Proving the preclusion of data manipulation using parallel data acquisition in chromatography

David Thomas Marehn[1,2,3, a *], Detlef Wilhelm[1,b], Heike Pospisil[2,c] and Roberto Pizzoferrato[3,d]

[1]AnaTox GmbH & Co. KG, Fuerstenwalde, 15517, Germany

[2]TUAS Wildau, High Performance Computing in Life Sciences, Wildau, 15745, Germany

[3]Universitá degli Studi di Roma "Tor Vergata", Rome, 00133, Italy

[a]david_marehn@anatox.de, [b]wilhelm@anatox.de,
[c]pospisil@th-wildau.de, [d]pizzoferrato@uniroma2.it

Keywords: Chromatography, Data Acquisition, Parallelism, Pharmacopoeia

Abstract. Traceability has an enormous value for companies, but especially for those working in the regulated environment. It plays a special role in the field of pharmacy with respect to manufacturing, controlling and distributing batches of drugs. Through the guidance of Good Manufacturing Practice (GMP) traceability should be ensured. An increasing number of pharmaceutical companies are member of one of the global pharmacopoeias (United States Pharmacopeia, European Pharmacopeia and Japanese Pharmacopeia). The specifications of these pharmacopoeias describe the best practice in documentation, control, qualification and risk management.

But however, the pharmacopoeias are written very generally and do not distinguish between the vendors of the analytical instruments. Here, we analyze how chromatographic analyses and data acquisition rely on a specific vendor of the device and the chromatography data system (CDS), the controlling software. We present a way to compare the data acquisition of different CDSs communicating with HPLC instruments.

A newly developed software called *Data Collector* allows the acquisition of data from a HPLC detector parallel to the controlling CDS in the same run. Two HPLC systems and two different CDSs using a well defined sample standard have been tested. The direct comparison of the acquired data precludes unexpected data manipulations of both tested CDSs and shows that there are primarily deviations between the CDSs due to time variations only which depend on the sampling rate. All in all the *Data Collector* can be used for the traceability of data acquisition.

Introduction

The majority of pharmaceutical companies work in accordance with one of the global players of the existing pharmacopeias: United States Pharmacopeia (USP) [1], European Pharmacopeia (EP) [2] and Japanese Pharmacopeia (JP) [3]. All established types of the separation technique chromatography are listed and described in the general chapters of these pharmacopoeias [4,5,6]. The general chapters about the chromatography deal with their specification, performance, evaluation and system suitability tests. On the other hand the exact parameters of a chromatography technique used to analyze a specific medical substance are defined in the corresponding chapter, in the so called monograph. All given parameters here are defined generally and are not specified to a vendor of the chromatographic system.

Each vendor of an analytical instrument (as for example a high performance liquid chromatography (HPLC) or gas chromatography (GC) system distributed e.g. from *Agilent Technologies*) creates its own specification and declares tolerances, that will be ensured by its product [7]. That means deviations of measurement results between instruments supplied by different vendors are

unavoidable. In contrast to the described challenge it is required that chromatography data systems (CDSs) controlling one given instrument should provide at least the same raw data even if the following data processing (peak detection and integration) could differ.
Based on this assumption we tried to figure possible variances between two different CDSs acquiring the detector signal data of one HPLC system. Additionally we focused on unknown manipulations on the acquired data by the chromatography data systems.

Materials and Methods

In practice a laboratory assistant cannot communicate with one HPLC system using two different chromatography data systems (CDSs) at the same time because the first software that establishes a connection will lock the instrument. That is why it is not possible to acquire signal data of a HPLC detector by two CDSs within the same run.
Therefore, a new software has been written called *Data Collector* which is able to connect to a HPLC system manufactured by *Agilent Technologies*. It is constructed as a tool that searches for a detector in the system and acquires its signal data parallel to the running CDS. The *Data Collector* do not parameterize and control the HPLC itself. The necessary communication is based on the freely available *Licop* library provided by *Agilent Technologies* [8]. The whole configuration using the *Data Collector* is shown in Fig. 1. The tool runs in parallel to the CDS which parameterize the system and start/stop experiments. The only prerequisite is an available LAN connection to the HPLC system which allows two instances.

