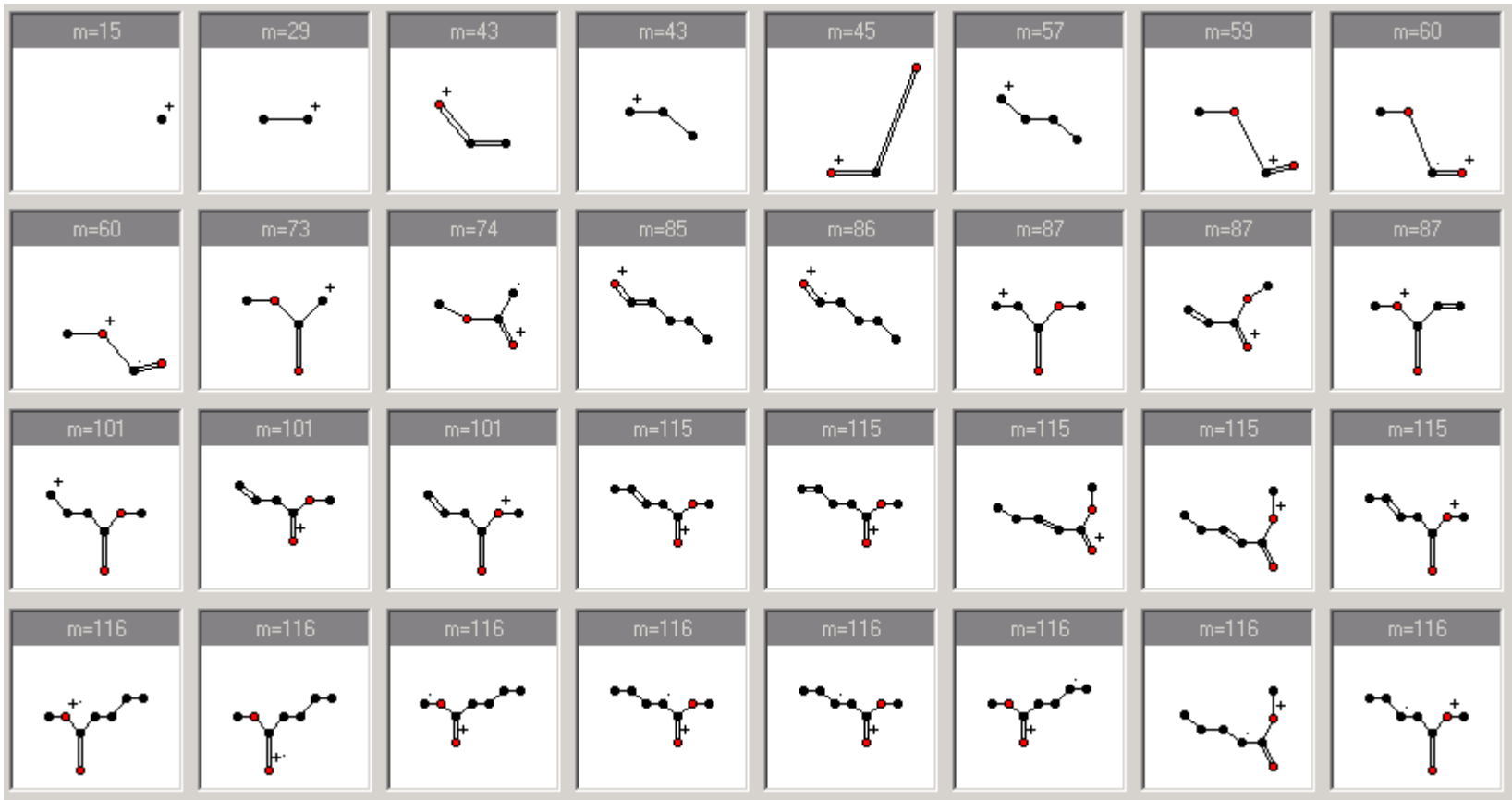
Example: MS Fragmentation Network



Between 7 and 162 fragment ions for the 1313 candidate structures

σ - ionization and H-rearrangement on five atoms were not considered in the above figure.

