HHS Public Access

Author manuscript

*J Pharm Sci*. Author manuscript; available in PMC 2019 July 22.

Published in final edited form as:

*J Pharm Sci*. 2017 November ; 106(11): 3257–3269. doi:10.1016/j.xphs.2017.06.022.

# Chemical Stability of the Botanical Drug Substance Crofelemer: A Model System for Comparative Characterization of Complex Mixture Drugs

**Asha Hewarathna**[1], **Olivier Mozziconacci**[1], **Maulik K. Nariya**[2], **Peter A. Kleindl**[1], **Jian Xiong**[1,3], **Adam C. Fisher**[4], **Sangeeta B. Joshi**[1,3], **C. Russell Middaugh**[1,3], **M. Laird Forrest**[1], **David B. Volkin**[1,3], **Eric J. Deeds**[5,6,7], **Christian Schöneich**[1,*]

[1]Department of Pharmaceutical Chemistry, University of Kansas, Lawrence, Kansas 66047

[2]Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66047

[3]Macromolecule and Vaccine Stabilization Center, University of Kansas, Lawrence, Kansas 66047

[4]Center for Drug Evaluation and Research, Office of Pharmaceutical Quality, U.S. Food and Drug Administration, Silver Spring, Maryland 20993

[5]Center for Computational Biology, University of Kansas, Lawrence, Kansas 66047

[6]Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047

[7]Santa Fe Institute, Santa Fe, New Mexico 87501

## Abstract

As the second of a 3-part series of articles in this issue concerning the development of a mathematical model for comparative characterization of complex mixture drugs using crofelemer (CF) as a model compound, this work focuses on the evaluation of the chemical stability profile of CF. CF is a biopolymer containing a mixture of proanthocyanidin oligomers which are primarily composed of gallocatechin with a small contribution from catechin. CF extracted from drug product was subjected to molecular weight—based fractionation and thiolysis. Temperature stress and metal-catalyzed oxidation were selected for accelerated and forced degradation studies. Stressed CF samples were size fractionated, thiolyzed, and analyzed with a combination of negative-ion electrospray ionization mass spectrometry (ESI-MS) and reversed-phase-HPLC with UV absorption and fluorescence detection. We further analyzed the chemical stability data sets for various CF samples generated from reversed-phase-HPLC-UV and ESI-MS using data-mining and machine learning approaches. In particular, calculations based on mutual information of over 800,000 data points in the ESI-MS analytical data set revealed specific CF cleavage and degradation products that were differentially generated under specific storage/degradation conditions, which were not initially identified using traditional analysis of the ESI-MS results.

*Correspondence to: Christian Schöoneich (Telephone: (785) 864-4880; Fax: (785) 864-5736). schoneic@ku.edu (C. Schöoneich).

This article contains supplementary material available from the authors by request or via the Internet at http://dx.doi.org/10.1016Zj.xphs.2017.06.022.

**Keywords**

crofelemer; complex mixture; chemical stability; oxidation; mass spectrometry; HPLC; mutual information scores; machine learning

## Introduction

Crofelemer (CF) is a naturally occurring botanical biopolymer extracted from the stem bark latex of the *Croton lechleri* tree.[1] It is an FDA-approved drug for the treatment of noninfectious diarrhea in adult patients with HIV/AIDS undergoing antiretroviral therapies and acts by inhibiting 2 chloride channels located in the gastrointestinal tract.[2] CF is a mixture of proanthocyanidin oligomers, also known as condensed tannins, with an average degree of polymerization ranging from 7, with an average molecular weight (MW) of 2300 Da.[3] The main constituents of CF are (+)-gallocatechin (GC) and (−)-epigallocatechin (epiGC), with a minor contribution from (+)-catechin (C) and (−)-epicatechin (epiC).[4] A representative structure of CF that contains B-type interflavan linkages is displayed in Figure 1a.

The C and GC components of CF are flavan-3-ol monomer units which are coupled through either A-type or B-type linkages to form oligomers.[5] The linkages through C4-C8 or C4-C6 carbon-carbon bonds are referred to as B-type linkages, whereas linkages involving either one of C4-C8 or C4-C6 carbon-carbon bond, and an additional ether bond between C2 and O7 or C2 and O5 positions, are referred to as A-type linkages.[6] The 2 asymmetric carbons at positions 2 and 3 result in 4 diastereomers per individual monomer. Both C and GC are *trans* in configuration, whereas epiC and epiGC are *cis* in configuration. Proanthocyanidins are characterized by 2 aromatic rings (rings A and B), which are linked by an O-hetero-cycle (ring C). The hydroxylation pattern on the B-ring distinguishes different monomers. Proanthocyanidins that are composed exclusively of 3′,4′-dihydroxy substitution on the B-ring are called procyanidins. Prodelphinidins contain exclusively GC with a 3′,4′,5′-trihydroxy substituted B-ring. In addition, C and GC are linked by C-C bonds through C4 and C8 or C4 and C6 to form B-type proanthocyanidins that sometimes undergo esterification with other groups such as glucose or gallic acid at C3.[7] A-type proanthocyanidins are relatively rare and are formed by an additional ether linkage between C2 and O7 or C2 and O5, imposing conformational stability.[8,9] The 3 asymmetric carbons at positions 2, 3, and 4 result in different epimers. Collectively, this structural diversity results in CF being a highly heterogeneous mixture of proanthocyanidin oligomers.

Proanthocyanidins are naturally occurring secondary plant metabolites that are prevalent in bark, nuts, seeds, fruits, vegetables, and flowers.[10] They are strong antioxidants[11] that react with different types of reactive oxygen species such as singlet oxygen, superoxide radicals, and hydroxyl radicals.[12] Both GC and galloylated-GC are more efficient radical scavengers than C.[13,14] Biological properties of proanthocyanidins are related to their protein-binding capability, metal ion chelation, and antioxidant capacity, which are determined by the structure and degree of polymerization.[10] Antioxidant properties can be evaluated through a variety of chemical and biochemical methods,[14–17] including the ability to scavenge free

radicals and transfer hydrogen atoms or electrons. In addition, the complexation of metal ions can influence and, possibly, prevent metal-catalyzed redox reactions.[18–20]

The abundance of phenolic moieties makes CF highly susceptible to oxidation, not unexpected based on the known antioxidant potential of proanthocyanidins. Moreover, oxidation would likely affect multiple monomeric units of CF, potentially leading to a large variety of oxidation products. This is part of a series of 3 papers in this issue focused on developing an integrated mathematical model for comparative characterization of complex mixture drugs using CF as a model system. Here, we evaluated the chemical stability profile of CF (1) during storage at 25°C and 40°C and (2) via forced degradation studies by metal-catalyzed oxidation. The initial physicochemical and biological characterization of CF by a wide variety of analytical methods is described in the preceding paper.[21] In the present paper, for the analysis of the chemical stability of CF, we used reversed-phase HPLC coupled with UV absorption spectroscopy, fluorescence detection, and electrospray ionization mass spectrometry (ESI-MS), including MS/MS. These analytical data sets were used to calculate mutual information scores (MISs) to correlate specific degradation products with specific storage/degradation conditions. The combination of the CF chemical stability data sets from this work and the physicochemical and biological characterization data of CF in the preceding paper is used in the third paper of this series for data-mining and machine learning approaches to develop an integrated mathematical model for comparative characterization of complex mixture drugs, using CF as a model system.

## Materials

The compounds C, epiC, GC, epiGC, and β-mercaptoethanol (β-ME) were purchased from Sigma-Aldrich (St. Louis, MO). CF was extracted from commercially available Fulyzaq® tablets (Salix Pharmaceuticals, Raleigh, NC), and the extraction procedure is explained in detail in the preceding paper in this issue.[21]

## Methods

### Membrane Fractionation

Membrane fractionation of the CF extract is described in detail in the preceding paper in this issue.[21] Briefly, Amicon Ultra-0.5 mL centrifugal filters (Millipore, Billerica, MA) of 3-and 10-kDa molecular weight cutoff (MWCO) were used to fractionate CF. A sample of CF was added to an Amicon filter and centrifuged at 4000× g for 30 min at 4°C. A reddish-brown colored fraction (CF-Top) was retained in the insert and a colorless fraction (CF-Bot) was collected as the filtrate. Although the yields of the bottom fractions were low (for more discussion, please see the preceding paper in this issue[21]), this method was the only fractionation technique that enabled us to separate nonoxidized from oxidized CF, initially present in the extract from Fulyzaq tablets, as described in the following.

### Thermal Stability

Individually sealed aliquots (3-mL type-1 glass vials; West Pharmaceutical Services, Exton, PA), stoppered with rubber stoppers (West Pharmaceutical Services), crimped of 25 mg/mL unfractionated CF, 25 mg/mL CF-Top, and 0.5 mg/mL CF-Bot were prepared in ultrapure

water and incubated at 25 or 40° C for 1 month for accelerated stability studies. Three and 4 sample replicates were analyzed for MS and HPLC analyses, respectively.

### Thiolysis of CF

Thiolysis was performed according to a previously published protocol for procyanidins from selected foods.[22] A 50-μL sample of 6 mg/mL CF was thiolyzed in 1.2-mL low adsorption clear glass vials sealed with polytetrafluoroethylene/silicone septa (Waters, Milford, MA), using 0.6 N HCl and 0.7 M β-ME at 95° C for 5 min, followed by vacuum concentration for 5 min. For control experiments, 50 μL of 2 mg/mL C, epiC, GC, and epiGC standard solutions were thiolyzed separately, using the same protocol.