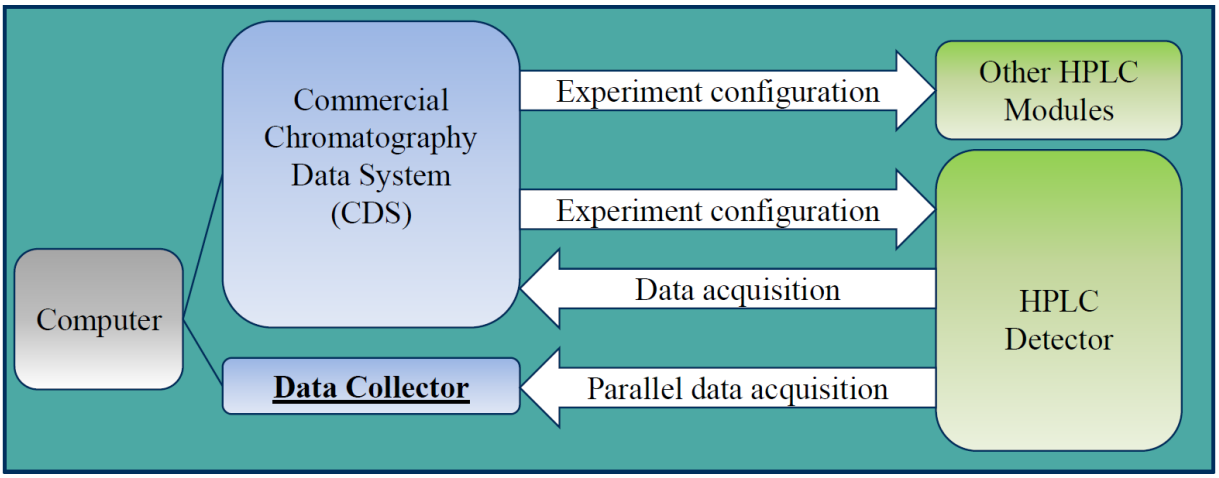

Figure 1. The *Data Collector* runs and acquires data parallel to the official CDS which communicate with the HPLC system and controls the experiment this way.

We used the presented *Data Collector* for repeated data acquisitions on two different HPLC systems. Both systems are manufactured by *Agilent Technologies*. In the further course of this article one system will be called the 1260 series system (G4225A Degasser, G1312B Binary Pump, G1367E Wellplate Autosampler, G1330B Autosampler Thermostat, G1316C Column Compartment and G4212B Diode Array Detector) and the other one the 1200 series system (G1322A Degasser, G1311A Quaternary Pump, G1329A Standard Autosampler, G1316A Column Compartment and G1315D Diode Array Detector) corresponding to the used modules.
An isocratic test sample containing four substances was used for each experiment due to an isocratic and isotherm configuration. The four substances of the sample are dimethyl phthalate, diethyl phthalate, biphenyl and o-terphenyl. These components were solved in methanol. For the separation of the components a mix of 35 vol.% HPLC-grade water and 65 vol. % Acetonitrile was used as mobile phase and a Zorbax xDB-C8 column supplied by *Agilent Technologies* was installed as stationary phase according to a reverse phase chromatography configuration. The column had a length of 50 mm, a diameter of 4.6 mm and a pore size of 1.8 µm.

The conditions of the chromatographic experiments were set up either by the commercial CDSs *OpenLab ChemStation*® (Rev. C.01.07 Build 27) developed and published by *Agilent Technologies* or by *Chromeleon*® (Rev. 6.80 SR15 Build 4656) developed and published by *Thermo Fischer*. The exact conditions have been chosen corresponding to the recommended specifications of the isocratic test sample: 1 ml/min flow, 1 µl injection volume, 40 °C column temperature and 254 nm detection wavelength. For every HPLC system and CDS several available detector sampling rates have been set running 10 repetitions per setup.

