PREDICTION OF MANDARIN JUICE FLAVOR:

A FLAVOROMIC APPROACH

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Joséphine Isabelle Marie Charve

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Dr Gary A. Reineccius, Advisor

May 2011

Acknowledgements

À mes parents,

À mes professeurs.

Table of Contents

List of Tables

List of Figures

Summary

The ability to predict flavor or sensory quality of food based on chemical composition has always been of interest for food scientists and food manufacturers, hoping it would serve as an alternative to sensory evaluations. Instrumental-sensory correlations have been used since the late 1970s to establish associations between chemical stimuli and flavor perceived, which can be used later for predictions. However, the task remains a challenge.

One major barrier to accurate predictions of flavor is due to the multi-dimensionality of the chemical stimuli involved in flavor perception. Even though the importance of both volatile and non-volatile compounds to flavor perception has been reported, there are few studies that have combined both to predict flavor. In addition, most works considered only aroma compounds or known sensory-active compounds, and often time overlooked at the contribution of other compounds.

This thesis introduces an alternative approach to predict flavor that may address some of the issues being neglected in the past and is referred to as flavoromics; this research adapts concepts and tools from metabolomics investigations. It is a non-targeted strategy combined with chemometrics which considers for study all (ideally) low molecular weight compounds in foods as candidate chemical stimuli in human flavor perception. This methodology presents substantial analytical and data processing challenges because of the expected diversity of compounds studied and the amount of instrumental information generated.

The feasibility of flavoromics to predict the intensities of various flavor attributes was tested on mandarin juice. Forty-six mandarin juices, from different cultivars and hybrids and harvest seasons, were characterized by both instrumental and descriptive sensory analyses. Volatiles and non-volatiles were analyzed by headspace solid-phase micro extraction gas chromatography (SPME-GC) and solid-phase extraction ultra high performance liquid chromatography (SPE-UHPLC) – time of flight mass spectrometry (TOF-MS), respectively. The developed methods were a compromise between the number of compounds extracted and detected, throughput, and repeatability. The capability of distinguishing samples based on mass spectral information collected from the different

instruments (GC- and UHPLC-TOF-MS) using chemometrics was confirmed. The descriptive sensory analysis of the mandarin juice samples revealed very different flavor profiles between and within hybrids and cultivars, and juices made from fruits with common genetic background (parents and hybrids) tended to share some sensory characteristics.

Compositional variations across mandarin juice samples and their sensory profile were correlated using partial least squares regression (PLSR), from which predictive models of sensory quality were developed. The explanatory and predictive performances of the models were improved when combining all instrumental data into one single data set as opposed to individual ones, thereby indicating that each individual subset conveyed complementary information and the fact of merging them improved the overall description of the sensory profile. The best PLS model was obtained with mid-level data fusion, for which a preliminary variable selection was done. The predictive power of the selected model was tested using a calibration and prediction sample sets (38 and 8 juices, respectively). A fairly robust model was obtained and a strong relationship between instrumental and sensory measurements was observed. The resulting model showed that prediction of sensory scores was possible to a certain extent for a majority of the sensory descriptors, demonstrating the applicability of using a data-driven approach to predict flavor irrespective of whether the chemical identity of the instrumental signals was known or not. The best predictions were obtained for the attributes of grapefruit, sour, fruity non-citrus, orange and pumpkin/fatty ($0.5 < Q^2Y < 0.7$), whereas tangerine, bitter and floral yielded the poorest ones ($Q^2Y < 0.35$).

The approach developed in this research might be a valuable tool in plant breeding programs to predict the sensory properties of newly developed hybrids, allowing faster screening. As well, other insights may be drawn from such a research approach, for instance by comparing the different hybrid populations against each other and identifying markers that contribute to their separation, or by revealing which compounds are associated with a particular sensory trait. This kind of information may assist in understanding the underlying genetic basis of the differences in sensory quality observed.

Keywords: flavoromics, chemometrics, non-targeted, flavor prediction, partial least squares regression, mandarin juice, sensory, volatiles, non-volatiles.

Schematic thesis outline

FLAVOROMIC RESEARCH APPLIED TO MANDARIN JUICE FLAVOR

Chapter 1
Background & significance
Research hypotheses &
objectives
Experimental approach &
research plan

Food	?	Flavor
chemical composition	chemical stimuli	*sensory attributes*

Chapter 2
Literature review

Mandarin juices
(differences in hybrids and harvest years)

Chapter 3 & 4
Development &
evaluation of
instrumental methods for
the untargeted analysis of
chemical stimuli of
orange juice flavor

Instrumental characterization

Sensory analysis

GC-, UHPLC-TOF-MS

Trained panel

Objective 1
To establish correlations
between the chemical
stimulus, or combination
of, and flavor attributes of
mandarin juices

Chapter 5
Sensory analysis of
mandarin juices

Volatiles & non-volatiles

PLS regression

Sensory profiles

Establish relationships between analytical
signal & sensory attributes
Cross-validation
Selection of most significant variables

Objective 2
To compare different
predictive models of
flavor attributes & to
select the best one in
terms of explanatory &
predictive performances

Chapter 6
Prediction of flavor using
a multi-instrumental
approach & chemometrics

Predictive models

Additional mandarin juices

Mass spectral library & database searches, MS/MS

Objective 3
To examine the selected
model to understand
which chemical signal(s)
correlates in a robust way
with specific flavor
attributes

External validation

Identification of the chemical stimuli contributing to model

Chapter 7
Conclusions & remarks

Objective 4
To identity some of the
compounds that
contributed the most to
the prediction of specific
flavor attributes

Chapter 1 : Introduction

This first chapter provides a general introduction to the research presented in this thesis. After a brief overview of the significance of this research to the flavor field, and more particularly to the understanding of the chemical stimuli involved in flavor perception, the objectives and hypotheses of the research are presented. The experimental approach taken, its novelty and potential limitations are presented as well. Chapter 2 will present the literature more focused on this research.