### Metal-Catalyzed Oxidation

To generate oxidized CF via a complementary oxidation reaction, CF, CF-Top, and CF-Bot samples were exposed to $CuCl_2$ and ascorbic acid/ascorbate at 37°C for 3 days. Mixtures of CF, CF-Top, or CF-Bot with $CuCl_2$ and ascorbic acid/ascorbate were prepared in ultrapure or deionized water. The pH values of these reaction mixtures were measured, but not controlled, that is, they were left at the values which were naturally adjusted by combination of the components of the mixtures. The ratio of CF:ascorbic acid/ascorbate:Cu(II) was kept at 50:10:1 between CF, CF-Bot, and CF-Top fractions. In CF and CF-Top, 10-mM CF (based on an average molecular weight of 2000 g/mol[4,23]) was treated with 2-mM ascorbic acid/ascorbate and 0.2 mM $CuCl_2$ at pH 3.90. Instead, the relatively less concentrated CF-Bot (0.1 mM) was reacted with 0.02-mM ascorbate and 0.002-mM $CuCl_2$ (pH 6.80). In a separate experiment, metal-catalyzed oxidation (MCO) was carried out by incubating 50 μL of 10-mM CF with 1.4-mM $CuCl_2$ and 14-mM ascorbic acid/ascorbate (pH 3.25) at 45° C overnight, and the reaction mixture analyzed by reversed-phase (RP)-HPLC.

### RP-HPLC Analysis

A C18 column (Vydac 218TP, 250 × 4.6 mm, 5 m) was preheated to 40° C and equilibrated with a mixture of 95% mobile phase A (0.1% formic acid and 0.5% 2-propanol in water) and 5% mobile phase B (70% methanol, 30% water, 0.1% formic acid, and 0.5% 2-propanol). The solutes were eluted by a linear gradient changing mobile phase B in the following way: 5%−30% within 55 min, followed by 30%−90% within 5 min at a flow rate of 0.8 mL/min. The solutes were monitored with UV detection at 280 nm and fluorescence detection ($\lambda_{ex}$ = 276 nm, $\lambda_{em}$ = 316 nm).

### HPLC-MS Analysis

The CF samples were analyzed by a Micromass Q-Tof Premier mass spectrometer (Micromass Ltd., Manchester, UK) connected to an Acquity UPLC system (Waters). The instrument was operated in the negative-ion ESI mode and the following instrument parameters were used: capillary voltage, 2.8 kV; desolvation temperature, 250°C. The samples were injected onto a Agilent ZORBAX (+) phenyl-hexyl column (3 × 50 mm, 1.8 μ), and the solutes were eluted with a linear gradient changing mobile phases A (0.1% formic acid and 0.5% 2-propanol in water) and B (70% methanol, 30% water, 0.1% formic

acid and 0.5% 2-propanol) as follows: 15%−50% B within 6.5 min followed by 50%−90% B within 3 min, at the flow rate of 0.3 mL/min.

### Mathematical Modeling

As outlined previously, the mass spectrometry data set nominally consists of 35 distinct samples: 5 fractions × 4 stability times × 2 temperatures for each stability time (except day 0). Each of these samples had several replicates, but the 10-kDa and 3-kDa bottom fractions had too few replicates; a total of 21 separate sample types were thus available for this analysis. For each retention time and *m/z* value, we constructed a column vector consisting of the ion flux counts corresponding to all replicates from all of these samples. We created a 2-D histogram of the joint probability distribution *p(x,y),* such that the *y*-bins are the 21 categorical variables (i.e., distinct samples) and the *x*-bins estimate the probability of a particular range of ion fluxes. Once we had the joint probability distribution, we calculated the MIS using the standard definition[24]:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

where $X$ and $Y$ are the sets of possible $x$ and $y$ bins, and $p(x)$ and $p(y)$ are the corresponding marginal probability distributions. We used 6 bins for the ion fluxes ($x$) for Supplementary Figure S5, and although the choice of bin number in this case is essentially arbitrary, we observed a strong correlation between MIS calculated using 4, 6, 8, and 10 bins in ion flux count (Spearman's rank correlation coefficient close to 0.9 in all cases). The choice of bin number for the ion flux variable thus had little effect on our result. All calculations were performed using the function MIS from sklearn.metric in Python. We used the base 2 logarithm for this calculation so that the MISs in all figures have units of bits.

## Results

### Sample Preparation

The extraction of CF from commercially available tablets into an aqueous solution has been described in the preceding paper.[21] After removal of the enteric coating, the Fulyzaq tablets as well as an aqueous extract exhibited a reddish-brown color. Because CF is a mixture of proanthocyanidin oligomers, it is expected to be colorless. Hence, the reddish-brown coloration might be an indication of the presence of oxidized proanthocyanidins as constituents of CF. On account of the complexity of the procyanidin mixture, we tested a membrane filtration method to further refine the mixture. Filtration by a 10-kDa MWCO filter separated the reddish-brown CF into 2 fractions, a colorless (CF-Bot) and a reddish-brown fraction (CF-Top) (Supplementary Fig. S1). The reddish-brown fraction presumably contains a substantial portion of polymer-like molecules which cannot pass through the 10-kDa MWCO membrane, whereas low-MW molecules, monomers, and proanthocyanidin oligomers may pass through the membrane. The compound that is responsible for the reddish-brown color of CF appears to have a higher MW than 10 kDa. Similarly, filtration of CF by a 3-kDa MWCO filter resulted in a reddish-brown fraction which was retained and a colorless fraction which was collected as the filtrate.

## Thiolysis

Thiolysis has been used for the analysis of proanthocyanidin polymers to measure the degree of polymerization and the composition of terminal and extension units.[25] In this procedure, the terminal units (see Supplementary Figure S7, Supplementary Information) are released as flavanol monomers while the extension units add a thiol to the 4-position.[26] For our experiment, β-ME was selected as the thiol. The presumed mechanism for thiolysis of a representative dimeric proanthocyanidin using β-ME is also displayed in Supplementary Figure S7. The thiolysis products of CF were further separated and analyzed by RP-HPLC and ESI-MS analysis.

## Thermal Stability of CF

Samples of 25 mg/mL CF, 25 mg/mL CF-Top, and 0.5 mg/mL CF-Bot were prepared in ultrapure water and incubated at either 25 or 40° C for 2, 7, and 30 days to evaluate the chemical stability of proanthocyanidins. Subsequently, the samples were thiolyzed and analyzed by RP-HPLC with UV detection at 280 nm and HPLC ESI-MS as described in the following.

## RP-HPLC-UV Analysis

Chromatograms of CF samples, incubated at 40° C for 0 days (control) and 30 days, subjected to thiolysis, are shown in Figure 2a. Thiolysis produces β-ME-derivatized proanthocyanidin and free flavanol molecules (see Supplementary Figure S7, Supplementary Information). Qualitatively, the RP-HPLC profile of products generated from the sample that was incubated for 30 days at 40° C is similar to that of the control (0 days). Peaks eluting at 7.5,14.9,15.5, and 27 min were assigned to GC, C, epiGC, and epiC by comparison to authentic standards. However, after 30 days of incubation, the amounts of GC, C, epiGC, and epiC recovered by thiolysis decreased compared to the amount recovered from control samples, suggesting a chemical transformation of CF. Importantly, the peaks with retention times (RTs) = 14.1, 25.3, 29, and 42.5 min also showed a time-dependent reduction of peak area (Fig. 2b). These peaks were collected for analysis by LC-MS.

We additionally monitored the changes in terminal monomer (substituted at C6 or C8 to the successive monomer unit in the original oligomer chain) composition of CF samples that were incubated at different temperatures over time (Fig. 3). Thiolysis of a control sample revealed that CF contains 30.8 ± 3.1% GC, 34.2 ± 5.0% epiGC, 15.1 ± 2.7% C, and 20.0 ± 1.6% epiC as terminal units. Figures 3a and 3b display the changes of terminal C, GC, epiC, and epiGC in CF samples that were incubated for up to 30 days at 25 and 40°C, respectively. The amount of each monomer decreases over time under both conditions; noteworthy, the epiGC content showed the most rapid decline during incubation at 40° C.

CF-Top fractions obtained after membrane filtration of CF (CF-Top) with a 10-kDa MWCO filter were incubated at 25°C and 40°C for 2, 7, and 30 days. The data are displayed in Figures 3c and 3d. Control samples of CF-Top contained 27.1 ± 3.8% GC, 36.1 ± 7.6% epiGC, 15.4 ± 3.5% C, and 21.0 ± 4.7% epiC. As observed for unfractionated CF, epiGC showed the most rapid temperature-dependent decline at 40°C (Fig. 3d).

## HPLC ESI-MS Analysis of Collected UV Peaks

Fragmentation reactions of proanthocyanidin ions in the gas phase are well documented.[27,28] The fragmentation mechanisms reported for polyphenolic compounds,[27,28] including proanthocyanidins, can be used to explain the fragment ions observed for ESI-MS

analysis of CF. Here, 3 main mechanisms are described, referred to as retro-Diels Alder reaction, heterocyclic ring fission, and quinone methide (QM) fission; these were reported for a representative B-type proanthocyanidin containing 2 GC units and are illustrated in Supplementary Figure S8, Supplementary Information. The hitherto unidentified peaks eluting at 14.1 and 42.5 min in the RP-HPLC chromatograms (Fig. 2a) were collected multiple times, and the fractions were pooled, and lyophilized, before analyses by negative-ion ESI-MS. The MS analysis of the peak eluting at 42.5 min revealed 3 major peaks with m/z 1097.18, 1307.22, and 917.17 in the total ion chromatogram (TIC). The $MS^2$ spectra of those peaks are shown in Figures 4a–4c.

### Peak 1 (m/z 1097.18)

The ion with m/z 1097.18 can be rationalized as a procyanidin trimer, composed only of C units (Fig. 4c). Table 1 shows the ions and fragment ions of this compound together with their tentative structures. After thiolysis, each of the 3 C units is derivatized with 1 equivalent of β-mercaptoethanol. QM cleavage (Supplementary Figure S8, Supplementary Information) of the top unit results in the fragments with m/z 731.10 and 365.09. Ring closure of ring-C and the loss of 1 β-ME generates the dimeric ion with m/z 653.12. QM cleavage of this dimer forms the ion with m/z 287.06 from the upper C unit.

