Smart Templates for Peak Pattern Matching with Comprehensive Two-Dimensional Liquid Chromatography (LCxLC)

Stephen E. Reichenbach[a], Peter W. Carr[b], Dwight R. Stoll[b], and Qingping Tao[c]

[a]Computer Science & Engineering Department
University of Nebraska-Lincoln

[b]Department of Chemistry
University of Minnesota

[c]GC Image, LLC
Lincoln NE

HPLC 2008, Baltimore MD, 14 May 2008

Comprehensive Two-Dimensional Liquid Chromatography

Comprehensive Two-Dimensional Liquid Chromatography (LCxLC) is ever faster and more powerful.

The greater peak separation capacity of LCxLC is especially critical for important, but complex biochemical applications, including proteomics and metabolomics.

The paucity of efficient, convenient and sufficiently powerful data analysis tools is the greatest impediment to its wide application.

Peak Identification and Classification

Fundamental goal: identify, classify, and quantify constituent compounds from chromatographic peaks.

Traditional approaches for peak identification include:

- Retention-time windows
- Multispectral matching (e.g., mass spectra library search)

Retention-time windows must be small for "crowded" separations. Chromatographic variations may cause peaks to "drift" outside of the windows.

Multispectral matching may be uncertain for large chemical domains with chemically similar compounds.

Smart Templates for Peak Pattern Matching

New approach for peak identification and classification.

Smart Templates™ record:

- Multidimensional retention-time pattern of peaks.
- Analytical metadata, including peak identities, groupings, labels, etc.
- Rules for recognizing peaks (e.g., based on multispectral characteristics).

The Smart Template pattern is recognized in subsequent data and the analytical metadata are used to identify and classify peaks.

Background Correction

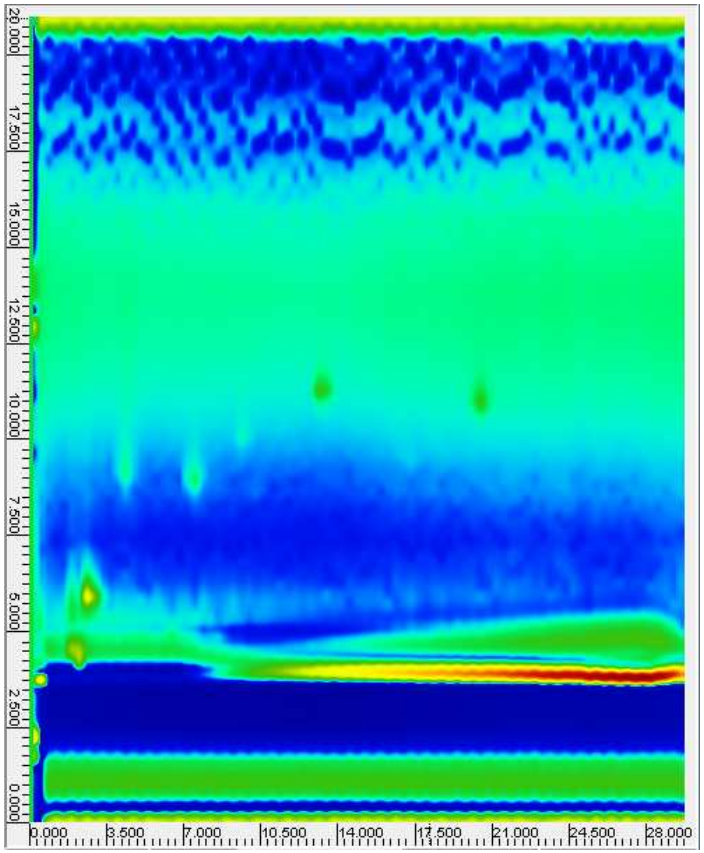
LCxLC data contains significant variations in the background.

Background must be corrected for accurate peak detection and quantitation.