Flavor researchers are continuously working on a means to have a better understanding of the mechanisms involved in flavor perception as flavor is one of the main drivers of food acceptance. Flavor perception is a complex phenomenon to study since it involves a variety of stimuli resulting from the stimulation of our senses (olfaction, taste, texture and vision) while eating or drinking (Keast et al. 2004). In addition, flavor perception is modulated by many intrinsic (food structure) and extrinsic factors (physiological and emotional status of the consumer) that cannot always be studied or reproduced in a laboratory setting. Chemical stimuli from the food itself are undoubtedly contributing the most to the formation of flavor perception and understanding how/ which chemicals relate to flavor perception is of primary interest for many flavor chemists. First, it would help in predicting the sensory quality (flavor) of the food produced based on its chemical composition. Second, it would provide information on which compounds need to be increased or decreased in a food item either by changes in the process or formulation (e.g., change of an ingredient) or through breeding (e.g., selection of tomato fruits with a given flavor characteristic) so that product acceptance is maximized.

Instrumental-sensory correlations have been used since the late 1970s for that purpose (Aishima and Nakai 1991), coinciding with the emergence of chemometrics, the main idea being to establish associations between the chemical stimuli (measured instrumentally) and the flavor perceived (measured by sensory analyses). However, the prediction of sensory attributes based on instrumental data is still a main challenge for scientists. The links between sensory and chemical stimuli are still unclear as well as their mechanisms of action. One of the main drawbacks of most investigations trying to relate instrumental-

sensory measurements has been the neglect of the multi-modal dimension of flavor perception. Even though sensory scientists have long acknowledged this fact, flavor chemists traditionally have focused their attention largely on aroma compounds (volatiles) and only recently, taste compounds (non-volatiles). Moreover, traditional research methods (e.g., gas chromatography-olfactometry, aroma extract dilution analysis) lack the ability to study relationships within and across stimuli (taste and aroma) and generally do not consider a chemical compound as being important to perception unless it makes a contribution alone. These approaches consider only the compound's contribution on an individual basis and its concentration although research has proven that even compounds present at sub-thresholds may trigger a sensation by interacting with other sub-thresholds compounds (Buttery 1999; Labbe et al. 2007). It is, therefore, apparent that predicting the sensory attributes of a food from instrumental analyses remains a challenge despite the efforts that food scientists have spent on linking chemical and sensory data. The present thesis attempts to address a fundamental question in flavor research: **Can we predict the flavor of a food through instrumental measurements?**

This research introduces an alternative approach to predict flavor perception that may address some of the issues being neglected in the past and is referred to as **FLAVOROMICS**. The term was coined by Reineccius (2008) at the 235th ACS meeting and referred to as "flavour metabolomics" by de Vos et al. (2008) during a workshop at the 12th Weurman symposium; to our knowledge, this is the first time that an application of flavoromics is being reported. Flavoromics research is considered to be an evolution of flavor research and integrates advances made in other research disciplines (metabolomics, bioinformatics and chemometrics) and developments in instrumentation. By analogy with comprehensive researches such as metabolomics, flavoromics is an analytical non-targeted methodology aimed at linking the chemical composition of foods with sensory quality (flavor) using chemometrics. This approach is unbiased (in theory) and gives an overall vision of the chemical entities potentially involved in the formation of flavor; as opposed to traditional research, it monitors a wide range of compounds instead of a particular group of compounds. Flavoromics addresses the necessity of using a comprehensive approach (holistic) to grasp the multi-dimensionality of the chemical stimuli involved in flavor

2

perception. The novelty of this methodology for flavor investigation is described later in the text. The feasibility of flavoromics to predict the intensities of various flavor attributes was tested on mandarin (also referred to as tangerine) juice. As a side note, the thesis research was initially designed to study orange juice flavor which is the reason why instrumental methodologies were developed with orange and not mandarin juices. Unfortunately, this research had to be stopped due to a lack of sensory diversity in the samples provided to us. Since the success of the prediction of flavor relies on the selection of samples used to build predictive models, we decided to continue the work with another citrus fruit, mandarin, for which a large set of diverse samples was available.

1.1. RESEARCH HYPOTHESES AND OBJECTIVES

Due to the nature of this study (non-targeted approach), this research is exploratory rather than hypothesis-driven (targeted approach). In other words, it does not look at approval or disproval of specific hypotheses like "traditional" research. In fact, non-targeted research usually generates hypotheses as an outcome of the experiment, and these hypotheses are tested in further experiments (targeted, this time). Instead of focusing on a particular subset of compounds defined *a priori* and studying how they relate to the sensory profile of mandarin juice, the entire juice matrix (in this thesis) was examined. Indeed, there is evidence in the literature that the presence of non-volatile components in orange juice (such as sugars, acids, lipids and polyphenols) has an effect on flavor perception, either by binding with the aroma compounds, or affecting their release from the food matrix (Ahmed et al. 1978a; Ahmed et al. 1978c; Hasegawa et al. 1996; Jung et al. 2000; Buettner and Schieberle 2001; Havekotte et al. 2009). Hence, it is legitimate to consider both volatiles and non-volatiles for analysis in order to understand how/ which components contribute to flavor perception.

The main goal of such an approach is to gather as much information as possible from known and unknown compounds, without being restricted to previous knowledge which can give an incomplete overview of the phenomenon studied (flavor perception). Even

though there are <u>no specific</u> hypotheses formulated initially (e.g., the concentration of compound X is affecting the perception of Y), a non-targeted research does not imply a "no" hypotheses at all. In fact, scientific background information provides directions for selecting the sample preparative techniques and instrumental platforms to use and also for speculating about some of the expected results (e.g., to assess the coherence of the observed results).

The aim of this research was to explore the relationships between the chemical composition (taste and aroma compounds) of various mandarin juices and their flavor attributes in order to develop predictive models of the flavor quality of mandarin juice, using an instrumental non-targeted approach combined with chemometrics.

1.1.1. Hypotheses

1) It is generally accepted that the sensory properties of a food are related to the chemical compounds in a food. It is hypothesized that chemometric techniques, particularly partial least squares regression, are appropriate for finding the covariance structure that links the changes in chemical composition of the mandarin juices with their sensory quality and therefore can be used to build predictive models of flavor.

2) Given that the perception of food flavor is a multimodal phenomenon involving a variety of chemical stimuli such as volatiles and non-volatiles, the integration of "all" low molecular weight compounds in the mandarin juice is the best manner to predict its flavor attributes. It is hypothesized that including the data sets of volatiles and non-volatiles analyses in a predictive model will improve the accuracy of prediction of the flavor attributes, as opposed to using volatiles only.

3) As the proposed approach is data-driven, it does not require spending time on compound identification to establish predictive models for flavor. It is hypothesized that this approach provides the possibility to quickly screen for compounds that are related to flavor, and that these compounds correspond to the variables (given as "retention time – mass-to-charge ratio" RT-m/z) having the strongest importance in the predictive models.