### Peak 2 (m/z 1307.22)

The product with m/z 1307.22 can be identified as a tetrameric structure, which contains 2 B-type linkages and 1 A-type linkage, and is derivatized with 2 β-ME molecules (Fig. 4b). Ions and fragment ions along with their tentative structures are listed in Supplementary Table S1. The fragmentation of the tetramer at the second B-type linkage from the top (indicated by an arrow in Supplementary Table S1) yields fragment ions with m/z 575.13 and m/z 653.13 through the QM reaction (Supplementary Figure S8, Supplementary Information). The ion at m/z 653.13 originates from QM cleavage followed by C-ring closure, with the elimination of 1 β-ME. It is also possible that subsequent C-ring closure with thiol elimination yields an ion with m/z 575.13, which subsequently converts into ions with m/z 125.02 and m/z 449.10 by heterocyclic ring fission (Supplementary Figure S8, Supplementary Information) of the upper C unit. The fragment with m/z 287.08 is generated by QM of the dimer (m/z 575.13).

### Peak 3 (m/z 917.17)

The spectrum in Figure 4a shows an important difference to the spectra of Figures 4b and 4c. Here, a prominent fragment ion in the low m/z region is observed with m/z 153.02 (instead of 125.02), suggesting the presence of a galloyl moiety. The molecular ion with m/z 917.17 would correspond to a GC dimer esterified with 2 galloyl groups (Supplementary Table S2). Loss of 2 gallic acid moieties (−152 Da) from the molecular ion leads to a dimer with m/z 611.14, which then undergoes the QM reaction to form GC (m/z 305.09). QM cleavage of

the molecular ion yields the fragment with m/z 459.11. The presence of galloylated species adds heterogeneity to the proanthocyanidin composition of CF.

MS analysis of the peak eluting at 14.1 min resulted in several signals (m/z 593.04, 625.03, 685.13, 696.04, 883.10, and 1145.18), and the most intense peak with m/z 1145.18 corresponds to a GC trimer derivatized with 3 β-ME units. The peaks eluting at 25.3 min and 29 min did not generate reliable data in the MS analysis, most likely due to their low yield.

## Thermal Stability of the Bottom Fraction of CF (CF-Bot)

Samples of 0.5 mg/mL CF-Bot prepared in ultrapure water were incubated for up to 30 days at 40° C. The samples were not thiolyzed on account of the low concentration recovered from the bottom fractions. However, CF-Bot samples were separated by RP-HPLC and monitored with UV and florescence detection at $\lambda_{ex}$ = 276 and $\lambda_{em}$ = 316 (Fig. 5a). When fluorescence detection was used, the control CF-Bot samples (0 days) exhibited a baseline shift, which was not observed for the samples exposed to 40° C for up to 30 days. This shift in the baseline may be related to the presence of polymer-like molecules which are degraded by prolonged exposure to 40° C. Similar to CF and CF-Top, a decrease in peak areas was observed for the CF-Bot fractions.

To identify oxidation products, if present, chromatographic profiles of temperature-stressed samples were monitored at λ = 340 nm (Fig. 5b). We observed a significant increase of a peak eluting at 6.8 min after 30 days of incubation at 40° C. On account of the low abundance, the identity of the product could not be verified by $MS^2$ analysis.

## MCO of CF

CF, fractions of CF, CuCl2, and ascorbate samples were prepared in ultrapure water or deionized water. MCO of CF, CF-Top, and CF-Bot was carried out with Cu(ll)/ascorbate at 37°C for 3 days. CF (10 mM) and CF-Top (10 mM) samples were treated with 0.2-mM $CuCl_2$/2 mM ascorbate, and CF-Bot (0.1 mM) samples were treated with 2 μM $CuCl_2$/0.02 mM ascorbate (both 1:10 molar ratio). After 3 days, the formation of insoluble material was observed in both CF and CF-Top fractions, and the initially, colorless CF-Bot fractions turned brown (see Supplementary Fig. S2).

For RP-HPLC analysis, a separate sample of 10 mM CF (based on an average molecular weight of 2000 g/mol[4,23]) was reacted with 1.4-mM Cu(ll)/14-mM ascorbate at 45°C overnight, and a control was similarly incubated in the absence of Cu(ll)/ascorbate. The sample that was subjected to MCO and a control were thiolyzed to enable analysis by RP-HPLC (see Supplementary Fig. S3 which compares a CF sample exposed to MCO to a control). Analogous to the temperature-stressed CF samples, MCO-exposed CF samples display a decrease in thiolytic products, monomers, and adducts.

## Data Analysis

As explained in the Methods section, CF samples were fractionated using 3-kDa and 10-kDa MWCO filters, and the unfractionated CF and its 4 fractions (3 kDa top and bottom and 10

kDa top and bottom) were incubated at either 25° Cor 40° C for the specified amount of time. Incubated samples (with 3 replicates per sample) of CF and the 4 fractions were then analyzed by RP-HPLC ESI-MS. Representative TIC and mass spectra for a sample of the 3 kDa top fraction, which was incubated at 25° C for 30 days followed by thiolysis, are shown in Supplementary Figure S4. The TIC contained 4 major peaks (Supplementary Fig. S4, top) and the mass spectra obtained by integrating the most intense 3 peaks on TIC are displayed in Supplementary Figure S4, bottom. The peaks 1 and 2 contained ions ranging from m/z 125.02 to 1145.18, whereas peak 3 contained ions ranging from m/z 125.01 to 1097.19. These MS data were used in the development of the mathematical model.

We next investigated whether there might be particular CF oxidation products or other detectable chemical changes specific to a particular CF fraction, incubation temperature, or stability time. One major difficulty in answering this question, however, is the sheer volume of the data: there are over 800,000 unique (RT, *m/z)* pairs for each individual sample. In addition, the “major” peaks (i.e., those with the highest ion counts) show very little difference in intensities between each of the samples, and as a result, it is essentially impossible to use visual inspection to find data points that are significantly different between the various treatments. We thus used an MIS analysis to identify differences in these data.

The MIS is a measure of the statistical correlation between random variables[24]; the higher the value of the MIS, the more information the ion flux at that particular (RT, *m/z)* value can tell you about the nature of the sample itself (see Methods for further details on this calculation). We calculated the MIS between the ion flux for every (RT, *m/z)* pair and the particular label of the samples (e.g., unfractionated, 25° C, 7 days). Supplementary Figure S5 shows the heat map of the MIS for all (RT, *m/z*) values. The region of highest overall information is for retention times between 9 and 11 min and *m/z* values between 700 and 1100, corresponding to fairly hydrophobic chemical species ranging in size from monomers to trimers. As is clear from Supplementary Figure S5, however, there are other points of high information content scattered throughout the data set, further highlighting the difficulty of finding these particular points by eye.

We ranked the (RT, *m/z)* pairs by MIS to better understand which chemical species are significantly different between the samples; Supplementary Figure S6 shows the MIS values for the top 100 points. To find the species that varied the most between samples, we chose the top 6 (RT, *m/z)* points according to MIS and plotted the ion flux for those points as a function of the stability time for each of the samples and temperatures (Fig. 6). As one can see, the abundance of these particular chemical species changes in a significantly different manner depending on the particular fraction and temperature in question. One consistent observation for all of these species is that they tend to have a somewhat higher concentration in the CF samples incubated at 25° C compared to 40° C. This can be rationalized with the oxidation sensitivity of the CF subunits, which degrade faster at 40°C compared to 25°C (representatively shown for monomers in Fig. 3). In addition, note that the ion flux counts in these cases are fairly low, indicating that the significant differences between samples in this case derive from (RT, *m/z)* pairs that would not generally be considered major peaks and thus could easily be missed in a standard analysis of mass spectrometry data. Although all of

these (RT, *m/z)* values vary significantly between the samples, we focused our further analyses on the (2.69 min, *m/z* 607.12) point because this is a nonmonomeric species.

The tentative molecular structures of the m/z values highlighted by MIS analysis are presented in Supplementary Table S3. We obtained the RT and m/z of components which change over time across different parameters (e.g., fractions, time, temperature) and extracted the mass spectra that correspond to these RTs. This includes a GC dimer (m/z 607.12) and its fragments. Here, we focus on the m/z 607.12 ion because other ions appeared to be fragments of the GC dimer, and the presence of this ion was confirmed under different conditions in different samples. For example, the mass spectrum extracted from a CF sample that was incubated at 25° C for 30 days is shown in Figure 7a. The formation of this ion could be correlated with the ions with high m/z, and a potential mechanism of formation is shown in Figure 7b. The ion with m/z 763.14 could be present in the solution, and fragmentation in the gas phase could generate m/z 685.13 and 607.12. Fragmentation of the ion during MS1 analysis might be due to fragmentation in the ion source. It is also possible that there is an equilibrium between these 3 ions in the solution.

Using our MIS analysis, we thus discovered a potential component of CF that changed over the degradation study. This component displays an ion with m/z 607.12 and tentatively corresponds to an oxidized GC dimer. The ion with m/z of 607.12 has an abundance of 5.20% (integration of ion signal) relative to the ion with m/z 305.09 of the gallocatechin monomer, that is, of a comparable peak in a nonstressed CF sample. Given the volume of the data, it is highly unlikely that a standard analysis would have discovered this particular species.

## Discussion

CF is a highly heterogeneous mixture of oligomers, and therefore, MW-based membrane filtration and thiolysis were introduced to reduce the complexity of CF. Fractionation with a 10-kDa MWCO separated reddish-brown CF into a colorless low-MW and reddish-brown high-MW fraction. A similar fractionation was observed when 3-kDa MWCO filters were used. The reddish-brown fraction appears to contain high-MW polymers. One reaction pathway leading to the formation of such polymers might be initiated by proanthocyanidin oxidation, followed by a series of secondary reactions to form high-MW species. However, it is also possible that the retention by the membrane is either due to hydrophobic interactions and hydrogen bonding between proanthocyanidin molecules and the membrane (regenerated cellulose) or through formation of higher molecular weight species through selfassociation. Spontaneous interactions between polyphenols and polysaccharides have been reported.[29]

Chromatographic separation challenges were encountered during our analysis because the heterogeneous CF eluted as a broad, unresolved peak despite the chromatographic method. Similar chromatographic separation challenges of proanthocyanidins were reported elsewhere.[10,30] Unlike oligomeric procyanidins, polymeric-procyanidins with a high degree of polymerization were difficult to separate by reversed-phase, normal-phase, and size-exclusion chromatography.[31,32] Thus, alternative methods, such as hydrophilic interaction liquid chromatography, have been introduced to analyze procyanidins.[33] One successful

strategy to address the complexity of proanthocyanidins would be to combine thiolysis or another fractionation method with routine HPLC and MS methods to achieve satisfactory resolution as evidenced in this work.