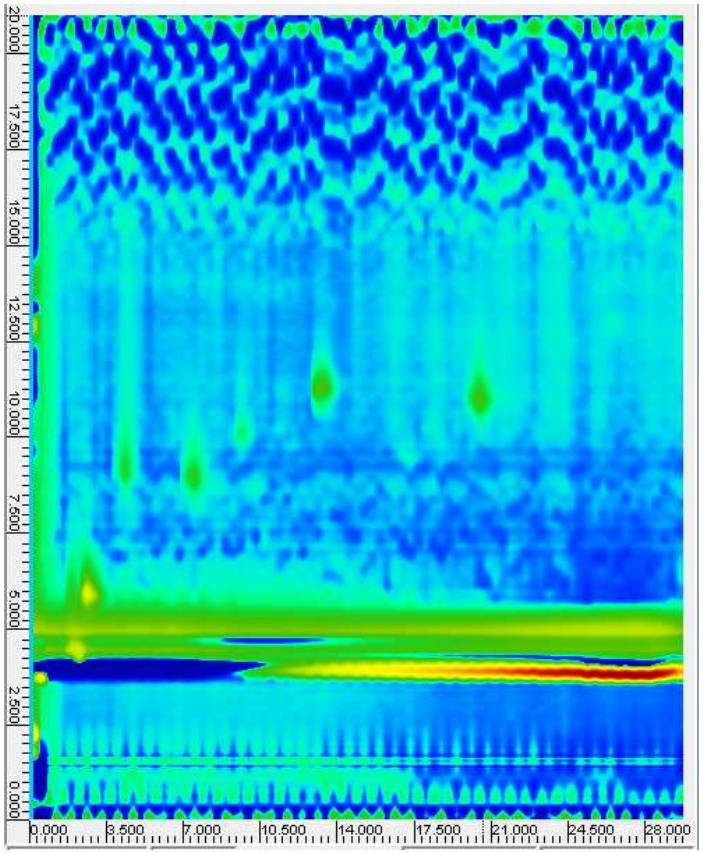
New method builds statistical models of the slowly varying background in each of the two dimensions of separation and then subtracts the background model value from the data.

Background correction in each "channel" of multispectral data.

Background Correction Example



Before



After

Peak Detection

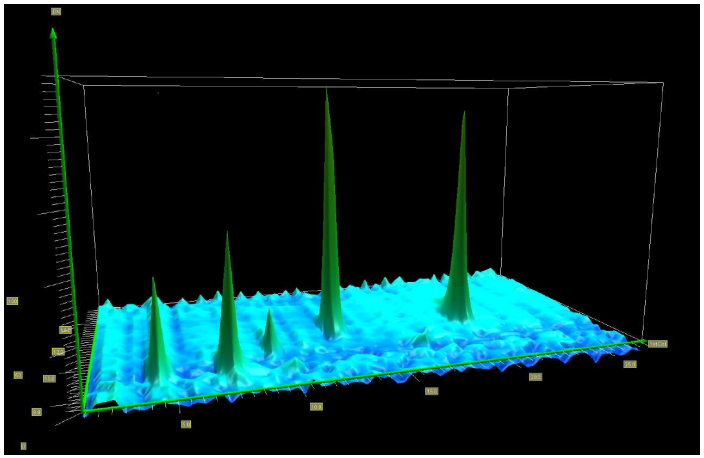
Simple two-dimensional approach detects each peak from its apex to surrounding minima. (Drain algorithm developed for GCxGC.)

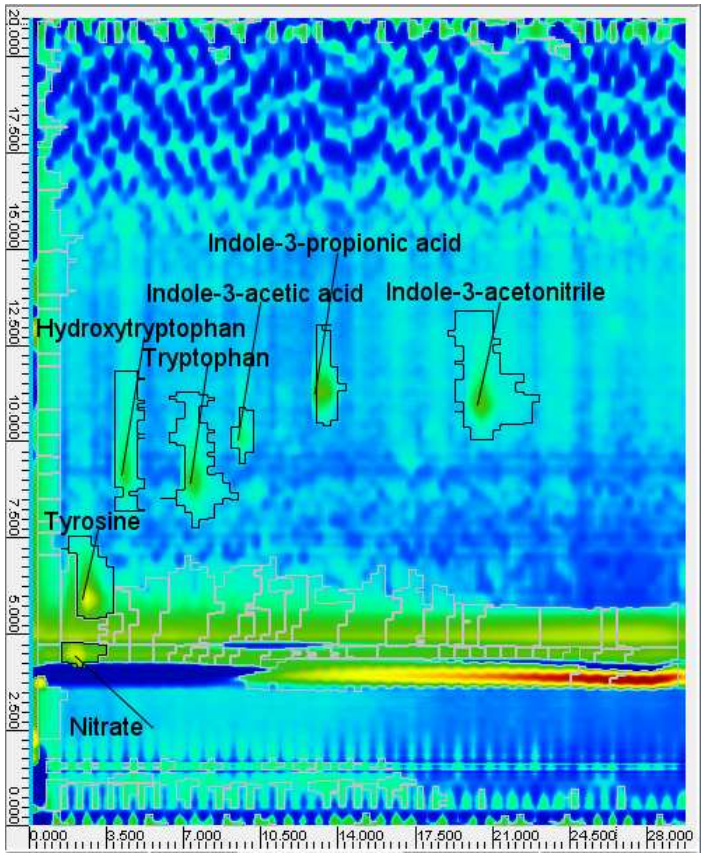
Example uses only the total intensity count (TIC) of the UV data. (Detection threshold for apex magnitude and footprint area.)