Example: Virtual Fragmentation



32 fragment ions, 17 molecular formulae, 15 integer masses

Matchvalue Calculation (General)

Requirements for a *perfect* spectra-structure matchvalue

- Values between 0 and 1 for any spectrum I and structure M: $MV(I,M) \in [0,1]$
- Value 1 for the true structure M^T: $MV(I, M^T)=1$
- Smaller values for false structures M^F: $MV(I, M^F) < MV(I,M^T)$

Calculating a spectra-structure compatibility matchvalue

- Determine the fragment ions' molecular formulae $\beta_1,\ldots,\beta_n$
- Calculate the theoretical isotope distributions $I_{\beta_1},\ldots,I_{\beta_n}$

Matchvalue Calculation (Detailed)

- Assuming that all fragment ions were generated, the experimental spectrum I is a linear combination of the fragment ions' theoretical isotope distributions:

$$I = \sum_{i=1}^{n} x_i I_{\beta_i}, \qquad x_i \geq 0$$

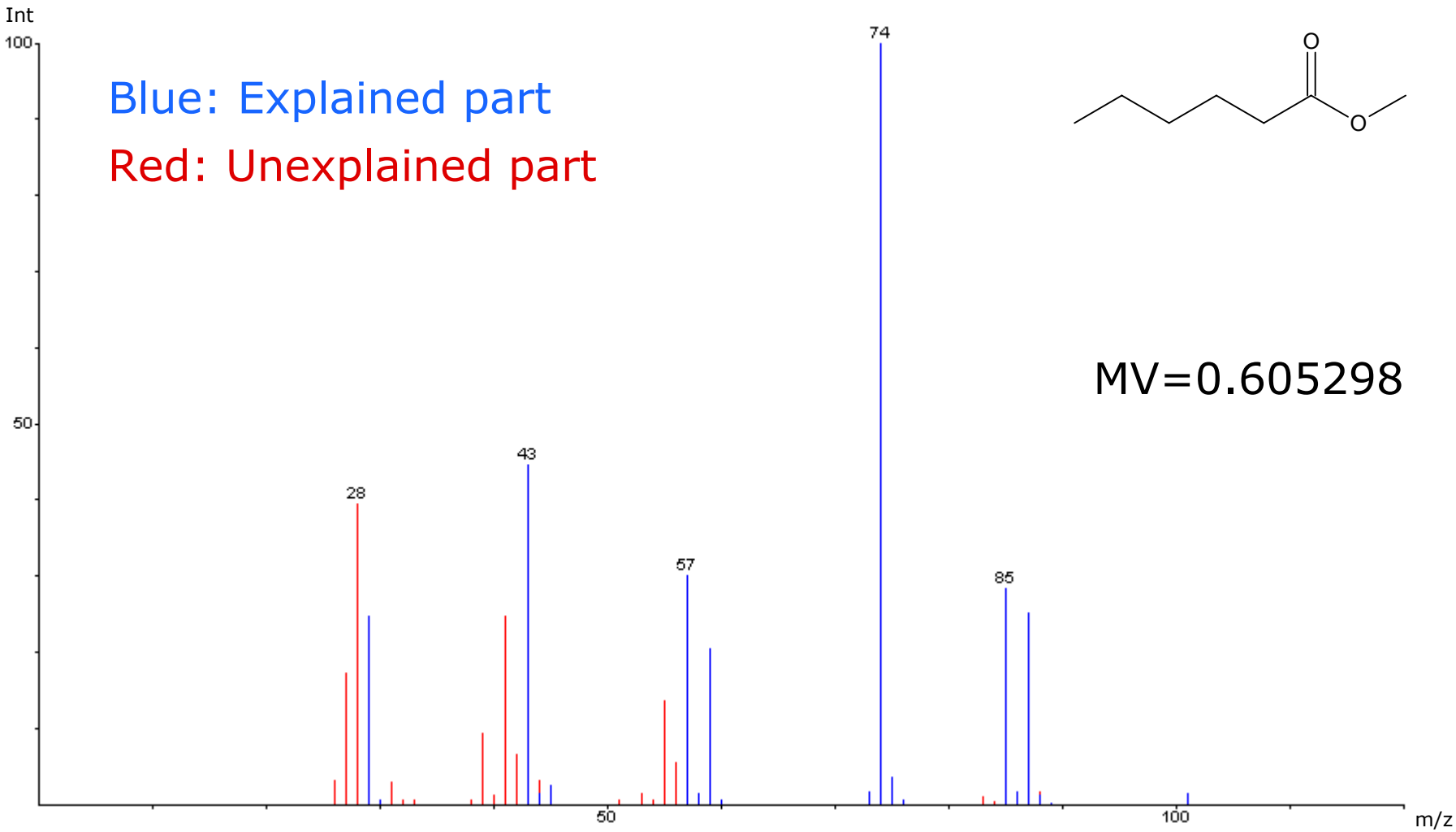
- Treat intensities x_i as unknowns and try to fit the experimental spectrum by a quadratic optimization:

$$\min_{\mathbf{x} \geq 0} \sum_{m} \left(I(m) - \sum_{i=1}^{n} x_i I_{\beta_i}(m) \right)^2$$

- Rescale values to the interval [0,1]:

$$MV(I, M) = 1 - \sqrt{\left(\sum_{m} (I(m))^2 \right)^{-1} \min_{\mathbf{x} \geq 0} \sum_{m} \left(I(m) - \sum_{i=1}^{n} x_i I_{\beta_i}(m) \right)^2}$$

Example: Explained Part of the Spectrum

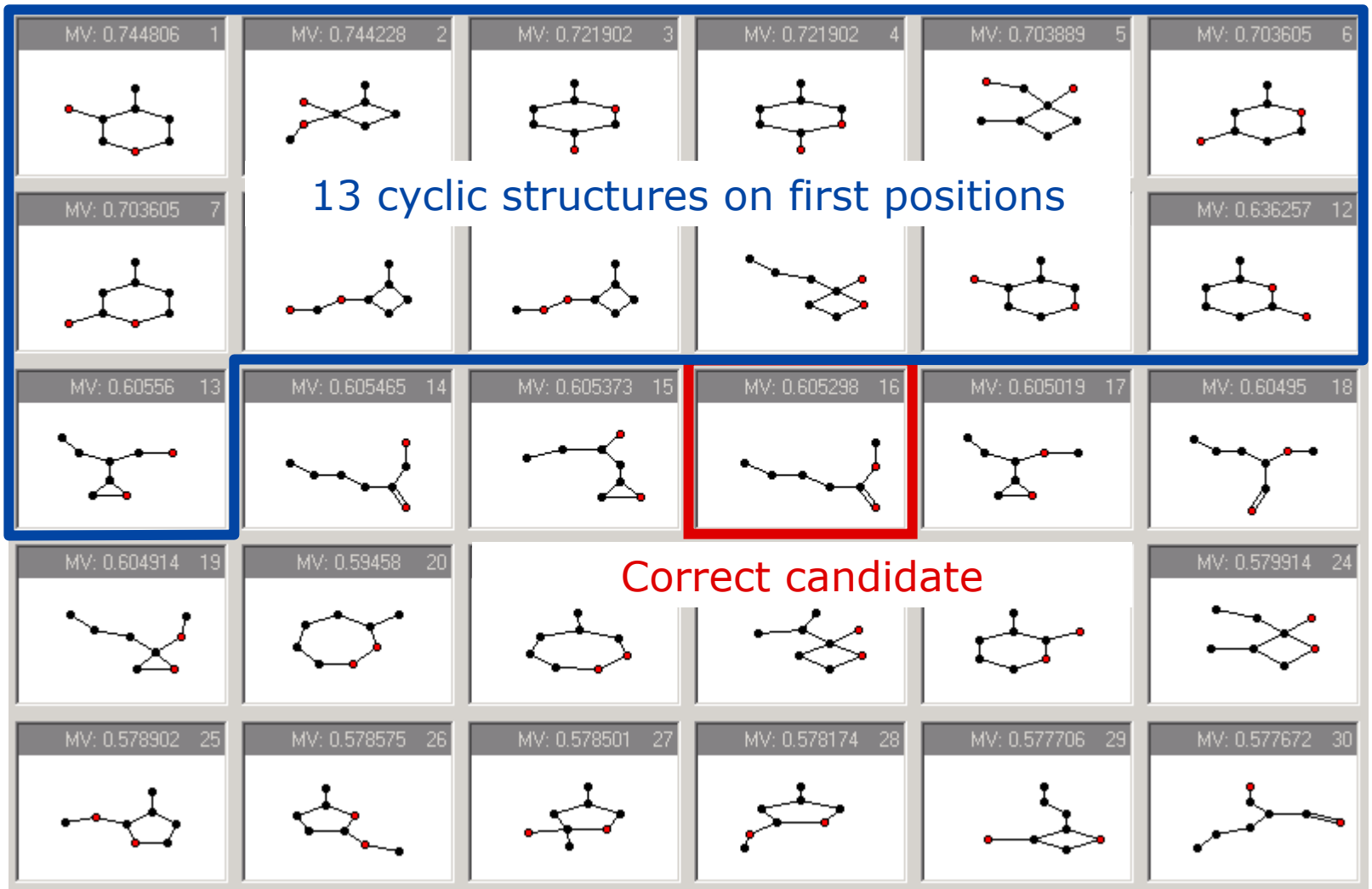


Blue: Explained part