In this analysis, CF was thiolyzed to obtain a good chromatographic separation. Thiolysis uses degradation of flavanols by a mercaptan in acidic medium.[26,34] Interflavanoid bond cleavage releases free monomers and forms an intermediary carbocation at the C4 position (Supplementary Figure S7, reaction 2; Supplementary Information), which subsequently reacts with a nucleophile to form flavanol adducts (Supplementary Figure S7, reaction 3; Supplementary Information). It should be noted that the intermediary C4 carbocation can deprotonate to yield a quinone methide, which will equally react with a nucleophile. Singly derivatized flavanol adducts can further react to form double-derivatized flavanol ad-ducts (Supplementary Figure S7, reactions 4 and 5; Supplementary Information). B-type linkages are cleaved by thiolysis, whereas A-type linkages and O-gallate substitutions at the C3 are resistant to thiolysis.[35,36] Heat-induced epimerization of C to epiC, and epiC to C is known.[37] It is important to note that released free flavanol monomers may undergo epimerization under thiolysis reaction conditions.[21]

The temperature stability of CF was assessed at 25° C and 40° C. We observed that thiolytic products, free flavanols, and β-ME-derivatized extension units were decreased significantly at 40° C over time, where epiGC was much more sensitive to temperature in comparison to the other isomers. This behavior is consistent with reports of the decrease of flavanols and polyphenol levels in plant foods as a result of thermal processing and drying.[38–42]

After temperature stress, the RP-HPLC peaks of interest were collected and analyzed by ESI-MS in the negative-ion mode, which permits detection of proanthocyanidins as deprotonated ions, to identify the constituents through their fragmentation patterns from $MS^2$ experiments. As expected, $MS^2$ analysis reveals that CF contains combinations of C and GC, resulting in a highly heterogeneous mixture of procyanidins and prodelphinidins. Interestingly, a dimeric prodelphinidin that is esterified with 2 gallic acid molecules was identified in temperature-stressed, thiolyzed CF samples. This gallocatechin gallate dimer is a possible constituent in proanthocyanidins, but not in CF. We did not analyze the similar fraction in a control sample. However, MS analysis of thiolyzed, temperature-stressed, and thiolyzed, control samples (without collecting individual peaks) revealed that all peaks contained similar fragment ions in both samples. Therefore, this gallocatechin gallate dimer should be present in control samples.

In addition, those samples contained a procyanidin dimer with an A-type linkage, which is an unexpected constituent in CF. A-type proanthocyanidins are resistant to thiolysis and therefore are expected to be released as free A-type dimers by thiolysis if present in terminal units or as β-ME derivatives if present in extension units. Here, the A-type procyanidin must be located in the terminal units because it is detected as a free dimer. A-type proanthocyanidins are yellow in color and believed to be generated by the oxidation of B-type proanthocyanidins, and it is known that the oxidation proceeds via a quinone methide mechanism.[43,44] Although we did not analyze the similar fragment of control samples, as explained previously, it is possible that A-type dimers are present in control samples. The

presence of oxidized species may account for the reddish-brown coloration in extracted CF samples.

CF contains dihydroxyl (C)- or trihydroxyl (GC)-substituted B-rings, which are susceptible to oxidation. Supplementary Figure S9 shows the formation of o-quinone by C oxidation.[45] O-quinones are highly reactive, and the nucleophilic addition of electron-rich A-rings to electron-deficient B-rings leads to conjugate formation.[46] Further conjugation leads to the formation of insoluble material such as tannins. Polyphenol oxidation is often accompanied by color changes and conjugate formation.

The reddish-brown coloration of CF cannot be attributed to structural and spectral properties of pure proanthocyanidin moieties; however, it appears to be a product of oxidation (e.g., a quinone). Oxidation-induced color changes have been associated with flavanols.[47] Dilute acid–treated proanthocyanidins are also reported to have characteristic red color.[48] Alternatively, the coloration of CF could be related to interactions between flavanol molecules and metal ions. The catechol ring chelates with metal ions to cause a red shift in the spectrum, thereby absorbing at higher wavelengths. Brown et al. reported that the 3′, 4′-hydroxyl groups on the flavonoids are important for the chelation of $Cu^{2+}$.[49] The complex formation of flavanols with metal ions and subsequent bathochromic shifts in spectral properties have been documented.[50]

The reddish-brown color and insoluble material formation during accelerated stability studies of CF was most likely due to nonenzymatic oxidation reactions. Conjugation yields insoluble yellow-brown tannins.[51] It is known that oxidation of proanthocyanidins form polymers and insoluble reddish-brown pigments, called phlobaphenes.[52] Nonenzymatic oxidation of proanthocyanidins proceeds via reaction with oxygen or ROS.[53] Mochizuki et al. reported that autoxidation of C initiates by 1-electron oxidation to generate semiquinone and superoxide radical anion.[54] Analysis of oxidation products of (+)-catechin model solutions revealed that different products are formed depending on the oxidation pathway.[55] For example, metal-catalyzed oxidation products showed different UV-Vis spectral properties in comparison to enzymatic oxidation or autoxidation conditions.

The large volume of data generated by MS analysis of CF and its fractions under different temperature conditions necessitated the use of data-mining approaches to better analyze the large data set. Similar to many modern measurement techniques, a major difficulty encountered with mass spectrometry is the sheer volume of data that it produces. In the case of our analysis of CF, each replicate of our 35 different samples results in flux counts for over 800,000 unique (RT, *m/z)* pairs, and most of these do not differ significantly between the samples. This is particularly true for those points with higher ion flux counts. In other words, the major peaks are more consistent between samples and thus less useful as a means of differentiating between chemical compositions.

Here, we used a straightforward application of information theory (namely the mutual information score or MIS) to overcome this problem. Our analysis revealed that points with the capacity to differentiate between various samples were scattered throughout the (RT, *m/z)* space. Although retention times between 9 and 11 min and *m/z* values between 700 and

1100 contained the highest information on average, the top 6 points by MIS were all outside of this region (Fig. 6, and Supplementary Figs. S5 and S6). This analysis allowed us to assign peaks and identify chemical species that differed significantly between samples, storage temperatures, and times, ultimately providing further insight into the oxidation processes occurring within these samples. Because all of these species correspond to relatively minor peaks, with low total ion flux counts, they could easily have been missed in a more traditional analysis of mass spectrometry data. For example, the peaks eluting at 2.46 and 2.69 min have abundance of 6.24% and 7.49%, respectively (integration of TIC), relative to one of the most intense peaks with m/z 1145.6 in a nonstressed CF sample, tentatively corresponding to a gallocatechin trimer derivatized with 3 molecules of β-mercaptoethanol (during thiolysis). Future application of this MIS analysis to both mass spectrometry and other high dimensionality data[56] will likely provide greater insight into the complex chemical processes that impact naturally derived complex mixtures.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Funding for this work was made possible by the Food and Drug Administration through grant 1U01FD005285-01. Views expressed in this publication do not necessarily reflect the official policies of the Department of Health and Human Services nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government. The views expressed in this publication by the authors do not necessarily reflect the views or policies of the Food and Drug Administration (FDA). In addition, Peter A. Kleindl was funded in part by the National Institutes of Health grant R01CA173292.

## References

1. Cottreau J, Tucker A, Crutchley R, Garey KW. Crofelemer for the treatment of secretory diarrhea. Expert Rev Gastroent. 2012;6:17–23.
2. Tradtrantip L, Namkung W, Verkman AS. Crofelemer, an antisecretory antidiarrheal proanthocyanidin oligomer extracted from Croton lechleri, targets two distinct intestinal chloride channels. Mol Pharmacol. 2010;77:69–78. [PubMed: 19808995]
3. U.S. Food and Drug Administration. Center for Drug Evaluation and Research. Fulyzaq NDA 202292 chemistry review. Silver Spring, MD: FDA; 2012 Available at: http//www.accessdata.fda.gov/drugsatfda_docs/nda/2012/202292Orig1s000ChemR.pdf. Accessed August 3, 2017.
4. Crutchley RD, Miller J, Garey KW. Crofelemer, a novel agent for treatment of secretory diarrhea. Ann Pharmacother. 2010;44:878–884. [PubMed: 20388859]
5. Schofield P, Mbugua DM, Pell AN. Analysis of condensed tannins: a review. Anim Feed Sci Tech. 2001;91:21–40.
6. Lv Q, Luo F, Zhao X, et al. Identification of proanthocyanidins from litchi (Litchi chinensis Sonn.) pulp by LC-ESI-Q-TOF-mS and their antioxidant activity. PLoS One. 2015;10:e0120480.
7. Kimura H, Ogawa S, Akihiro T, Yokota K. Structural analysis of A-type or B-type highly polymeric proanthocyanidins by thiolytic degradation and the implication in their inhibitory effects on pancreatic lipase. J Chromatogr A. 2011;1218:7704–7712. [PubMed: 21803362]
8. Fulcrand H, Remy S, Souquet JM, Cheynier V, Moutounet M. Study of wine tannin oligomers by on-line liquid chromatography electrospray ionization mass spectrometry. J Agric Food Chem. 1999;47:1023–1028. [PubMed: 10552410]
9. Ferreira D, Bekker R. Oligomeric proanthocyanidins: naturally occurring O-heterocycles. Nat Prod Rep. 1996;13:411–433.