Results and Discussion

The simultaneous acquired data of each experiment can be used to compare the resulting raw data of the *Data Collector* and the official chromatography data systems (CDSs). Figure 2 shows overlaying chromatograms of *OpenLab ChemStation*® and *Data Collector* generated by the 1260 series system and using a sampling rate of 2.5 Hz. All four substances of the isocratic test sample are separated and are visible in the chromatograms. Even at the up- and downslope areas of the peaks there are no apparent deviations. A closer look on the exact raw data shows that the acquired signal values are complete equal except for different data accuracy given as the available number of decimal places. This behavior applies for all experiments done with the 1260 and 1200 series systems.

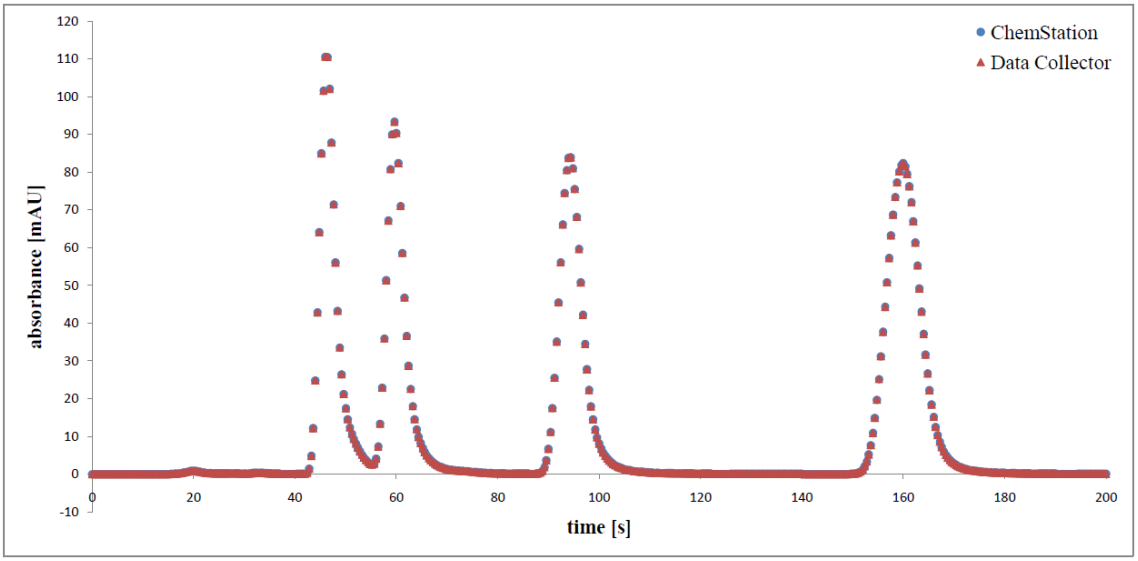

Figure 2. Overlaid chromatograms of *OpenLab ChemStation*® and *Data Collector* using a sampling rate of 2.5 Hz. Single data points are visible at the rising and falling areas of the peaks without apparent deviations.

A comparison of the raw data acquired by *Chromeleon*® and *Data Collector* resulted into the Fig. 3. The overlay of the signal generated by the 1260 series HPLC system using a sampling rate of 2.5 Hz can be seen here. In contrast to Fig. 2 apparent deviations between the raw data acquired by both software packages are visible at the up- and downslope areas of the peaks. In detail the signal acquired by *Data Collector* has a time delay compared to *Chromeleon*®.
A visualization of the signal differences is shown in Fig. 4. It represents another kind of plot for the signal values of both software packages at the time range from 80 to 120 seconds surrounding the third peak in the chromatograms of Fig.3. The signal values of *Chromeleon*® are plotted here against the exact signal differences between *Chromeleon*® and *Data Collector*. The specific formation of the data points in this plot describes the following behavior: When the peak is rising the signal difference increases until a maximum at the inflection point. Afterwards the difference reduces nearly to zero at the peak apex. For the falling area of the peak there is a negative difference reaching a maximum at the inflection point, too. That means the difference increases corresponding to the signal change in the chromatogram. A repetition of 10 runs using a sampling rate of 2.5 Hz

resulted into an average maximal deviation of 9.05±1.95 % (maximal difference value relative to the actual signal of *Chromeleon*®).