1.1.2. Objectives

1)   The first objective was to establish correlations between the chemical compounds (measured as m/z-RT) and specific flavor attributes of mandarin juice. The research combined results from a mass spectrometry-based characterization and a descriptive sensory analysis of mandarin juices, using partial least squares regression (PLSR) to link the chemical stimuli (instrument) and flavor attributes (sensory).

2)   The second objective was to develop different PLS models including volatiles only, non-volatiles only, or the combination of both and to compare them using the parameters R^2 and Q^2 as a measure of their explanatory and predictive abilities, respectively. The correlation structure between instrumental and sensory data was also examined by means of the score plot t_1/u_1. The best predictive model was validated through internal validation and external validation (additional set of samples).

3)   The third objective was to examine the selected model in order to understand which chemical signal (or combination of) correlates in a robust way with specific flavor attributes. The regression coefficients of the partial least squares models were examined to understand their influence on the models and the relationships between the instrumental signal and sensory responses.

4)   The fourth objective was to identify some of the compounds that contributed the most to the prediction of specific flavor attributes. Mass spectral library search, retention indices and high resolution mass spectrometry combined with a MS/MS strategy was used.

1.2.   EXPERIMENTAL APPROACH AND RESEARCH PLAN

1.2.1. Experimental approach: overview

A schematic overview of the experimental strategy followed in this thesis is shown in Figure 1.1. Briefly, a non-targeted instrumental approach in combination with chemometrics was developed and used to predict the intensities of flavor attributes of mandarin juices. This research strategy offered breadth in the compounds investigated

instead of focusing only on compounds that are already known to influence the flavor quality (biased). Several mandarin juices (differences in hybrids and harvest years), were characterized by both instrumental and descriptive sensory analyses. The chemical composition of the juices was characterized by GC- and UHPLC-TOF-MS in order to provide complementary data in terms of chemical compounds analyzed, and semi-quantitative data were collected. In parallel, the flavor of the mandarin juices was characterized by means of descriptive sensory analysis. This latter part of the study was not conducted at the University of Minnesota but was done at the University of Florida Citrus Research and Education Center and at the USDA/ARS, Citrus and Subtropical Products Laboratory (Winter Haven, FL).

In common with other comprehensive research, the instrumental data collected was defined by a large number of correlated variables (chemical markers) and few observations (orange juices) which made chemometrics a tool of choice to explore the data sets. Partial least squares regression (PLSR) was used to correlate instrumental measurements to sensory attributes. As opposed to univariate data analysis, PLS methodology can simultaneously measure the association of several variables (e.g., volatiles and non-volatiles) with the response (flavor attributes). Also, PLSR "quantifies" the magnitude of association between variables and response. The predictive models were built using, in combination or individually, all information from GC and UHPLC analyses. The models were reduced afterward so that only relevant variables (chemical signals: m/z-RT and associated intensities) were included. The predictive models developed were compared and examined for their accuracy, and to reveal which chemical signals contributed to specific flavor attributes of mandarin juice. Subsequently, the identity of the compounds associated with those signals was attempted. The sensory relevance of the identified compounds was not tested in the present research and will need to be evaluated in the future.
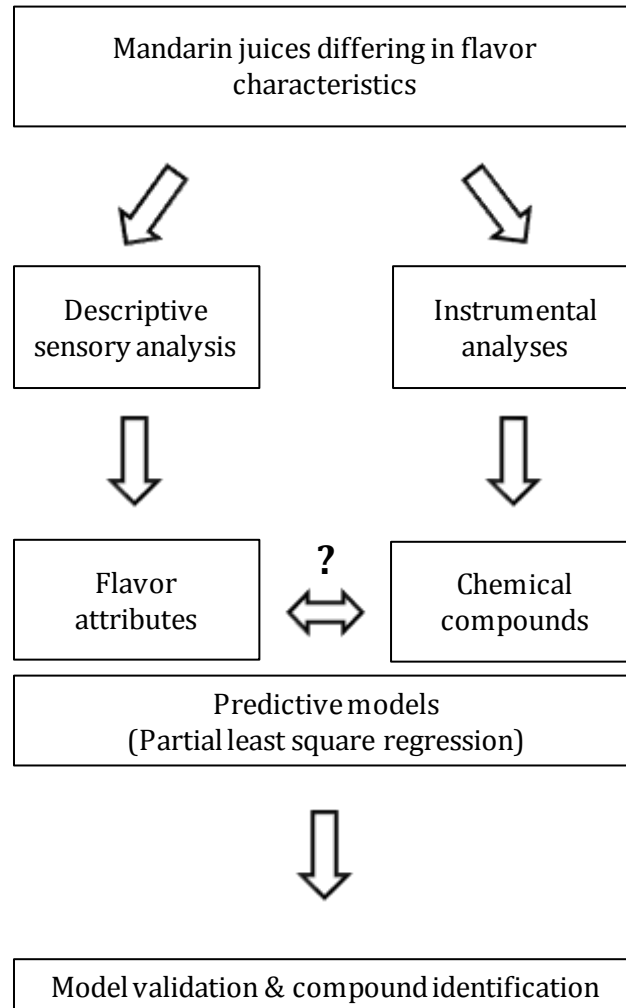
Figure 1.1: Schematic overview of the experimental strategy. (Note: the descriptive sensory analysis was conducted at the USDA/ARS, Citrus and Subtropical Products Laboratory, Winter Haven, FL).

1.2.2. Research plan

   To achieve the objectives outlined in the previous section, the research was organized as follows:

1)   Develop instrumental methods capable of measuring the largest unbiased pool of

compounds present in orange juice including volatiles and non-volatiles. (As stated earlier, the work was initially aimed at orange juice flavor but was later transferred to mandarin juice. We decided that the methods developed for the orange juice were valid for the mandarin juice since both fruits belong to the genus *Citrus* therefore having close chemical composition).

    a. The methodology must maximize the number of compounds extracted, separated and detected with minimum bias, be readily automated to collect large amounts of data and involve little sample preparation to maximize throughput. These goals tend to be conflicting or mutually exclusive thus requiring compromise in ultimately developing methods. The favored instrumental methods use GC-TOF-MS and UHPLC-TOF-MS for the analysis of volatile and non-volatile compounds, respectively.

    b. Assess if the developed methods were robust; that is, able to detect chemical differences between orange juices and repeatable.

    c. Evaluate our ability to manage, and analyze the massive amounts of data coming from different sources/instruments (GC- and UHPLC-TOF-MS) using spectral processing and chemometrics software.