10. Valls J, Millan S, Marti MP, Borras E, Arola L. Advanced separation methods of food anthocyanins, isoflavones and flavanols. J Chromatogr A. 2009;1216: 7143–7172. [PubMed: 19691963]
11. Cadiz-Gurrea MD, Fernandez-Arroyo S, Segura-Carretero A. Pine bark and green tea concentrated extracts: antioxidant activity and comprehensive characterization of bioactive compounds by HPLC-ESI-QTOF-MS. Int J Mol Sci. 2014;15:20382–20402. [PubMed: 25383680]
12. Ricardo-Da-Silva JM, Darmon N, Fernandez Y, Mitjavila S. Oxygen free-radical scavenger capacity in aqueous models of different procyanidins from grape seeds. J Agric Food Chem. 1991;39:1549–1552.
13. Jeong WS, Kong ANT. Biological properties of monomeric and polymeric catechins: green tea catechins and procyanidins. Pharm Biol. 2004;42:84–93.
14. Rice-Evans CA, Miller NJ, Paganga G. Structure-antioxidant activity relationships of flavonoids and phenolic acids. Free Radic Biol Med. 1996;20: 933–956. [PubMed: 8743980]
15. Apak R, Ozyurek M, Guclu K, Capanoglu E. Antioxidant activity/capacity measurement. 1. Classification, physicochemical principles, mechanisms, and electron transfer (ET)-Based assays. J Agric Food Chem. 2016;64:997–1027. [PubMed: 26728425]
16. Apak R, Ozyurek M, Guclu K, Capanoglu E. Antioxidant activity/capacity measurement. 2. Hydrogen atom transfer (HAT)-Based, mixed-mode (electron transfer (ET)/HAT), and lipid peroxidation assays. J Agric Food Chem. 2016;64: 1028–1045. [PubMed: 26805392]
17. Apak R, Ozyurek M, Guclu K, Capanoglu E. Antioxidant activity/capacity measurement. 3. Reactive oxygen and nitrogen species (ROS/RNS) scavenging assays, oxidative stress biomarkers, and chromatographic/chemometric assays. JAgricFood Chem. 2016;64:1046–1070.
18. Karamac M Chelation of Cu(II), Zn(II), and Fe(II) by tannin constituents of selected edible nuts. Int J Mol Sci. 2009;10:5485–5497. [PubMed: 20054482]
19. Andrade RG Jr, Dalvi LT, Silva JM Jr, Lopes GK, Alonso A, Hermes-Lima M. The antioxidant effect of tannic acid on the in vitro copper-mediated formation of free radicals. Arch Biochem Biophys. 2005;437:1–9. [PubMed: 15820211]
20. Lopes GK, Schulman HM, Hermes-Lima M. Polyphenol tannic acid inhibits hydroxyl radical formation from Fenton reaction by complexing ferrous ions. Biochim Biophys Acta. 1999;1472:142–152. [PubMed: 10572935]
21. Kleindl PA, Xiong J, Hewarathna A, et al. The botanical drug substance as a model system for comparative characterization of complex mixture drugs. J Pharm Sci. 2017;106(11):3242–3256. [PubMed: 28743606]
22. Gu L, Kelm M, Hammerstone JF, et al. Fractionation of polymeric procyanidins from lowbush blueberry and quantification of procyanidins in selected foods with an optimized normal-phase HPLC-MS fluorescent detection method. JAgricFood Chem. 2002;50:4852–4860.
23. Frampton JE. Crofelemer: a review of its use in the management of non-infectious diarrhoea in adult patients with HIV/AIDS on antiretroviral therapy. Drugs. 2013;73:1121–1129. [PubMed: 23807722]
24. Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken, NJ: Wiley Interscience; 2006.
25. Guyot S, Marnet N, Drilleau J. Thiolysis-HPLC characterization of apple procyanidins covering a large range of polymerization states. J Agric Food Chem. 2001;49:14–20. [PubMed: 11170553]
26. Kennedy JA, Jones GP. Analysis of proanthocyanidin cleavage products following acid-catalysis in the presence of excess phloroglucinol. J Agric Food Chem. 2001;49:1740–1746. [PubMed: 11308320]
27. Li HJ, Deinzer ML. Tandem mass spectrometry for sequencing proanthocyanidins. Anal Chem. 2007;79:1739–1748. [PubMed: 17297981]
28. Gu L, Kelm MA, Hammerstone JF, et al. Liquid chromatographic/electrospray ionization mass spectrometric studies of proanthocyanidins in foods. J Mass Spectrom. 2003;38:1272–1280. [PubMed: 14696209]
29. Manach C, Scalbert A, Morand C, Remesy C, Jimenez L. Polyphenols: food sources and bioavailability. Am J Clin Nutr. 2004;79:727–747. [PubMed: 15113710]

30. Monagas M, Quintanilla-Lopez JE, Gomez-Cordoves C, Bartolome B, Lebron-Aguilar R. MALDI-TOF MS analysis of plant proanthocyanidins. J Pharm Biomed Anal. 2010;51:358–372. [PubMed: 19410413]
31. Hammerstone JF, Lazarus SA, Mitchell AE, Rucker R, Schmitz HH. Identification of procyanidins in cocoa (Theobroma cacao) and chocolate using highperformance liquid chromatography mass spectrometry. J Agric Food Chem. 1999;47:490–496. [PubMed: 10563922]
32. Yanagida A, Kanda T, Shoji T, OhnishiKameyama M, Nagata T. Fractionation of apple procyanidins by size-exclusion chromatography. J Chromatogr A. 1999;855:181–190. [PubMed: 10514983]
33. Montero L, Herrero M, Prodanov M, Ibanez E, Cifuentes A. Characterization of grape seed procyanidins by comprehensive two-dimensional hydrophilic interaction x reversed phase liquid chromatography coupled to diode array detection and tandem mass spectrometry. Anal Bioanal Chem. 2013;405: 4627–4638. [PubMed: 23224621]
34. Karonen M, Leikas A, Loponen J, Sinkkonen J, Ossipov V, Pihlaja K. Reversed-phase HPLC-ESI/MS analysis of birch leaf proanthocyanidins after their acidic degradation in the presence of nucleophiles. Phytochem Anal. 2007;18:378–386. [PubMed: 17624904]
35. Yokota K, Kimura H, Ogawa S, Akihiro T. Analysis of a-type and B-Type highly polymeric proanthocyanidins and their biological activities as nutraceuticals. J Chem. 2013 Article ID 352042.
36. Prieur C, Rigaud J, Cheynier V, Moutounet M. Oligomeric and polymeric pro-cyanidins from grape seeds. Phytochemistry. 1994;36:781–784.
37. Seto R, Nakamura H, Nanjo F, Hara Y. Preparation of epimers of tea catechins by heat treatment. Biosci Biotechnol Biochem. 1997;61:1434–1439.
38. Crozier A, Lean MEJ, McDonald MS, Black C. Quantitative analysis of the flavonoid content of commercial tomatoes, onions, lettuce, and celery. J Agric Food Chem. 1997;45:590–595.
39. Raynal J, Moutounet M, Souquet JM. Intervention ofphenolic-compounds in plum technology. 1. Changes during drying. JAgricFood Chem. 1989;37:1046–1050.
40. Jood S, Chauhan BM, Kapoor AC. Polyphenols of chickpea and blackgram as affected by domestic processing and cooking methods. J Sci Food Agric. 1987;39:145–149.
41. Price KR, Bacon JR, Rhodes MJC. Effect of storage and domestic processing on the content and composition of flavonol glucosides in onion (Allium cepa). J Agric Food Chem. 1997;45:938–942.
42. Khanal RC, Howard LR, Prior RL. Effect of heating on the stability of grape and blueberry pomace procyanidins and total anthocyanins. Food Res Int. 2010;43: 1464–1469.
43. Chen L, Yuan P, Chen K, Jia Q, Li Y. Oxidative conversion of B- to A-type procyanidin trimer: evidence for quinone methide mechanism. Food Chem. 2014;154:315–322. [PubMed: 24518348]
44. He F, Pan QH, Shi Y, Duan CQ. Chemical synthesis of proanthocyanidins in vitro and their reactions in aging wines. Molecules. 2008;13:3007–3032. [PubMed: 19052525]
45. Janeiro P, Brett AMO. Catechin electrochemical oxidation mechanisms. Anal Chim Acta. 2004;518:109–115.
46. Tanaka T, Matsuo Y, Kouno I. Chemistry of secondary polyphenols produced during processing of tea and selected foods. Int J Mol Sci. 2010;11:14–40.
47. Li H, Guo A, Wang H. Mechanisms of oxidative browning of wine. Food Chem. 2008;108:1–13.
48. Venter PB, Senekal ND, Kemp G, et al. Analysis of commercial proanthocyanidins. Part 3: the chemical composition of wattle (Acacia mearnsii) bark extract. Phytochemistry. 2012;83:153–167. [PubMed: 22917955]
49. Brown JE, Khodr H, Hider RC, Rice-Evans CA. Structural dependence of flavonoid interactions with Cu2+ ions: implications for their antioxidant properties. Biochem J. 1998;330(Pt 3):1173–1178. [PubMed: 9494082]
50. Makris DP, Rossiter JT. Heat-induced, metal-catalyzed oxidative degradation of quercetin and rutin (quercetin 3-O-rhamnosylglucoside) in aqueous model systems. J Agric Food Chem. 2000;48:3830–3838. [PubMed: 10995278]
51. Cheynier V, Ricardo-da-Silva JM. Oxidation of grape procyanidins in model solutions containing trans-caffeoyltartaric acid and polyphenol oxidase. J Agric FoodChem. 1991;39:1047–1049.

52. Pourcel L, Routaboul JM, Cheynier V, Lepiniec L, Debeaujon I. Flavonoid oxidation in plants: from biochemical properties to physiological functions. Trends Plant Sci. 2007;12:29–36. [PubMed: 17161643]

53. Waterhouse AL, Laurie VF. Oxidation of wine phenolics: a critical evaluation and hypotheses. Am J Enol Viticult. 2006;57:306–313.

54. Mochizuki M, Yamazaki S, Kano K, Ikeda T. Kinetic analysis and mechanistic aspects of autoxidation of catechins. Biochim Biophys Acta. 2002;1569:35–44. [PubMed: 11853955]

55. Oszmianski J, Cheynier V, Moutounet M. Iron-catalyzed oxidation of (+)-catechin in model systems. J Agric Food Chem. 1996;44:1712–1715.

