Chromaligner: a web server for chromatogram alignment

San-Yuan Wang[1,2], Tsung-Jung Ho[1,2], Ching-Hua Kuo[2,3,*] and Yufeng J. Tseng[1,2,3,4,*]

[1]Department of Computer Science and Information Engineering, [2]The Metabolomics Group, [3]School of Pharmacy, College of Medicine and [4]Graduate Institute of Biomedical Electronic and Bioinformatics, National Taiwan University, Taipei, Taiwan

Associate Editor: Jonathan wren

ABSTRACT

Summary: Chromaligner is a tool for chromatogram alignment to align retention time for chromatographic methods coupled to spectrophotometers such as high performance liquid chromatography and capillary electrophoresis for metabolomics works. Chromaligner resolves peak shifts by a constrained chromatogram alignment. For a collection of chromatograms and a set of defined peaks, Chromaligner aligns the chromatograms on defined peaks using correlation warping (COW). Chromaligner is faster than the original COW algorithm by k^2 times, where k is the number of defined peaks in a chromatogram. It also provides alignments based on known component peaks to reach the best results for further chemometric analysis.

Availability: Chromaligner is freely accessible at http://cmdd.csie .ntu.edu.tw/~chromaligner

Contact: kuoch@ntu.edu.tw; yjtseng@csie.ntu.edu.tw

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 BACKGROUND

Chromatography is the most important separation technique. With the improvement of chromatographic technology and chemometric analysis, there is a growing interest to use the entire chromatogram to detect differences between pre-classified sample sets. Fingerprint analysis (Cianchino *et al.*, 2007) and metabolomics (Nordstrom *et al.*, 2006) are two rapid growing research fields that use the entire chromatogram for authentication of herbal medicine, clinical diagnostic and evaluation of drug toxicity or efficacy. Most samples used for fingerprint analysis and metabolomics are composed of a large number of diverse components and slight retention time shifts could result in the incorrect detection of differences between pre-classified samples. Retention time shifts in chromatograms may be originated from variation in flow rate, temperature and column aging of the chromatographic system (Christin *et al.*, 2008). It is important to correct the retention time shits before chemometric analysis of a chromatographic data. For the reasonable chemometric analysis with chromatograms, specifically the fingerprint analysis of plant extracts (Cianchino *et al.*, 2007) or metabolomics analysis (Nordstrom *et al.*, 2006), certain known peaks, as the key chromatogram features, should be aligned together before any statistical treatment.

Current publicly available tools for chromatogram alignment such as ChromA (Hoffmann and Stoye, 2009), VIPER (Monroe *et al.*, 2007) and Metalign (Lommen, 2009) are designed for chromatographic technique coupled to mass spectrometers. Due to the high instrumental cost of mass spectrometry, many laboratories use UV and fluorescence detectors for signal detection. However, current alignment tools cannot be applied to such data sets.

The two most commonly used alignment algorithms are correlation optimized warping (COW) (Nielsen *et al.*, 1998; Tomasi *et al.*, 2004) and dynamic time warping (DTW) (Tomasi *et al.*, 2004) that focus only on the comprehensive optimal solution only, and failed on chromatograms with large peak shifts as commonly seen in capillary electrophoresis (CE). Most alignment methods are implemented in MATLAB code that is typically difficult for a wet lab analytical chemist to use. Several algorithms are implemented as web servers, for example, Tomasi *et al.* (2004) provide one employing the COW and DTW functions for chromatogram alignments. However, users can only compare two chromatograms with the exact length and less than 2000 data points at one time. This restriction makes it difficult to align common chromatogram of 30 min (more than 3000 points) or longer that are generated by a HPLC or a capillary electrophoresis. Our goal is to use a constrained chromatogram alignment algorithm to resolve those issues. Here, we provide a web interface, Chromaligner, for chromatogram alignment algorithms with constraints using COW. It can be used to align retention time for chromatographic methods coupled to spectrophotometers.

2 METHOD

Among the current alignment methods, COW has the advantage of matching small and large peaks, while preserving their shape but the processing is typically slow. Chromaligner uses a constrained chromatogram alignment based on COW to reduce the complexity and the processing time.

Prior to aligning the chromatogram, Chromaligner corrects the baseline and selects the peaks. The chromatogram is partitioned into walking windows of 25 s. To determine whether a point is a peak, that point must first be the maximum in a window. Then, the standard deviation of intensities of each walking window is ranked and the third quartile of the standard deviation is selected. If the difference of the maximum intensity and minimum intensity in this walking window exceeds the third quartile, the point is considered a peak.

To reduce the processing time and to improve accuracy, Chromaligner uses either user-defined peaks (Fig. 1a) or it automatically selects peaks by matching UV spectrum from a diode array detector (DAD). If the Pearson's correlation of peaks from the UV spectrum between the test and target chromatograms is >0.98, within 1 min, the corresponding peak(s) will be selected. Chromaligner further divides the chromatograms into segments,

*To whom correspondence should be addressed.
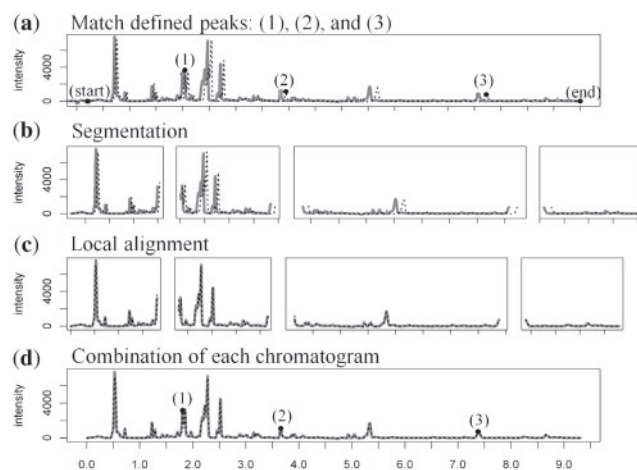
Fig. 1. Chromaligner workflow.

with respect to these peaks (Fig. 1b) and then, performs COW on each segment (Fig. 1c).