2)   Establish the sensory profile of the mandarin juices by means of descriptive sensory analysis using a trained panel.

    a. Develop a lexicon (list of descriptors) that describes the flavor of mandarin juice and to train the assessors for tasting and using the lexicon. (This part was done by the USDA/ARS, Citrus and Subtropical Products Laboratory).

    b. Evaluate the assessors and panel performance in terms of ability to discriminate between samples, repeatability in scoring and agreement with the panel.

    c. Select the sensory attributes that offer significant statistical differences between the samples.

3)   Establish correlations between the instrumental data and sensory scores of mandarin juices.

a. Characterize by instrumental measurements (GC- and UHPLC-TOF-MS) the compounds present in mandarin juices from different hybrids and harvest years.

b. Characterize by quantitative descriptive sensory analysis the mandarin juice samples.

c. Establish linear multivariate models that relate instrumental measurements (X) and sensory measurements (Y) using partial least squares regression and to compare their quality (goodness of fit and predictability) when including in the models only volatiles vs. including volatiles and non-volatiles.

d. Examine if including only variables having the strongest influence on the model (based on their VIP value) improves the quality of the models, and if so, use those for subsequent predictions.

e. Validate the predictive models by means of internal and external validation.

4) Identify the chemical compounds which are strongly associated with some specific flavor attributes of the mandarin juice.

a. Conduct a library search for the mass spectra obtained by GC-TOF-MS; for LC-TOF-MS: conduct a search on publicly available databases using the accurate mass measurements.

b. Conduct tandem mass-spectrometry (MS/MS) analyses to gain further structural information on the compounds (from UHPLC-MS analyses).

c. Confirm compound identity by the injection of reference standards, when commercially available.

1.3. NOVELTY OF THE RESEARCH AND POTENTIAL OUTCOMES

1.3.1. Novelty

The originality/novelty of the present research for flavor investigation resides in the following facts:

✓ It is a comprehensive approach that considers the stimuli of both volatiles and non-volatiles as inputs of flavor perception, instead of focusing only on compounds that are

already known to influence flavor quality (unbiased). This approach potentially allows one to identify new flavor contributors and also to have a better prediction of flavor as it includes inputs from more chemical compounds.

✓ It is a data-driven approach that does not require compound identification (at least in the initial stages of the research) to establish predictive models of flavor. The instrumental measurements provide an objective and repeatable mean to assess the presence and relative concentration of the compounds (given as m/z-RT and associated intensity) with no need for identification. This provides the possibility to more quickly screen for compounds that are related to flavor since only those compounds linked to perception are identified in subsequent experiments (it would be a tedious and costly task to identify and validate all possible candidates). Additional sensory studies are necessary to confirm and to understand better the nature/mechanisms of their contribution to flavor.

✓ It uses chemometric tools such as multivariate multi-block techniques to establish predictive models, meaning that datasets from the different instrumental analyses are merged into a unique matrix. This gives a complete overview of the juice chemical composition since it considers simultaneously the contribution of many variables (volatiles and non-volatiles) to the overall flavor profile. This may increase the accuracy of the predictive models since all chemical stimuli measured are included; it may also reveal interactions between them.

✓ It integrates knowledge from other research fields such as metabolomics and chemometrics. For instance, it borrows sample preparative techniques, instrumentation (accurate mass measurement, UHPLC) and statistical analyses that are not commonly used in flavor research.

1.3.2. Outcomes

Flavoromics can be used differently depending on the final objective of the researcher. Two main applications serve to provide a better understanding of flavor

perception: flavor prediction and discovery of flavor contributors. The present thesis examines the first one.

1.3.2.1. Prediction

One application of flavoromics is to use data on volatiles and non-volatiles to predict flavor quality. The creation of predictive models of flavor profile, using instrumental measurements as inputs, may help food manufacturers in designing foods with a specific flavor profile or to control the quality of the product being produced. The fact of considering as many compounds as possible in the models takes care of possible correlations within and across stimuli, generally not accounted for by traditional research methods. It is important to highlight that the predictive models are used to predict a food's quality characteristics (objective assessment) and not its acceptance/preference (subjective assessment).

1.3.2.2. Discovery

Another application of flavoromics is to uncover/identify the compounds resulting from using different processes, raw materials or formulations and evaluate their relation to differences in flavor attributes. This is achieved by highlighting the chemical markers that are statistically different between food products and examining their relationships with the differences in sensory quality. Once identified, a researcher may investigate the origin of the compounds (e.g., how are they formed?); for instance, to enhance or decrease their formation if they are associated with desirable or undesirable flavor attributes, respectively. Using a comprehensive approach (enlarging the spectrum of compounds studied), this research strategy offers the potential to identify a broader base of compounds that "shape" our perception for a given flavor attribute.

1.4. CHALLENGES OF THE RESEARCH

Flavoromics is considered to be an additional tool for flavor research in terms of how to approach the question of flavor perception and certainly is not without challenges. This section is a concise exposé of the expected challenges and how to approach them; the

limitations and weaknesses of flavoromic research are discussed in the final chapter of this thesis.

✓   A primary challenge is the fact that it is not possible to measure and detect all chemical components present in a mandarin juice using a single methodology. Therefore, methods development considered complementary platforms and detection modes in an attempt to detect as many compounds as possible in the shortest amount of time. To compensate for the time spent in data collection, the time involved in sample preparation and data processing (no identification of compounds) were minimized.

✓   A direct consequence of using multiple instrumental analyses is the large amount of data generated which introduces a second challenge. In fact, the extent and quality of the outcome of the study depends directly on the ability to handle and extract relevant information from the instrumental data. A lot of tools and techniques have been developed in the last decades for this purpose. Data handling can be roughly divided in two steps: data pre-processing and data analysis. We chose to put time into acquiring knowledge on signal processing methods (MS signal deconvolution, spectral alignment, noise filtering, etc.) and chemometric techniques (normalization, scaling, merging of data sets, etc.), as well as becoming familiar with the software available for these tasks.