56. Nariya MK, Kim JH, Xiong J, et al. Comparative characterization of crofelemer samples using data mining and machine learning approaches with analytical stability data sets. J Pharm Sci. 2017;106:3270–3279. [PubMed: 28743607]

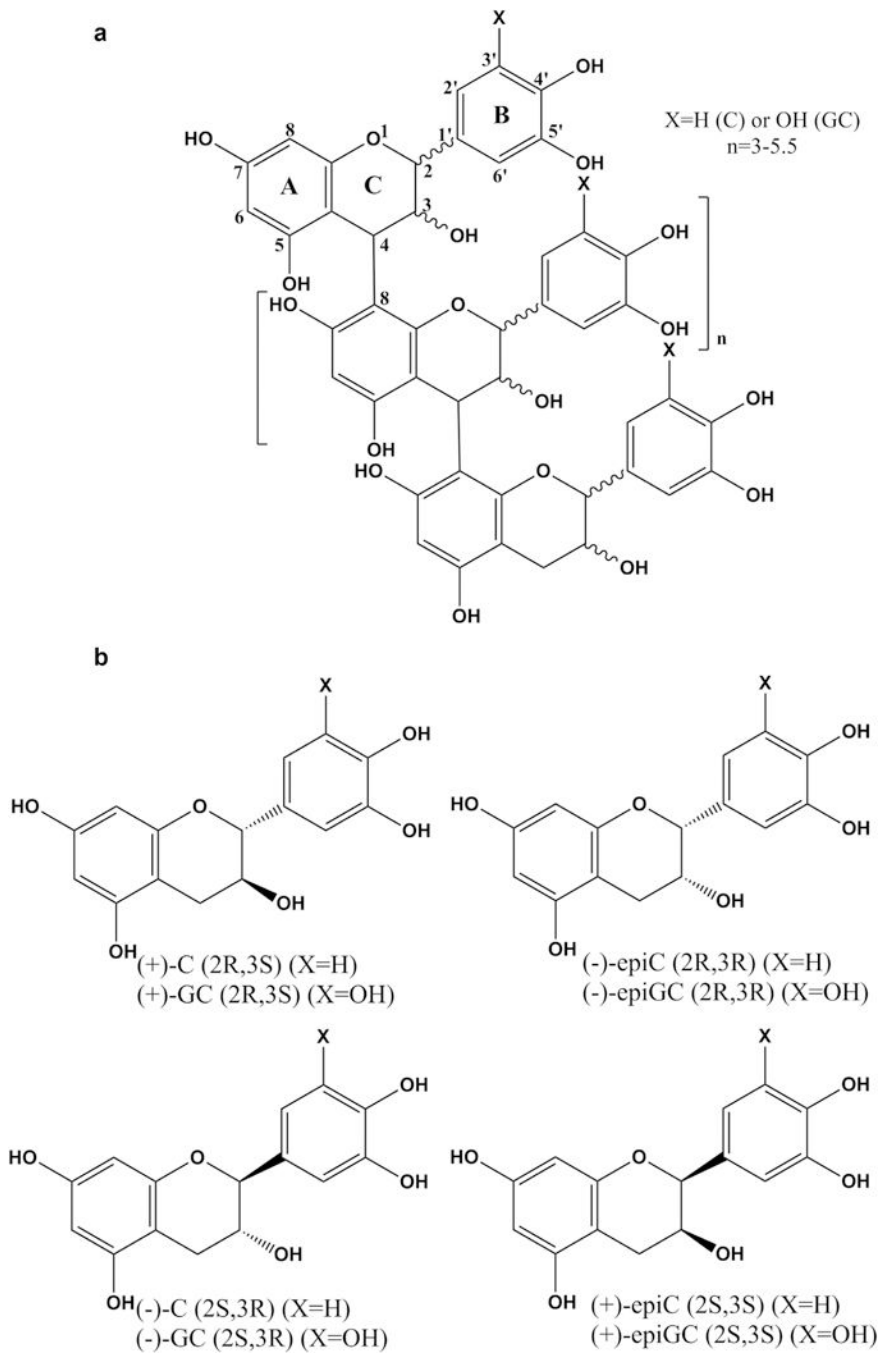



**Figure 1.**

(a) A representative structure of CF. Two aromatic rings are marked A and B, and the O-heterocyclic ring is marked C. (b) Individual constituents of CF.

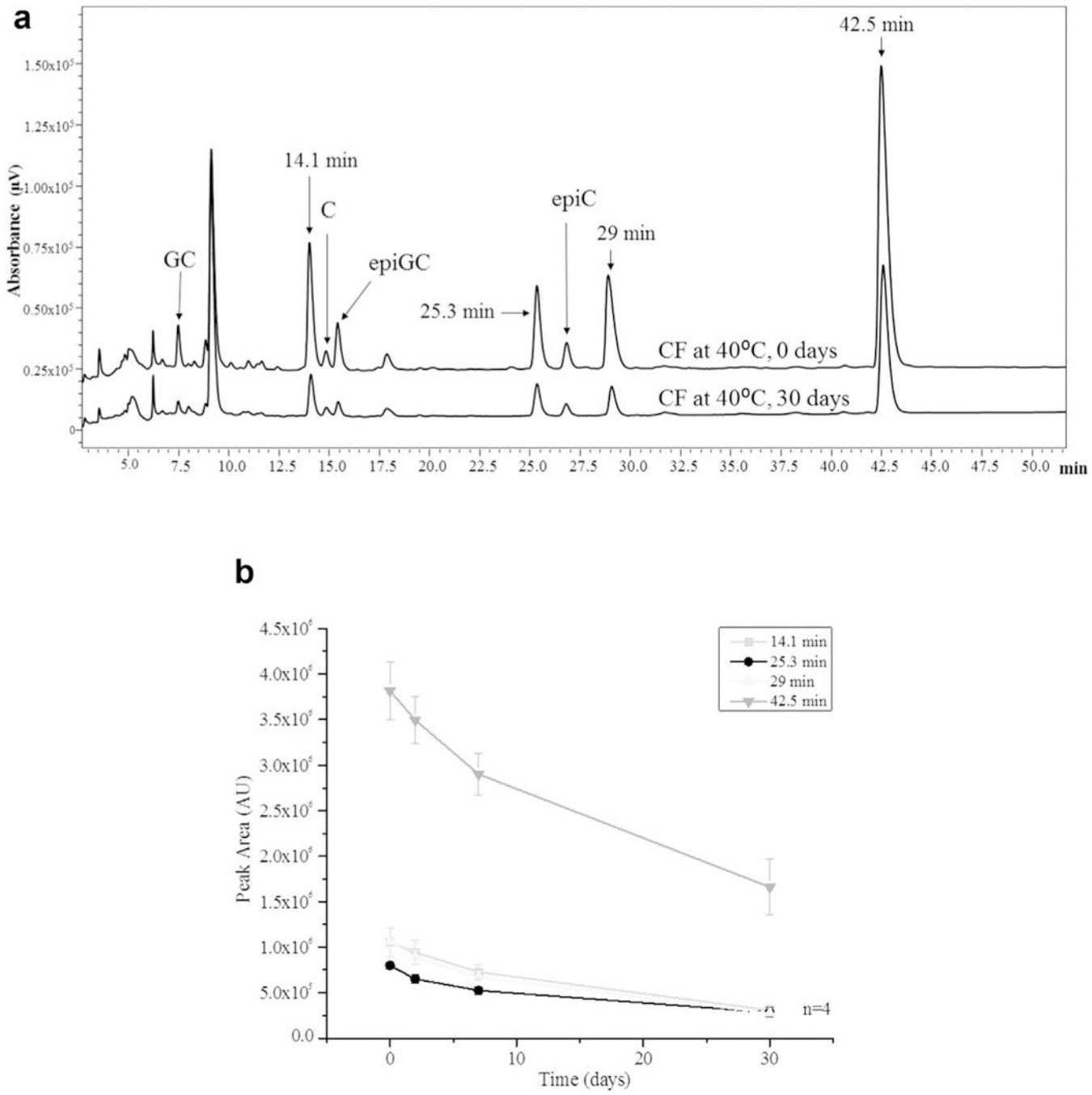



**Figure 2.**

(a) Comparison of RP-HPLC chromatograms at 280 nm of 25 mg/mL CF incubated at 40°C for 0 and 30 days. The temperature-stressed samples were thiolyzed before chromatography. Peaks at 7.5,14.9,15.5, and 27 min were assigned to GC, C, epiGC, and epiC, respectively. (b) The change of CF products eluting at 14.1 min, 25.3 min, 29 min, and 42.5 min over 0, 2, 7, and 30 days at 40°C as monitored by RP-HPLC and UV detection at 280 nm. The values shown are the mean ± SE for 4 sample replicates.