For a chromatogram of length L, COW and DTW can resolve the alignment in $O(L^2)$ time through dynamic programming (DP) (Wang *et al.*, 2005). For Chromaligner, if the number of defined peaks is k, the segment length to be aligned is reduced to $O(L/k)$. Therefore, the time complexity for each segment is reduced to $O(L^2/k^2)$. However, COW requires manual tuning of the section length and the slack number to achieve better alignment. Our experience showed that the local optimal solution using COW is generally obtained, while the section length is between 20 and half of L with the slack number between 1 and 5. By heuristically tuning a chromatogram, COW needs to tune $O(L)$ times to achieve the optimal alignment in $O(L^3)$ time. To completely align a chromatogram, Chromaligner tunes $O(L/k)$ times for each segment and continues to align a total of $k+1$ segments for a chromatogram; making the time complexity $O(L^3/k^2)$. Thus Chromaligner aligns chromatograms k^2 times faster than COW.

By DP nature, both DTW and COW require $O(L^2)$ in space. However, the size of the DP table in COW is dependent on the section length (Nielsen *et al.*, 1998). A section length of m results in a DP table size of L^2/m and reduces the space complexity to $O(L^2/m)$ for COW which is smaller than the $O(L^2)$ of DTW. The space complexity of Chromaligner is $O(L^2/k^2m)$ due to the fact that each segment length L is reduced to L/k.

3 INPUT AND OUTPUT

Chromaligner requires three sets of files in plain text (tab delimited with a .txt extension): an index file, raw data and parameter files. The raw data file must contain retention time to intensity information. UV spectrum from a DAD can be provided for auto-picking peaks for Chromaligner. For additional details on the file format, please visit the Chromaligner website.

Chromaligner is able to export results in chromatograms. The aligned files are transformed into retention time and intensity data for further chemomatric analysis. Overlaid chromatograms that depict shifts between the target and sample are also provided.

4 CASE STUDY

The World Health Organization (WHO) has suggested chromatographic fingerprinting for quality evaluation of the herbal medicines. We applied Chromaligner to chromatograms obtained from different instruments and herbal extracts. Chromatograms of

Table 1. Aligned results of Scutellaria and Rhubarb using COW and Chromaligner

| | | INITIAL | After warping | | | |
| | | | COW | | Chromaligner | |
	Instruments	R	R	T	R	T
Scutellaria-1	HPLC-DAD	0.44	NA	NA	0.92	4.8
Scutellaria-2	HPLC-DAD	0.68	NA	NA	0.96	4.8
Rhubarb-1	CE-UV	0.16	0.24	61.9	0.87	2.4
Rhubarb-2	CE-UV	0.27	0.53	61.9	0.63	2.4

R, Pearson's correlation; T, execution time (s); NA: not available.

Scutellaria were obtained from an HPLC-DAD instrument. Using the auto-picking function in Chromaligner to select the peaks of DAD (which is absent from original COW), the correlation greatly improved to a great extent after alignment. We also used Chromaligner to analyze a well-known traditional Chinese medicine and cooking herb, rhubarb (Table 1). The chromatograms of rhubarb were obtained from a CE-UV instrument. Compared to COW, Chromaligner provided better alignments within a shorter execution time. Thus, Chromaligner is an efficient and accurate tool for chromatogram alignment.

ACKNOWLEDGEMENTS

REFERENCES

Christin,C. *et al.* (2008) Optimized time alignment algorithm for LC-MS data: correlation optimized warping using component detection algorithm-selected mass chromatograms. *Anal. Chem.*, **80**, 7012–7021.

Cianchino,V. *et al.* (2007) Fingerprint analysis and synthetic adulterant search in Hedera helix formulations by capillary electrophoresis. *Pharmazie*, **62**, 262–265.

Hoffmann,N. and Stoye,J. (2009) ChromA: signal-based retention time alignment for chromatography-mass spectrometry data. *Bioinformatics*, **25**, 2080–2081.

Lommen,A. (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal. Chem.*, **81**, 3079–3086.

Monroe,M.E. *et al.* (2007) VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics*, **23**, 2021–2023.

Nielsen,N.-P.V. *et al.* (1998) Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr A*, **805**, 17–35.

Nordstrom,A. *et al.* (2006) Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. *Anal. Chem.*, **78**, 3289–3295.

Ramos,L.S. (1994) Characterization of mycobacteria species by HPLC and pattern recognition. *J. Chromatogr. Sci.*, **32**, 219–227.

Tomasi,G. *et al.* (2004) Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemometrics*, **18**, 231–241.

Wang,J.C.P. *et al.* (2005) Application of fast fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets. *Anal. Chem.*, **17**, 5655–5661.