✓   An additional challenge for the success of flavoromics resides in the experimental design and more particularly in the samples used. Indeed, the success for the prediction of flavor depends strongly on the samples selected to develop the models. It is essential to choose samples that offer a large range of sensory quality (and therefore different chemical compositions) so that robust predictive models can be established. With that acknowledged, it is important noting that for the present thesis, there was little control over sample choice as they were provided by the USDA/ARS, Citrus and Subtropical Products Laboratory. However, these samples have shown diversity in previous sensory studies conducted at the USDA research center. Another important remark is that it was not possible to evaluate simultaneously the samples by sensory and instrumental analyses. The juices used for this thesis are coming from a three-year study (2006-2009) conducted by the breeding program of the University of Florida Citrus Research and Education Center. Even though samples

were kept frozen (-18 °C) until being shipped to us for the instrumental analyses, some degradation may have occurred during storage thereby introducing differences between the samples tasted and analyzed instrumentally.

✓ Finally, another difficult task is the identification of the chemical compounds highlighted by the multivariate analyses (i.e., variables being the most influential in the predictive models), especially for variables selected from liquid chromatography analyses. Indeed, there are only limited mass spectral libraries available for LC-MS as opposed to gas chromatography. The use of high mass-accuracy mass spectrometers in combination with MS/MS may facilitate compound identification but confirmation of its identity by injecting a standard (if commercially available!) or by NMR spectroscopy is still needed.

1.5. FLAVOROMIC RESEARCH APPLIED TO MANDARIN JUICE FLAVOR

While in the long term flavoromics can be broadly applied to food products, the feasibility of using flavoromics to predict flavor and to provide a better understanding of flavor contributors was initially tested on mandarin juice. Even though mandarins are the second largest citrus crop (after oranges) produced worldwide, the mandarin juice market in the US is small since mandarins are mostly eaten as a fresh fruit. For instance, less than 6 % of its total production (20.3 million metric tons in 2010) (USDA Foreign Agricultural Service 2011) is processed as juice, concentrate, canned fruit or blend. Consequently, the number of studies available on mandarin flavor is limited.

Producing fruit juices with highly desirable sensory qualities (e.g. flavor) is one strategy to increase consumption and is also one of the goals of the citrus breeding programs. Indeed, the juice's flavor results from a complex mixture of chemical components whose presence and levels are largely dependent upon raw material (and processing operations). The characterization of the flavor profile of mandarin hybrids and the identification of compounds responsible for their characteristic flavor traits plays a key role in helping the breeder to select fruits with desirable sensory quality. As well, it brings additional

understanding on the flavor genesis pathways when those compounds ("flavor markers") are related to the genetic background of the hybrids.

It is apparent that the flavoromic approach can benefit the citrus breeding program in several ways. First, it can be used as a methodology to predict the flavor quality of mandarin juice from instrumental data. As of now, researchers utilize expert sensory panels to characterize the flavor of the juices. However, sensory panels are expensive and time-consuming and it is not possible to use them extensively over a long time period or for large sample sets. Second, it offers the possibility to identify compounds ("flavor markers") associated with specific flavor characteristics (desirable or not). Then, the breeder can trace back to the gene responsible (molecular marker) for the presence of the flavor marker by relating inter-varietal similarities/ differences in the chemical composition to the genetic background. This may provide knowledge on the inheritance mechanisms of genes (e.g., dominant genes) involved in flavor biosynthesis and consequently may accelerate the selection of parents to grow fruits with the desired sensory properties. This second application of flavoromics (relation with genetic background) was not considered in this thesis as it is beyond my area of expertise.

The present research seeks to establish a quantitative relationship between the chemical composition of mandarin juices and their sensory attributes using a non-targeted approach, which was used later for prediction. A set of mandarin juices from various hybrids and harvest years and therefore, different flavor profiles, was characterized instrumentally and by means of descriptive sensory analysis. Instead of studying a specific set of compounds such as aroma compounds (targeted approach), "all" compounds were considered as far as possible. In this way, the investigation was not limited to already known flavor contributors and it opened the possibility to improve flavor prediction and to identify new flavor stimuli that were overlooked in previous works. To our knowledge, there is no study that uses a similar approach to understand the contribution of individual compounds and groups of compounds to mandarin juice flavor attributes.

Chapter 2 : Literature review

This second chapter provides the reader a concise review of facts and findings from the literature considered relevant to this project. It includes an overview of the mechanisms of flavor perception and how it is studied, a presentation of comprehensive research and their methodologies, and lastly, a look at the chemistry of citrus flavor.

2.1. FLAVOR PERCEPTION

2.1.1. A multimodal phenomenon

Historically, flavor chemists have largely focused their research on the aroma of foods but with time, they have come to understand that other sensory stimuli also contribute to flavor. Today it is well accepted that flavor perception includes inputs from volatile (aroma) and non-volatile compounds (taste, chemestetics). A presentation of the major compounds in foods involved in creating flavor perception and the mechanisms that affect flavor perception follows, as well as an overview on how flavor scientists are investigating the phenomenon of flavor perception.

2.1.1.1. Sensory inputs of flavor perception

In developed countries, flavor is the primary driver in food liking (Schultz and Wahl 1981) and results from many external and internal factors (Figure 2.1). Among those, chemical stimuli coming directly from the food play an important role in the perception of flavor. The most common known contributors are aroma, taste and chemesthetic compounds (irritants).

Aroma compounds (volatiles) are responsible for the odor of foods. They are carried by the inhaled air to the olfactory receptors of the nasal cavity via two pathways: orthonasal and retronasal (Linforth and Taylor 2006). The orthonasal pathway is the direct route through the nostrils, whereas the retronasal pathway is indirect, and implies travel through the nasopharynx (Figure 2.2). The binding of volatiles to the receptor cells triggers nerve

impulses (olfactory sensation) that are transmitted to the brain, where they are decoded and finally interpreted (olfactory perception) (Bell 1996). It is generally recognized that foods may contain a large number of volatiles but of these volatiles, only a limited number can be perceived and fewer likely contribute to the characteristic aroma of a food (key odorants). The sensory threshold of an odorant depends on its vapor pressure (influenced by the food matrix and other compounds present) and a human's inherent ability to detect it. A good illustration of the broad range in sensory thresholds exhibited by odorants is β-pinene and ethyl-2-methylbutanoate with orthonasal odor thresholds in a orange juice matrix of 37.2 mg/L and 8.10^{-5} mg/L, respectively (Plotto et al. 2004; Plotto et al. 2008).