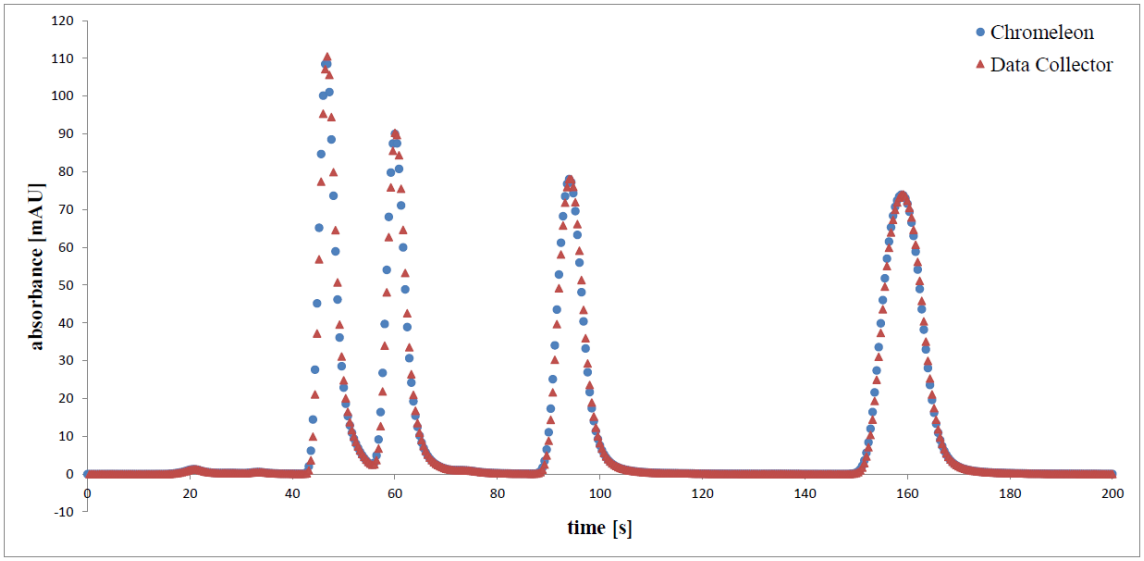

Figure 3. Overlaid chromatograms of *Chromeleon*® and *Data Collector* using a sampling rate of 2.5 Hz. A time delay of the data points at the rising and falling areas of the peaks is visible.

A second plot of acquired data using a higher sampling rate of 20 Hz has been overlaid in order to check for any dependencies (see Fig. 4). This ends up into another formation of the data points as the average maximal deviation decreased to approximate 1.85±0.43 % and the apex signal value increased. The second phenomenon describes the dependency of the raw data on the sampling rate and signal filtration as mentioned by Fahab et al. [9]. The decreased deviation between the signals shows that there is a real existing time delay between *Chromeleon*® and *Data Collector* but it is much lower as assumed when using a sampling rate of 2.5 Hz.