Multivariate chemometric methods may be able to unmix (e.g., deconvolve) coeluted peaks based on multispectral signatures.

Peak Detection Example

Region with indoles shown in 3D perspective (below). Detected target peaks outlined in black (right).





Spectral Identification

Spectral matching based on similarities or differences between a spectrum and reference/library spectra.

Spectral matching may be uncertain.

For example, spectra of 5 indole standards in detected peaks matched with database of UV absorbance spectra of 26 indoles. Correct spectral match from 33% (indole-3-acetic acid) to 100% (indole-3-acetonitrile).

Spectral matching is insufficient for complex mixtures.

Templates and Matching

Templates record the retention-time pattern of peaks along with analytical metadata (peak identifications, groupings, etc.).

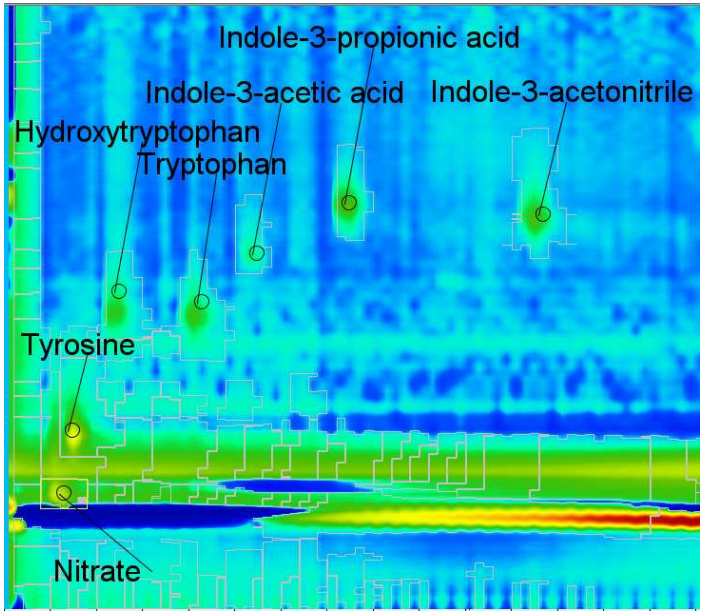
Goal of matching is to transform the template pattern in the retention-time plane (e.g., shifting and scaling) to match the detected peaks in another chromatogram.

Matching criteria is the number of peak correspondences between the template and the target.

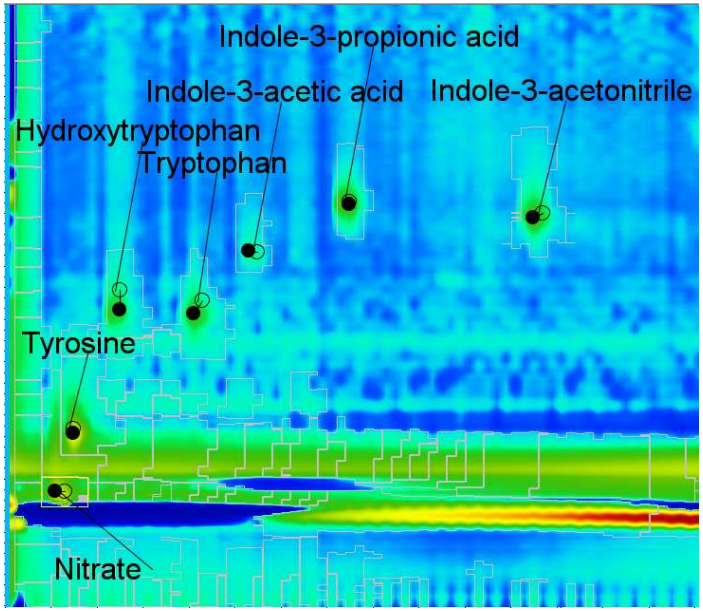
Matching is subject to geometric transformation parameters and correspondences are subject to retention-time window.