Taste compounds are most often non-volatile compounds; they encompass a diversity of molecules (100-20,000 Da) including, for example, sugars, salts, acids and nucleotides. These compounds contribute to the five basic tastes: sweet, salty, sour, bitter, and umami (Delwiche 1996; Laing and Jinks 1996). Taste compounds interact primarily with taste receptors located primarily on the surface of taste buds of the tongue (Lindemann 1996), and are transported to the taste buds by saliva. Similar to olfaction, the binding of taste compounds to specific taste receptors leads to taste perception. The transduction of sour and salty tastes involves the opening of ion channels at the surface of the taste receptor cells, whereas bitter, umami and sweet tastes result from the binding of taste compounds to surface receptor sites (e.g., G proteins) and subsequent transduction cascades (Lindemann 2001). Genes encoding for taste receptors were recently identified (Matsunami et al. 2000; Max et al. 2001). This is of particular interest for the discovery of bitter-masking compounds or design of new sweeteners (e.g., reduce calorie content) since knowing the binding-site structure could help in tailoring taste compounds that activate or inhibit specific taste receptors.
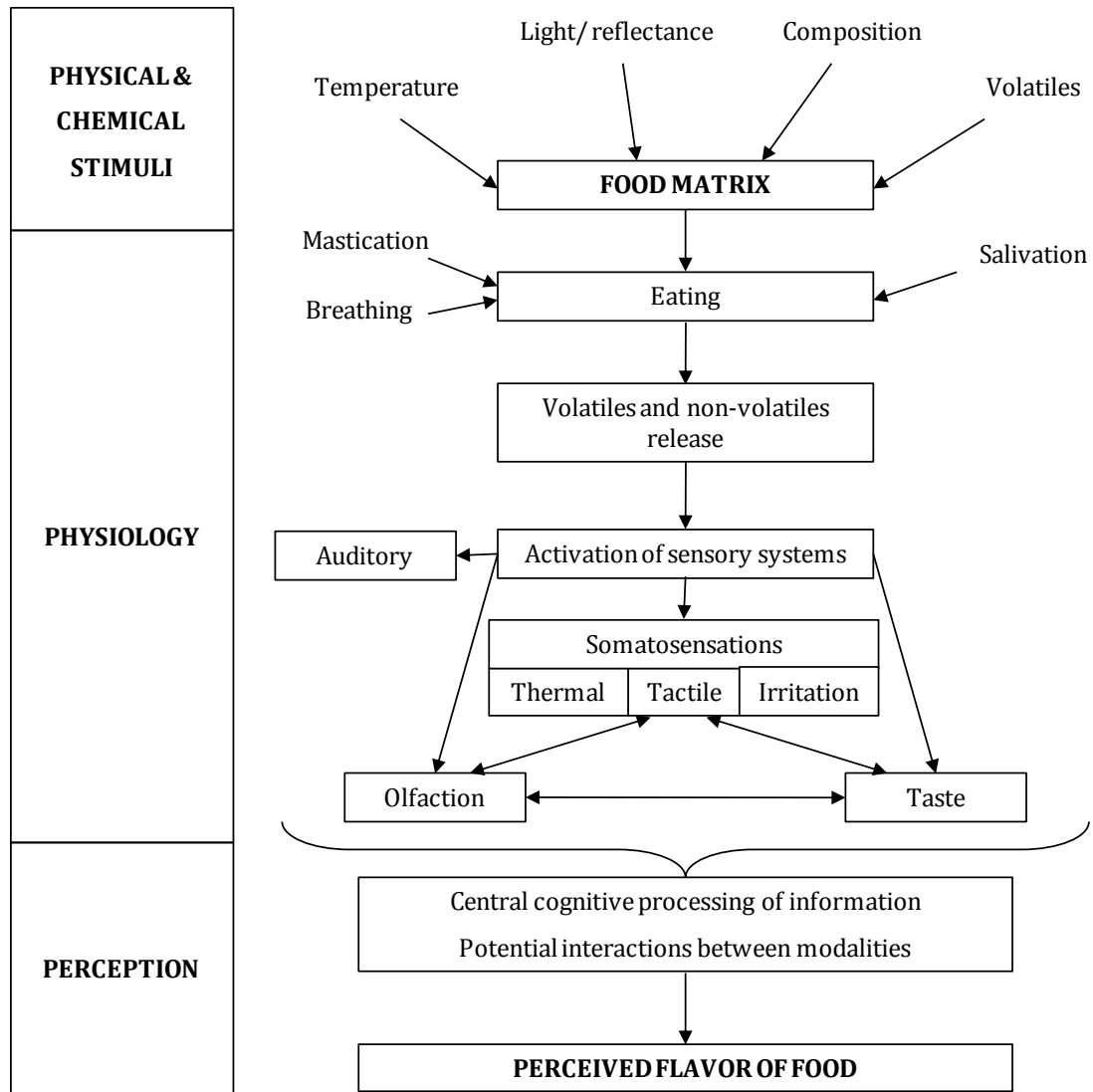
Figure 2.1: Flow diagram outlining stages and factors involved in flavor perception (adapted from Keast et al. 2004). (Copyright 2004 Wiley. Used with permission.)

Lastly, chemesthetic compounds (irritants) are responsible for oral and nasal chemosensory sensations (chemesthesis) and also contribute to flavor perception. They generally have a molecular weight in the range of volatile compounds (<400 Da). The burn from hot peppers and mustard (capsaicin and allyl isothiocyanate, respectively), the coolness from mint (menthol) and the tingling/stinging of carbonation (carbon dioxide) are