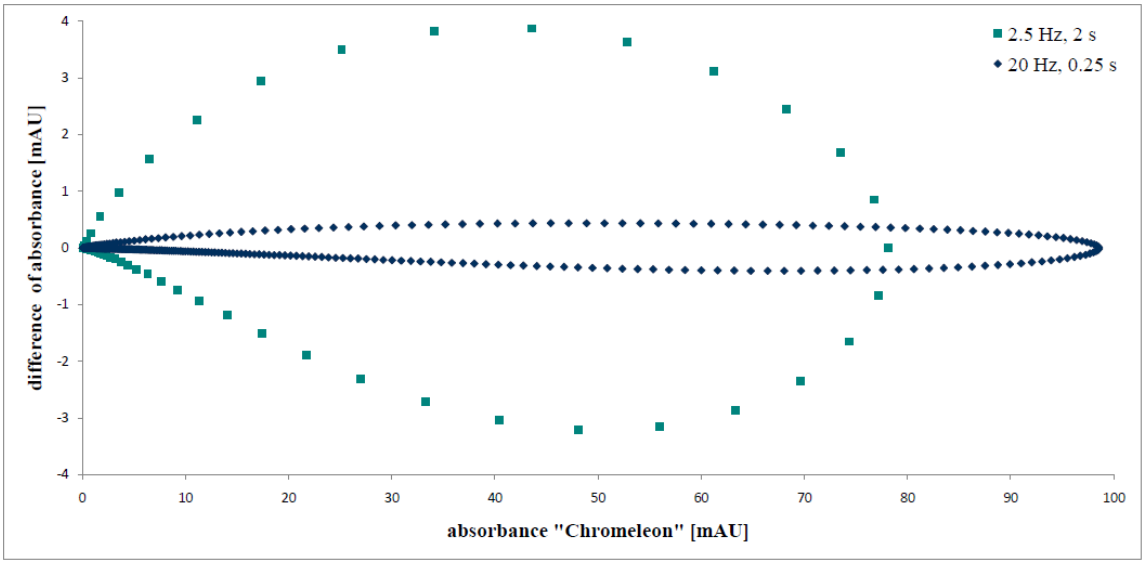

Figure 4. The raw data acquired by *Chromeleon*® of the time range from 80 to 120 seconds in Fig. 3 surrounding the third peak against the exact signal difference. Using a sampling rate of 2.5 Hz and response time of 2 s leads to a greater deflection at the rising and falling area of the peak than for 20 Hz and 0.25 s.

Summary

The parallel data acquisition using the new written *Data Collector* has shown that there are differences in the raw data between the tested CDSs *OpenLab ChemStation*® and *Chromeleon*®. These differences seem to be based on the underlying driver communicating with the HPLC system which will lead to a small time delay between the acquired data. This delay is marginal and has no significant influence on the raw data if a high sampling rate is used. This behavior has to be considered when chromatograms acquired by *OpenLab ChemStation*® and *Chromeleon*® will be compared.
Additionally, because of the fact that the source code of the *Data Collector* is known it has been proven that there are no manipulations on the raw data by the commercial CDSs before they store the raw data on the file system or in a database. Otherwise the deviations between *Data Collector* and the other CDSs would be more considerable.

Achnowledgements

We thank the AnaTox GmbH and Co. KG supporting this research by providing all required materials, chemicals and HPLC systems.

References

[1] The United States Pharmacopeial Convention, Convention Membership. Information on http://www.usp.org/about/convention-membership

[2] European Directorate for the Quality of Medicines, Membership and Observership, 2017. Information on https://www.edqm.eu/en/membership-observership

[3] Pharmaceuticals and Medical Devices Agency, Japanese Pharmacopoeia. Information on https://www.pmda.go.jp/english/rs-sb-std/standards-development/jp/0009.html

[4] The United States Pharmacopeial Convention, <621> Chromatography, in: The United States Pharmacopeia and National Formulary, First Supplement to USP 40-NF 35, 2017, pp. 1-12.

[5] European Directorate for the Quality of Medicines & HealhCare, 2.2.46. Chromatographic Separation Techniques, in: European Pharmacopeia, 5th ed, Strasbourg, 2005, pp 69-73.

[6] The Ministry of Health, Labour and Welfare, 2.01 Liquid Chromatography, in: The Japanese Pharmacopoeia, 17th ed, 2016, pp 41-44.

[7] Agilent Technologies, EQUIPMENT QUALIFICATION PLAN (EQP), 2014.

[8] Agilent Technologies, LICOP components for software development, 2016. Information on http://www.agilent.com/en-us/firmwareDownload?whid=56791

[9] M.F. Wahab, P.K. Dasgupta, A.F. Kadjo, and D.W. Armstrong, Sampling frequency, response times and embedded signal filtration in fast, high efficiency liquid chromatography: A tutorial, Analytica Chimica Acta 907 (2016), 31-44.