Template Matching Example

Standards: Template (#1/64) & Target (#20/64)



Overlay



Matching

Retention-Time Variability

Chromatographic variability changes peak patterns.

Template Sequence #	Target Sequence #	Translation(1)	Translation(2)	Scaling(1)	Scaling(2)
1	2	0.0000	−0.0711	1.0000	1.0119
2	20	−0.2493	−0.1014	0.9924	0.9788
20	38	−0.1069	0.0278	0.9990	1.0032
38	63	−0.2007	0.1883	0.9851	1.0286
63	64	0.0000	−0.0458	1.0000	1.0042
1	64	−0.5480	−0.0036	0.9771	1.0273

Retention-Time Variability

Observations:

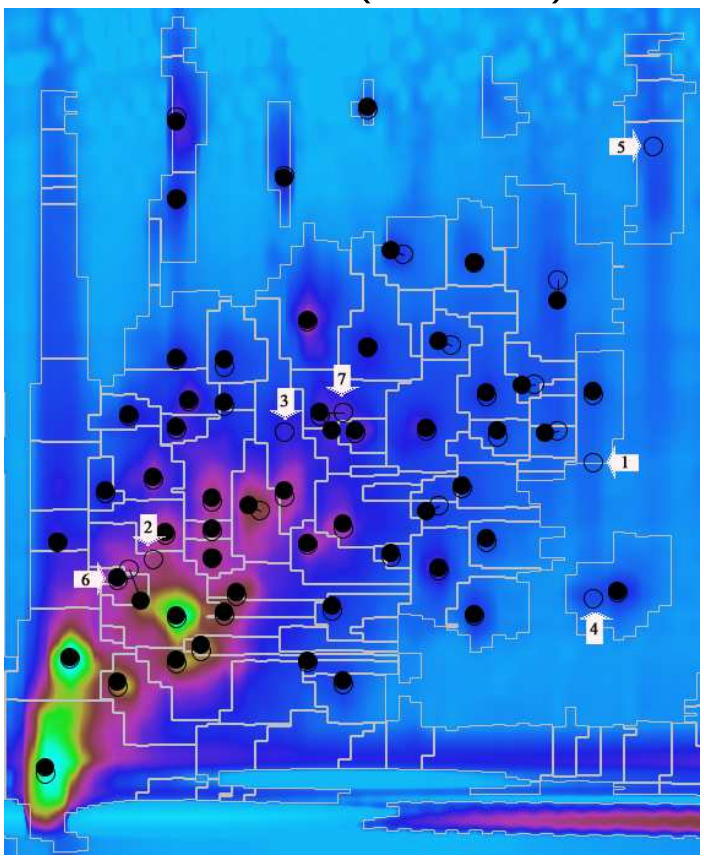
- Adjacent runs have small variability.
- First column translation and scaling are monotonically non-increasing, with larger cumulative effect.

With this simple example, matching parameters can be increased to find the correct correspondences.

More complex data presents more difficult pattern matching, requiring smarter matching.

Template Matching Errors

Urine sample (control): Template (#11/64) & Target (#15/64)



Arrow 1: Peak error, peak not detected cannot be matched.
Arrows 2–4: Peak error, merged peak not detected cannot be matched.
Arrow 5: Match error, peak too distant not matched.
Arrows 6–7: Match error, merged peak causes incorrect peak match.

Smart Templates

Smart Templates use chemical rules to constrain peak correspondences during template matching.

Rules in Computer Language for Identifying Chemicals (CLIC).

CLIC has functions for multispectral comparison; multispectral library search; multispectral characteristics (e.g., spectral peak ordering, ratios), chromatographic peak statistics (e.g., TIC, symmetry, retention time); and arithmetic, relational, and logical operators.

Rules reduce matching errors with incorrect peaks. Rules allow larger retention-time windows to reduce matching errors from peak "drift".

Automated Rules for Smart Templates

Evaluate multispectral comparisons of template spectrum with: (1) correct spectra and (2) incorrect spectra.