examples of chemesthesis (Green 1996). Mouthfeel sensations, such as astringency, can also be considered part of chemesthetic sensations. The first reference to « a common chemical sense […] as distinct and well defined as smell and taste » was described in 1912 (Parker 1912), and was later coined as trigeminal response until being renamed chemesthesis. A more general term was preferred because nerves other than the trigeminal nerve are responsible for this sensation (i.e., glossopharyngeal and vagus nerves for oral chemesthesis) (Rentmeister-Bryant and Green 1997). Chemesthetic sensations result from the stimulation of receptors usually associated with pain (nociceptors), thermal perception (thermoreceptors), and touch (mechanoreceptors). These receptors are primarily located in the oral, nasal and ocular mucosae (Rentmeister-Bryant and Green 1997).
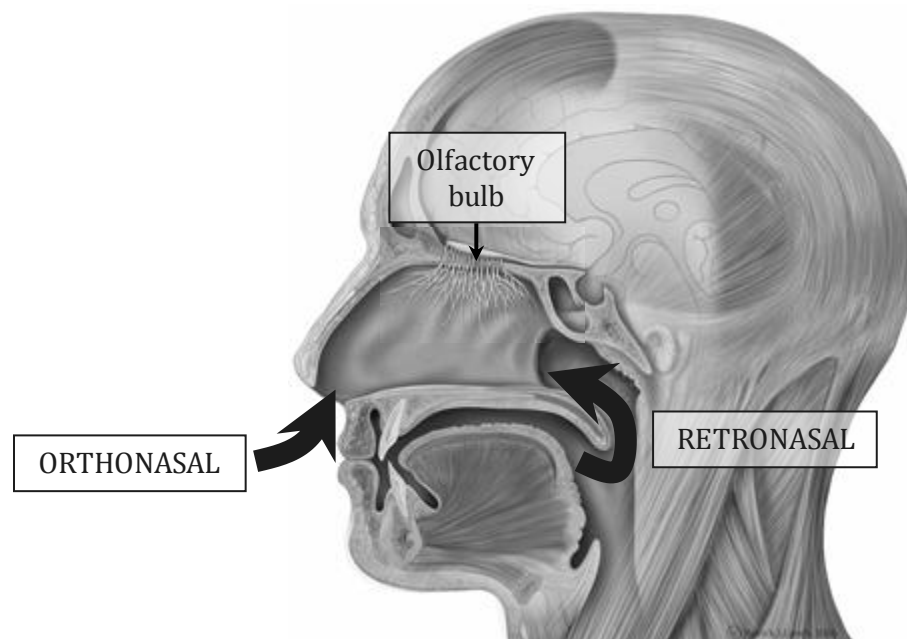
Figure 2.2: Schematic representation of orthonasal and retronasal pathways (adapted from Lynch 2006).

2.1.1.2. Factors influencing flavor perception

The perception of flavor results principally from the stimuli of chemical compounds released from the food matrix upon food consumption and transported to the sensory receptors in the nasal and oral cavities (Overbosch et al. 1991). Flavor release has been extensively studied in the last decades. It is well documented that physiological factors (breathing, mastication, salivation (Harrison 1998; van Ruth and Roozen 2000)), physicochemical properties of the aroma compounds (van Ruth et al. 2000), and food matrix (composition, texture, viscosity (Pangborn and Szczesniak 1974; Bakker et al. 1996; Mestres et al. 2005; Boland et al. 2006)) affect flavor release, and consequently flavor perception.

Flavor perception is also influenced by the interaction between compounds contributing to the same (within-modality interactions) or to different senses (cross-modality interactions). These interactions affect the intensity and the flavor profile of the food. However, it is still unclear how each sense interacts with other sensory attributes of food. Aroma compounds in a mixture can affect the volatility (release) of each other or compete for receptor sites (Laing et al. 1984; Rospars et al. 2008). Consequently, even if chemical stimuli are present at sub-threshold levels they can have an impact on the perception (e.g., additive effect) (Buttery 1999; Labbe et al. 2007). Similarly, the perception of taste compounds in mixture is either enhanced or decreased, depending on the combination and concentration of the compounds (Breslin 1996). Many studies have looked at interactions between taste and aroma compounds. It is generally recognized that the effect of taste-aroma interactions on perceived intensity (enhanced or not) varies with the type of compounds, their combination (Noble 1996) and their congruency (Dalton et al. 2000). For example, the perception of fruity flavor can be increased in presence of sweet taste compounds rather than salty as they are usually associated with sweet taste. The relevance of such interactions in food applications is illustrated by the work from Davidson et al. (1999), who showed that the perceived mint flavor of chewing gum in time followed the release of sucrose rather than menthone (minty). Such findings are of particular interest when developing food products with controlled flavor release or, when adding or removing

non-volatile compounds (e.g., sodium chloride reduction). Interactions between tastants, odorants and irritants have been also reported (Prescott and Stevenson 1995).

2.1.2. Measuring flavor perception

Flavor perception of a food is the sensation formed upon stimulation of our senses while drinking or eating. There are many methods traditionally used in flavor research to study this complex phenomenon; they include instrumental and sensory analyses, and are either targeted at the food matrix itself, at the mechanisms of food consumption or at the final sensation perceived (Figure 2.3).
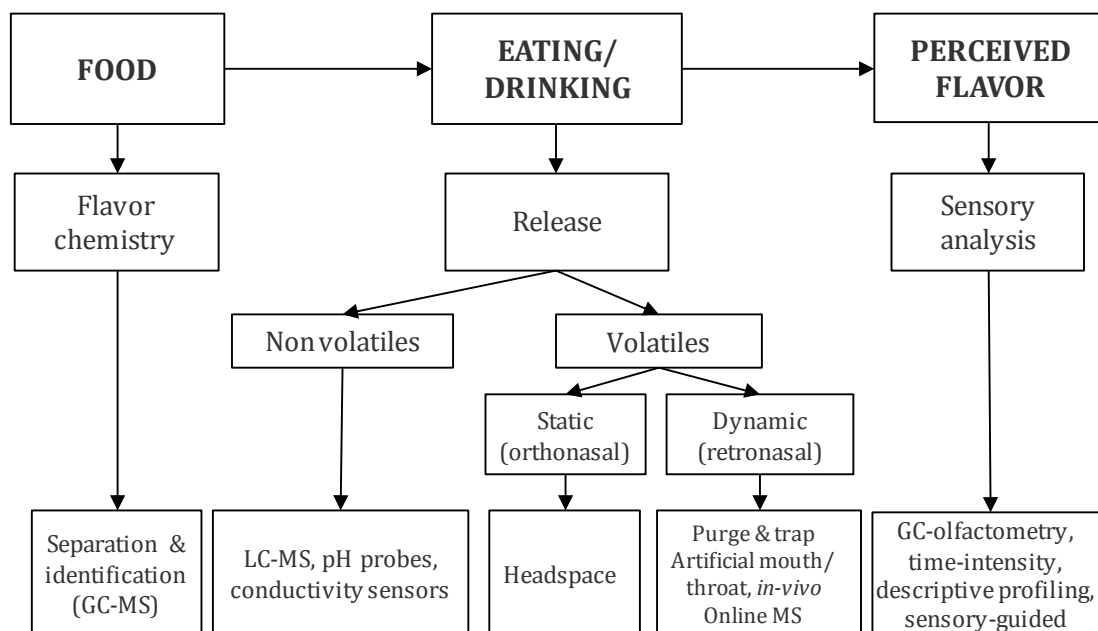
Figure 2.3: Overview of the methods traditionally used to investigate flavor perception.