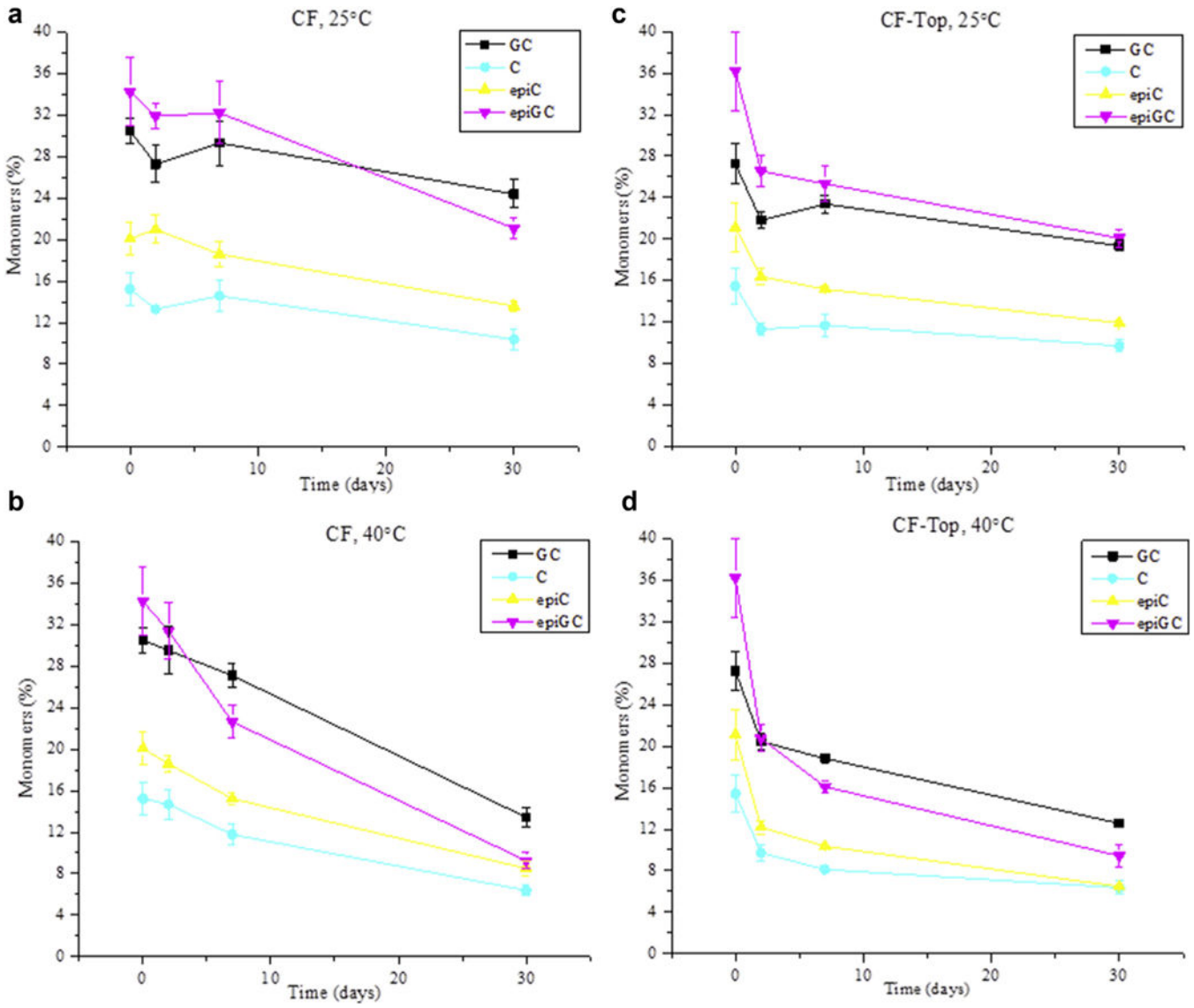



**Figure 3.**

Percentage of monomers, C, GC, epiC, and epiGC, generated over time in temperature-stressed CF samples. CF samples (25 mg/mL) were incubated at 25°C (a) and 40°C (b) followed by thiolysis. CF-Top (25 mg/mL) fractions were also incubated at 25°C (c) and 40°C (d) followed by thiolysis. The values shown are the mean ± SE for 4 sample replicates of temperature-stressed samples except 8 sample replicates for time zero. The total amount of terminal monomers was considered 100% at time zero.

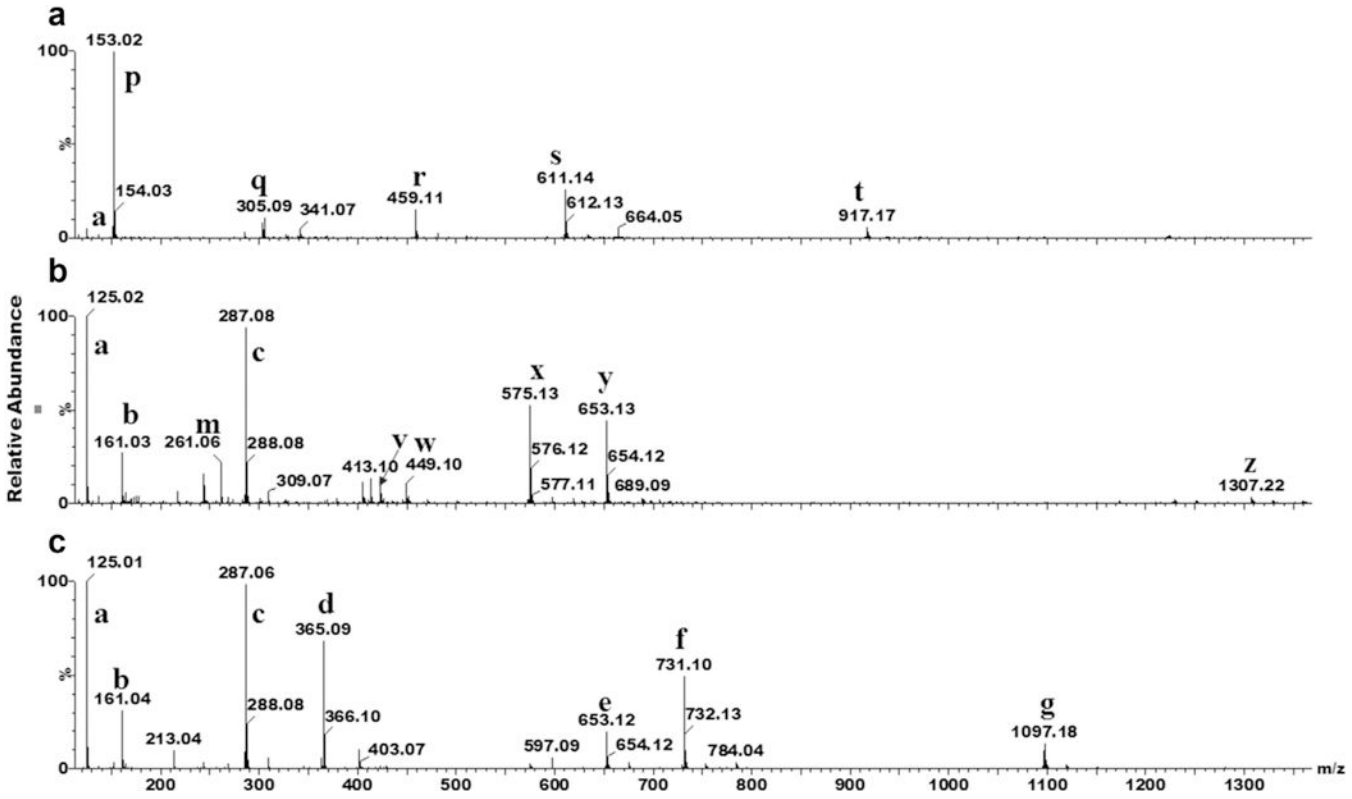



**Figure 4.**

CF samples (25 mg/mL) which were incubated at 40°C for 30 days were thiolyzed and analyzed by RP-HPLC at 280 nm and the peak eluting at 42.5 min was collected. The collected peak was lyophilized and analyzed by mass spectrometry. Three main peaks were observed on the TIC chromatogram and their $MS^2$ spectra are shown in a, b, and c. Refer Tables 1, and Supplementary Tables S1 and S2 for the structures and m/z of fragment ions.