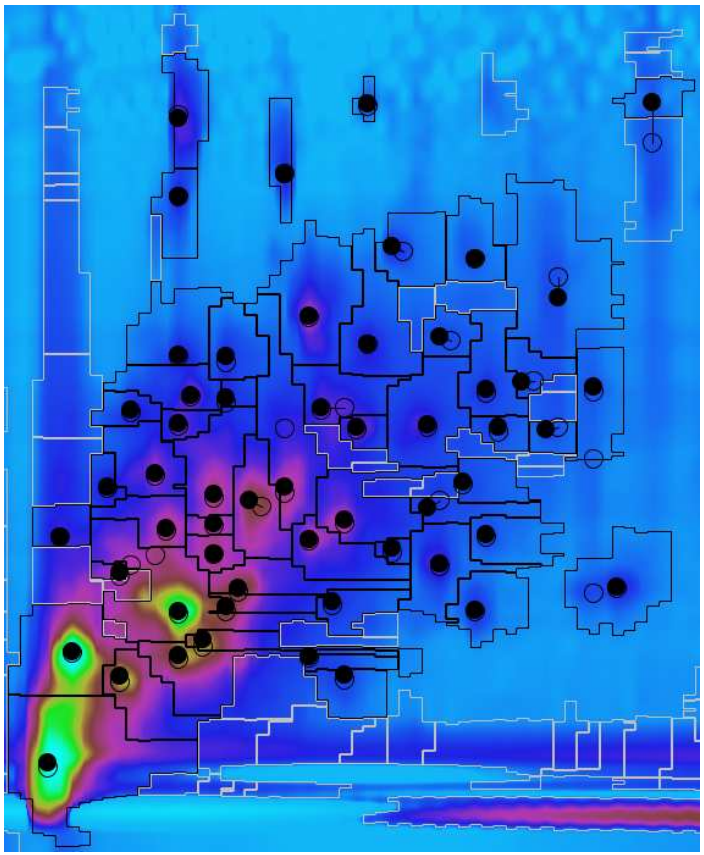
Construct a rule with a threshold that allows correct matches and disallows incorrect matches.

$$EuclideanDistance(\text{“}<ms>\text{”}) < 0.2$$

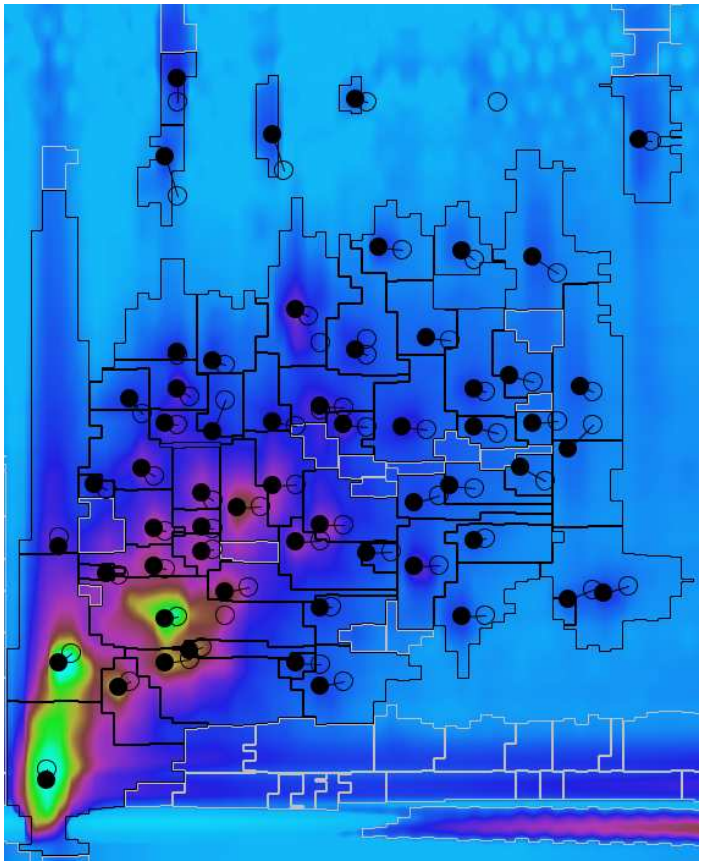
Different rule threshold for each template peak.

Smart Templates Examples

Template (#11/64)
& Target (#15/64)



Template (#3/64)
& Target (#63/64)



Conclusions

Smart Templates can automatically identify and classify peaks even with drift from variable chromatography.

Smart Templates record peak pattern, metadata, and chemical rules. Peak matching recognizes the template pattern in new chromatograms.

Applicable to LCxLC and other multidimensional separations.

Applicable to UV, mass spectrometric, and other detectors.

Questions ?

Steve Reichenbach: reich@unl.edu
LCxLC Software: www.LCxLC.com