Red: Unexplained part

MV=0.605298

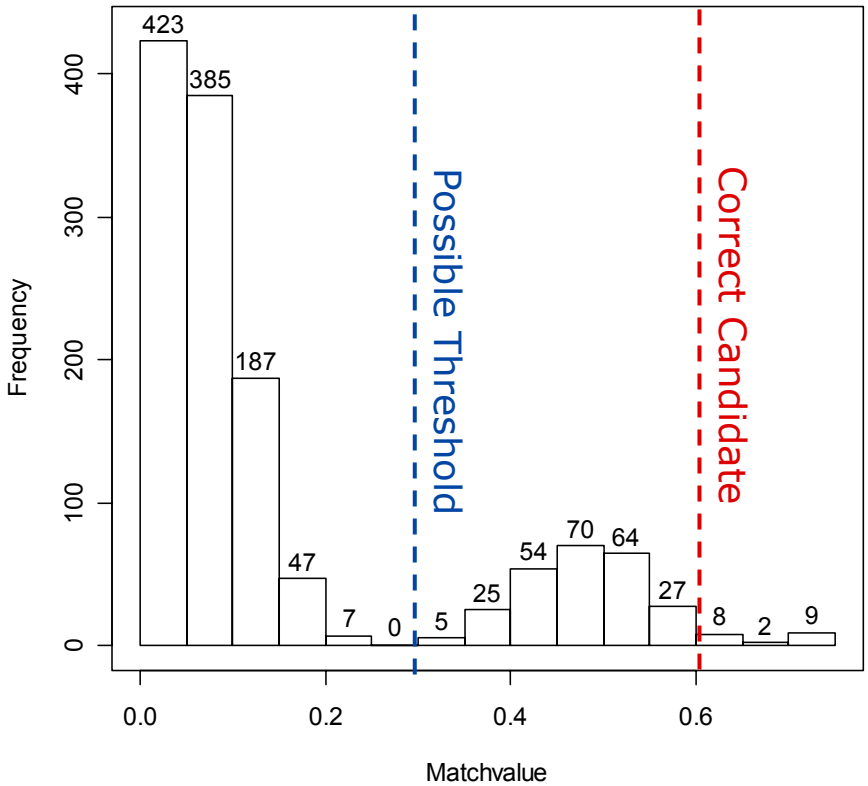
Example: Highest Ranked Candidates



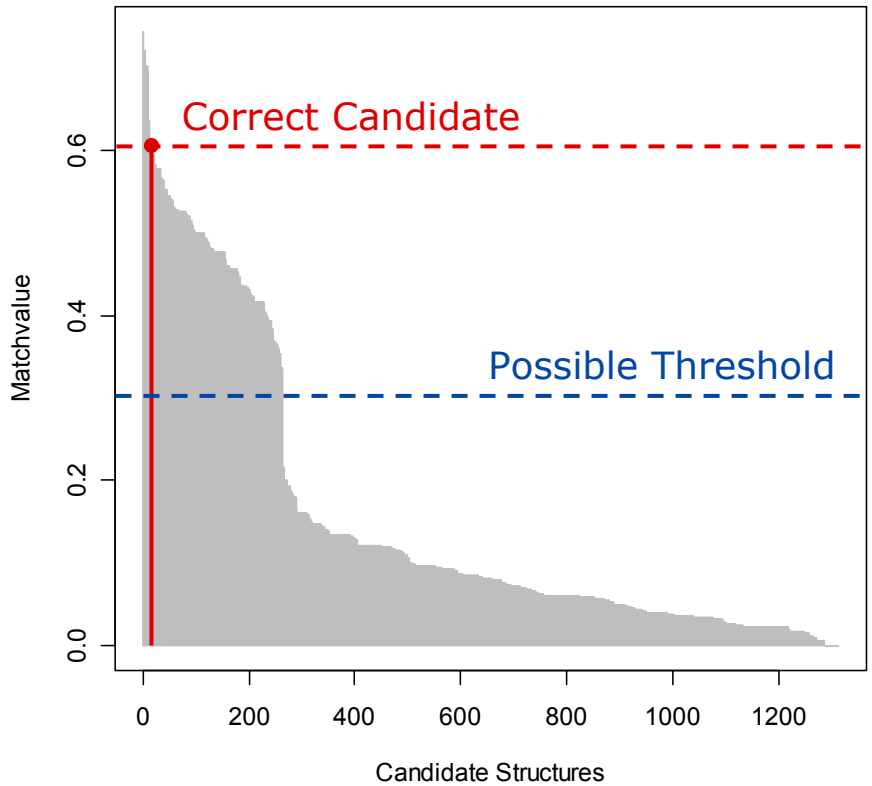
Note: Among the 1313 isomers of $C_6H_{12}O_2$ there are 672 cyclic and 641 acyclic structures.

Example: Matchvalues for the 1313 Candidates

Histogram



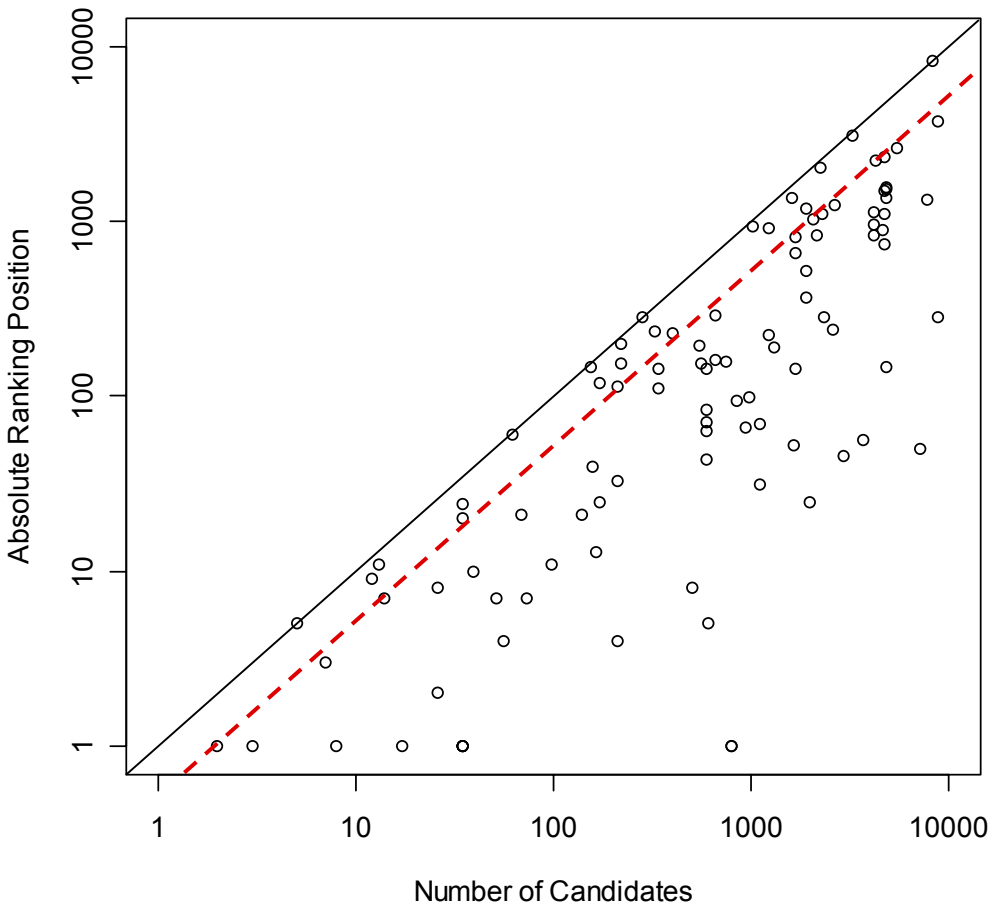
Distribution



A Computational Experiment

- Evaluate abilities of the spectrum-structure compatibility matchvalue on a larger set of data
- Random sample of 100 spectrum-structure pairs from the NIST MS database
  - Molecular mass at most 200 amu
  - More than 1 and at most 10000 constitutional isomers
- Calculate isomers, virtual fragmentation and matchvalues as shown before
- Evaluate rankings using
  - Absolute ranking positions
  - Relative ranking positions