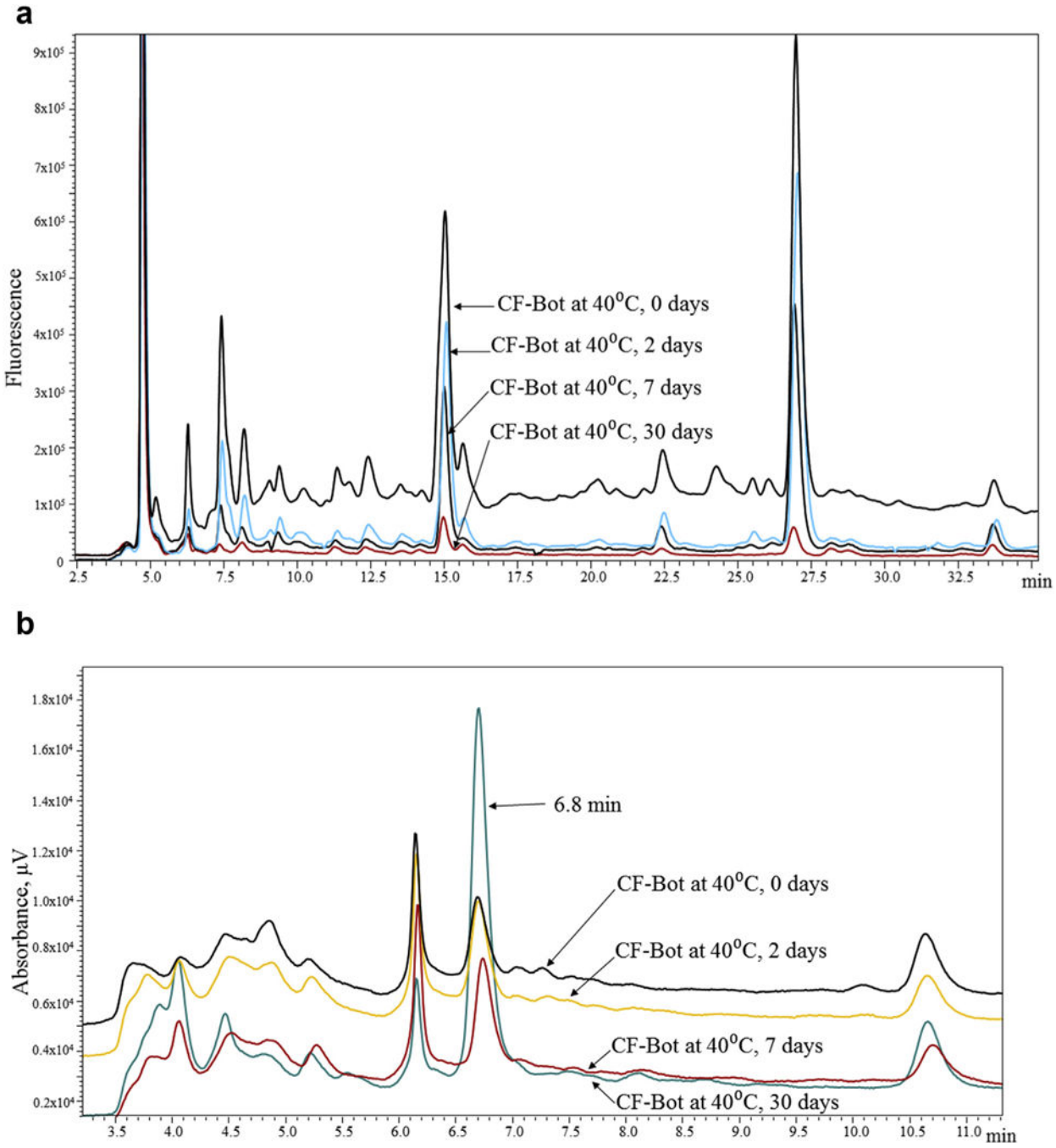



**Figure 5.**

(a) CF-Bot (0.5 mg/mL) fractions were incubated at 40°C for 0, 2, 7, and 30 days and analyzed by RP-HPLC and fluorescence detection at $\lambda_{em}$ = 276 nm and $\lambda_{em}$ = 316 nm. (b) CF-Bot (0.5 mg/mL) fractions were incubated at 40°C for 0, 2, 7, and 30 days and analyzed by RP-HPLC and UV detection at 340 nm. The peak eluting at 6.8 min has increased significantly after 30 days of incubation at 40°C. These samples are relatively low concentrated compared to CF and CF-Top and thus were not thiolyzed.

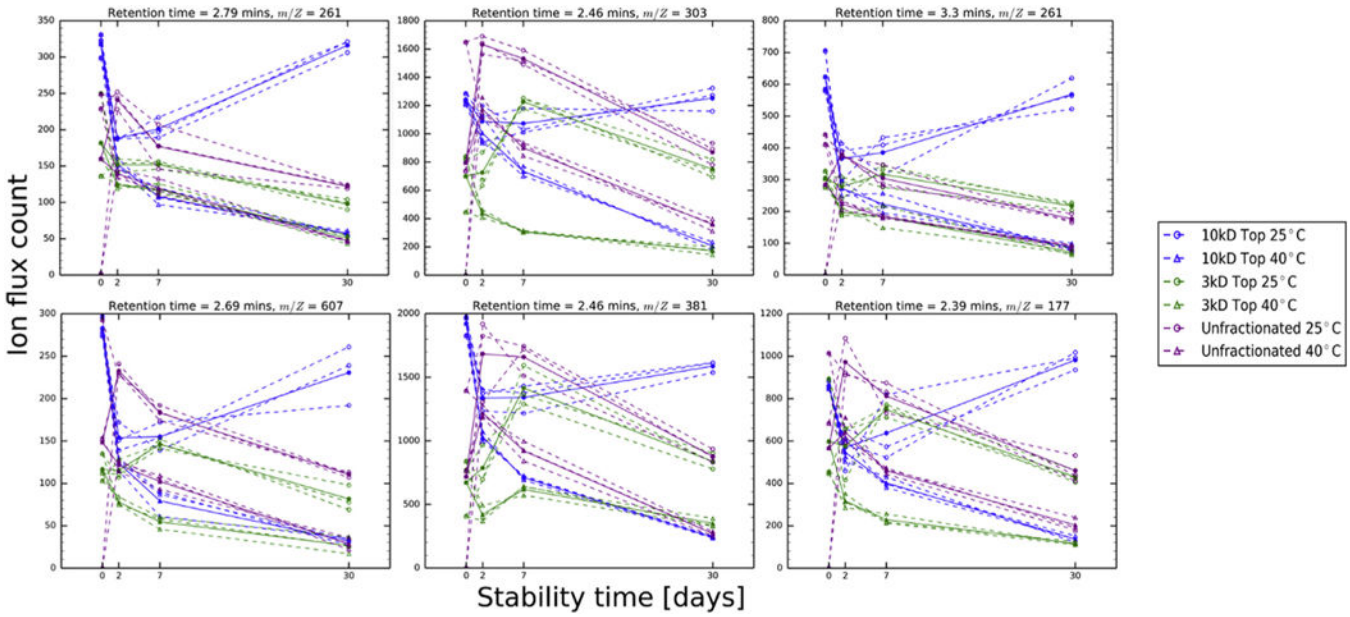



**Figure 6.**

Ion flux trajectories as a function of the stability time for the top 6 pairs of m/z and retention times with the highest mutual information score.

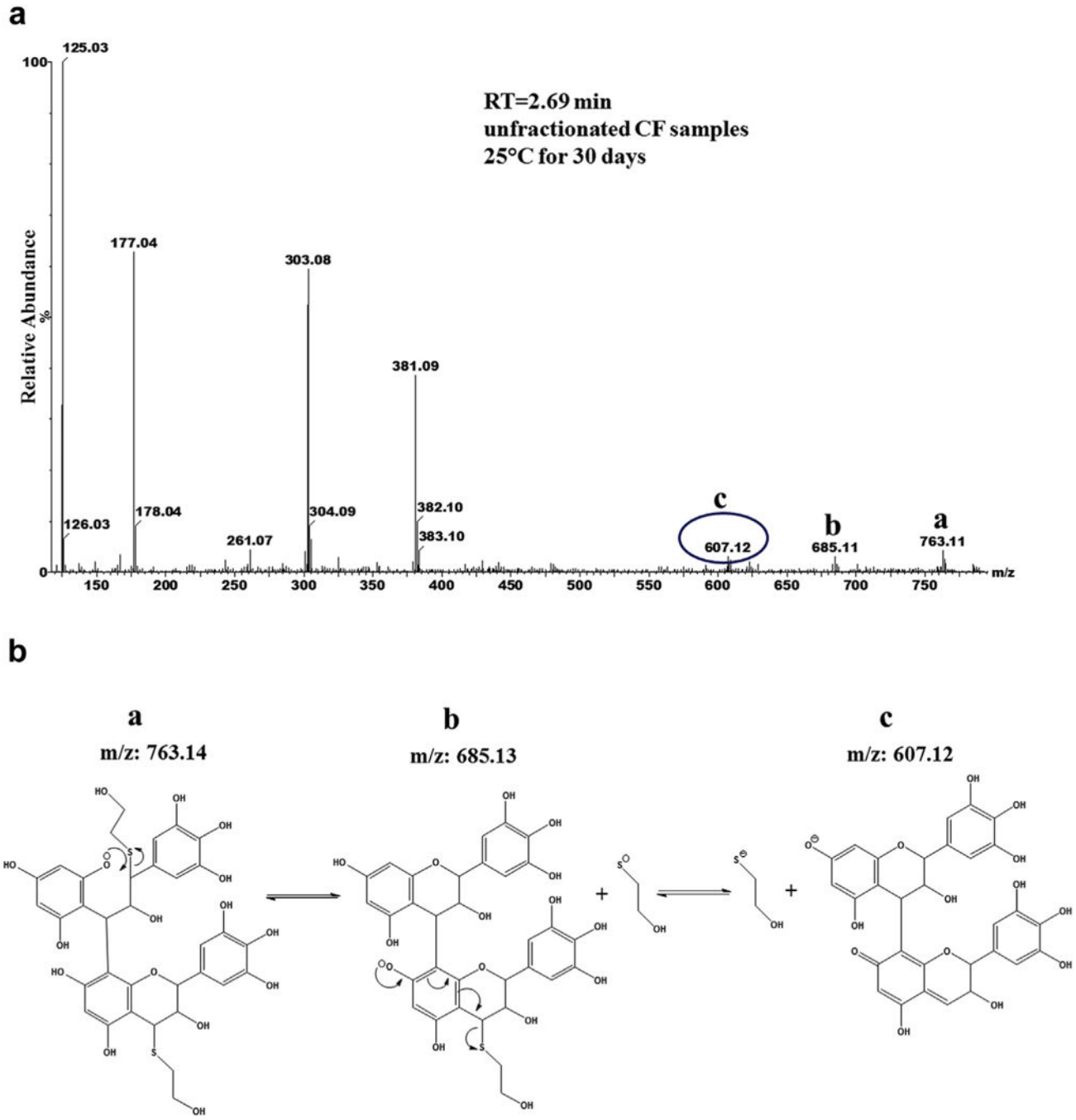
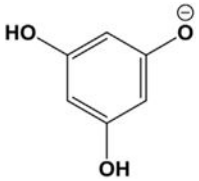
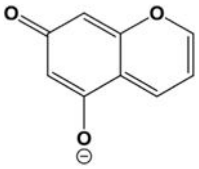
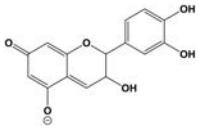
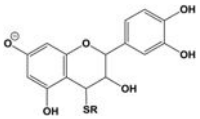
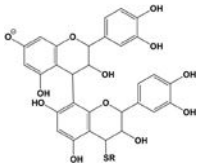
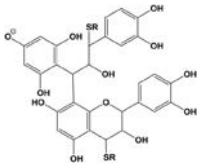
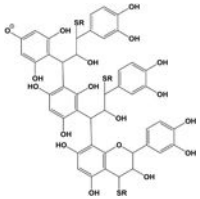



**Figure 7.**

A possible CF structural change under temperature stress over time as discovered by the mathematical model. (a) Mass spectrum (MS1) that was extracted at 2.69 min of 25 mg/mL CF-Top incubated at 25°C for 30 days. (b) A possible mechanism to explain the formation of m/z 607.12 in MS1.

**Table 1**

Negative-Ion ESI Fragment Ions of the Molecular Ion With m/z 1097.18 From Crofelemer (CF), Incubated for 30 Days at 40°C

| Ion m/z | Structure | Ion m/z | Structure |
|---|---|---|---|
| a 125.01 |  | b 161.04 |  |
| c 287.06 |  | d 365.09 |  |
| e 653.12 |  | f 731.10 |  |
| g 1097.18 |  | R = $-CH_2CH_2OH$ | |

The letters a-g correspond to the signals in the $MS^2$ spectrum.