Results: Absolute Ranking Positions

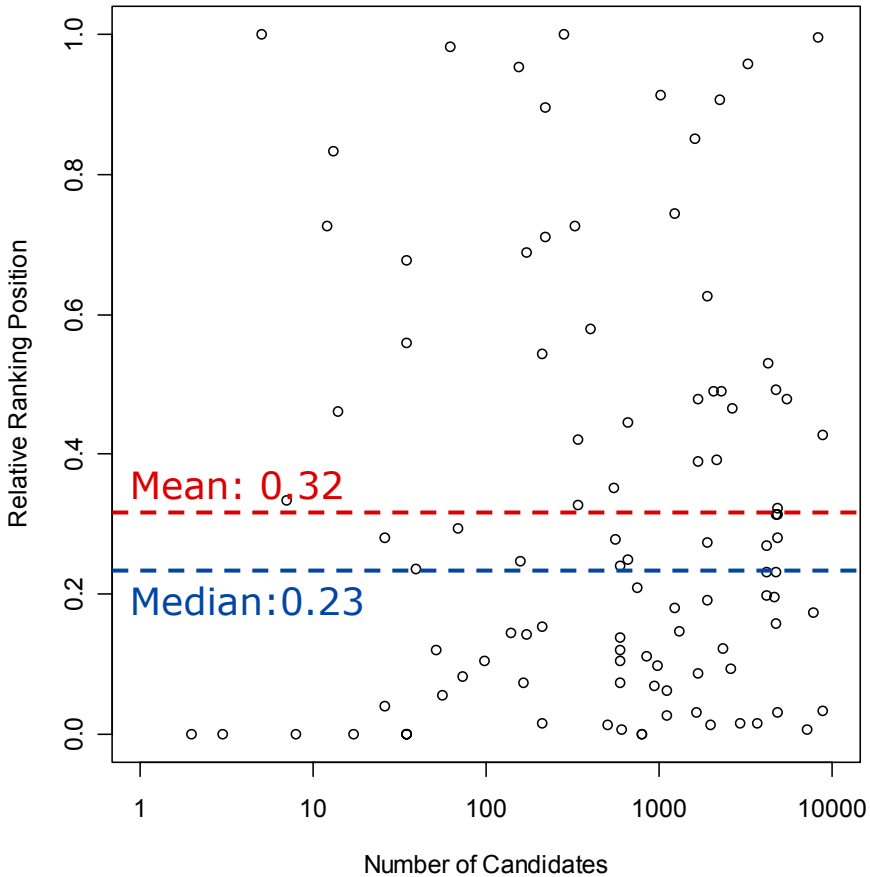


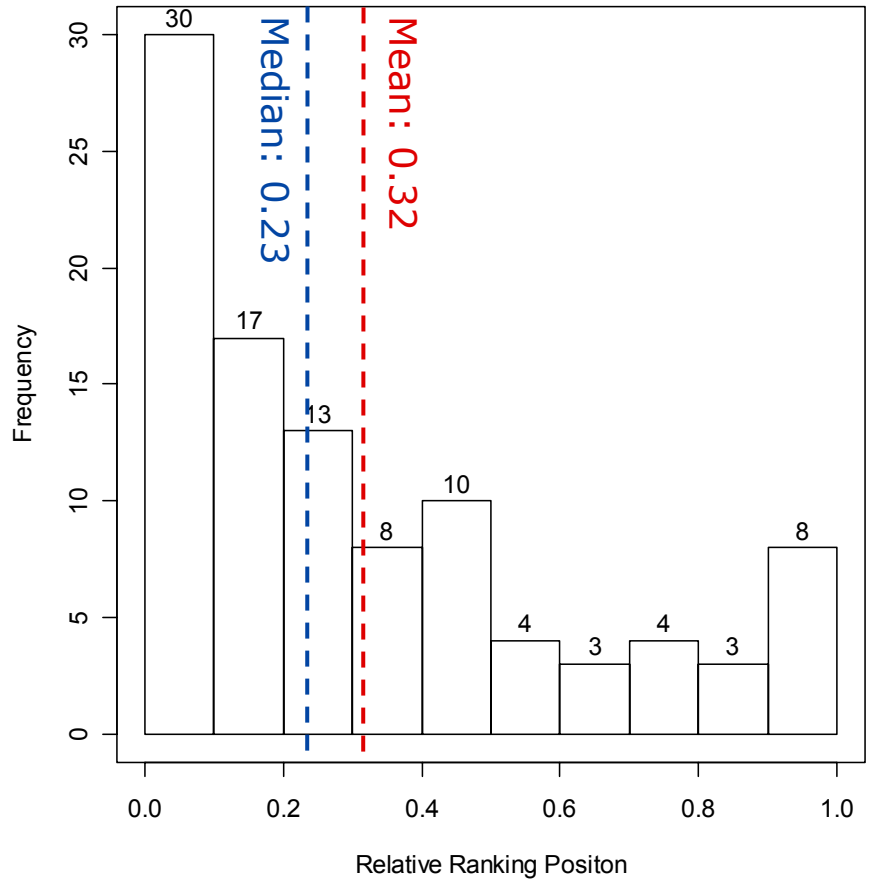
- Average candidates: 1716
- Computation time: 12 h 14 min
- Absolute ranking position 1: 9 instances
- Absolute ranking position less than ½ the number of candidates: 78 instances

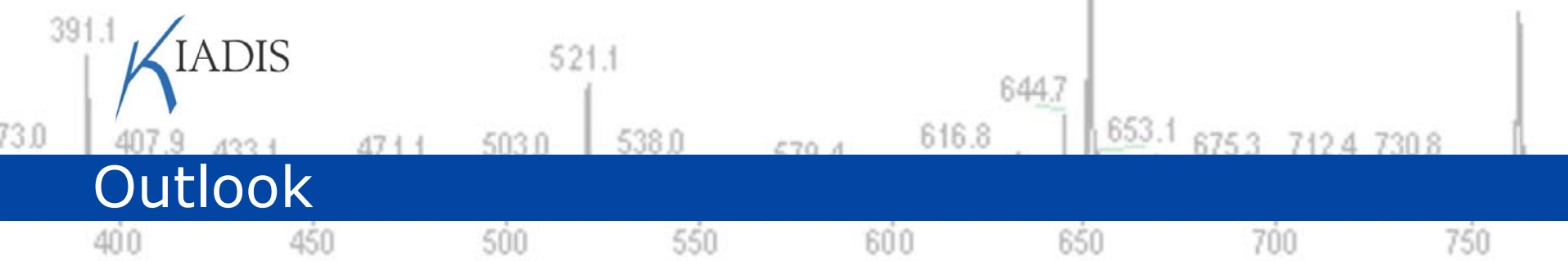
Points below the red line represent instances where the absolute ranking position is less than ½ the number of candidates.

Results: Relative Ranking Positions

$$RRP := \frac{position\ of\ the\ true\ candidate - 1}{total\ number\ of\ candidates - 1} \in [0, 1]$$







Outlook

Improve results on library spectra by

- Other (better) virtual fragmentation

  e.g. HighChem‘s MassFrontier, ACD/MS Fragmenter
- Other (better) matchvalues

  e.g. fuzzy peak patterns by Seebas and Pretsch
- Filtering unlikely chemical compounds

Application in a CombiChem and HRS environment

- Synthesis will provide structural restrictions for candidate generation
- QTOF MS/MS will provide accurate fragment masses for more selective structure ranking

Acknowledgments

- University of Bayreuth, Mathematics II

  Prof. A. Kerber, Prof. R. Laue, Dr. C. Rücker (MOLGEN)
- TU Vienna, Laboratory for Chemometrics

  Prof. K. Varmuza, Dr. W. Werther
- NIST, Mass Spectrometry Data Center

  Dr. S. Stein (NIST MS Library)
- Kiadis B.V., Groningen, Amsterdam

  Prof. H. Irth

THANKS FOR YOUR ATTENTION!