2.1.2.1. Sensory evaluation

Sensory methodologies have been largely used by flavor scientists to study flavor perception (Piggott et al. 1998) since they used human subjects as the measuring

instrument, which is the most "natural" assessment for flavor. There are roughly two categories of sensory techniques used for flavor analysis: "true" sensory and "instrumental sensory" techniques.

The "true" sensory techniques are based on the human's assessment (perception) to explain the flavor and use either trained or untrained panelists. Among these, two methods commonly used are descriptive analysis and time-intensity profiling (Pangborn 1981). Descriptive analysis of food flavor is used to characterize the flavor profile of a food or to evaluate a food for given attributes; for example, in storage studies to monitor the changes in flavor and the occurrence of off-flavors over time. The time-intensity profile methodology is used to follow the changes in perceived flavor intensity while a food is consumed (Guinard and Marty 1995). Its use has expanded with the development of computer-assisted systems that allow recording and responding rapidly to changes in intensity perceived by assessors.

Instrumental sensory techniques use the human ability to perceive flavor as a part of an instrumental setting. One common practice in flavor research involves gas chromatography-olfactometry (GC-O). This technique uses a human assessor to describe perceived odors of GC effluents ("sniffing") thereby providing some perspective of the sensory character of a given odorant (van Ruth 2001). Other examples of combined instrumental/sensory analyses are given later in the text.

One advantage of sensory analyses is that, generally, food is evaluated as a whole and thus the contributions and interactions of all stimuli are considered (unless a food model is used). Although human assessments reflect reality better than instrumental methods, sensory analysis has large variability inherent to humans (e.g., physiological differences, emotional experience with the food such as previous exposure) (Brown and Wilson 1996), is costly, time consuming in preparation, and seldom provides a chemical knowledge of the reason for the sensory output. These considerations drove food scientists to investigate alternatives to sensory analyses such as instrumental analyses which provide faster analysis, extensive usage, greater repeatability, sensitivity and objectivity. In contrast,

physical stimuli (mouthfeel, texture, etc.) and interactions between senses are often ignored.

2.1.2.2. Instrumental evaluation

Early research (1960-1970s) on flavor perception consisted of the analysis of aroma compounds in foods. Most of the studies were focused on the isolation, separation and identification of the key odorants (gas chromatography-olfactometry and aroma extract dilution analysis). However, these studies lacked the ability to study relationships within and across chemical stimuli (taste and aroma compounds) and generally did not consider a chemical compound as being important to perception unless it made a contribution alone.

A few decades later, flavor researchers placed greater emphasis on understanding how chemical or physical stimuli influence the flavor of foods. They started investigating the parameters influencing the release of aroma compounds from the food matrix and their associated effects on sensory perception. The role of major food components (lipids, proteins, carbohydrates) and other lesser abundant components (acids, bitter compounds) on flavor were initially studied by static headspace analysis (Druaux et al. 1998). Later, dynamic studies were used to account for the fact that flavor perception is temporal (i.e., it changes continuously as the food is consumed) and also for resistance to mass transfer. Devices designed to simulate the human mouth or the process of eating, often termed artificial mouths or throats, were developed (Lee 1986; Roberts and Acree 1995; Weel et al. 2004) to study flavor release and also to avoid the variability of human assessors. They have shown good sensitivity, reproducibility and often produced similar release profiles as human assessors (Roberts and Acree 1995; Deibler et al. 2001). In parallel, advances in instrumentation provided additional tools to investigate flavor release and consequently flavor perception. For instance, instruments such as Atmospheric Pressure Chemical Ionization - Mass Spectrometry (APCI-MS) and Proton Transfer Reaction - Mass Spectrometry (PTR-MS) have been used to follow the real-time release of aroma compounds from a food matrix to the oral or nasal cavities (Lindinger et al. 1998; Taylor et al. 2000).

It is only lately (1990s) that the influence of non-volatiles on the perception of flavor has been investigated with greater interest. It is possible that the interest given to the participation of non-volatiles to flavor was minimized earlier because they were mostly thought as taste components or precursors of flavor. However scientists long knew that non-volatiles affected the volatility of the aroma compounds but most of the flavor researches tended to neglect their impact in terms of flavor perception. Further, their studies were generally limited to major non-volatile compounds such as proteins, lipids and carbohydrates while neglecting the potential contribution of others (e.g., polyphenols and organic acids). The use of liquid chromatography and sample preparative techniques such as solid phase extraction, mostly reserved to taste research until then, are valuable additional tools for flavor research.

2.1.2.3. Combining instrumental and sensory evaluations

Instrumental-sensory correlations in food science have been used since the late 1970s (Aishima and Nakai 1991) with the objective of understanding better the relationships between the chemical compounds themselves and the sensory properties of the food. Using such approach allows to simultaneously study the influence/ interaction of several chemical stimuli (e.g., odor and taste compounds) and their direct impact on flavor perception. Many works have been published on this topic and included various types of foods like cheeses (Lawlor et al. 2003; Biasioli et al. 2006), baguettes (Quilez et al. 2006), tomatoes (Tandon et al. 2003; Abegaz et al. 2004), carrots (Kreutzmann et al. 2008), coffee (Lindinger et al. 2008) and orange juice (Shaw et al. 1999; Tenenhaus et al. 2005), among others.

Studies combining instrumental-sensory evaluations to study flavor perception can be roughly divided in two groups: mechanistic and predictive studies. The first group (mechanistic studies) encompasses research aimed at understanding the mechanisms of flavor perception (not at the molecular level) and how the different components of the food interact to form flavor. Most of these studies focused on studying taste-aroma or texture-aroma interactions and their relation to flavor perception (Taylor et al. 2003; Koliandris et al. 2008). For instance, the simultaneous release of aroma and taste compounds from foods during eating was studied by coupling real-time mass spectrometric techniques (APCI-MS,

PTR-MS) with saliva sampling (swabs, conductivity sensors, spit out), and was related to time-intensity profiles (Davidson et al. 1999). Another example was the development of a Multichannel Flavour Delivery System (Hort and Hollowood 2004) which allowed the delivery of controlled amounts and combinations of aroma compounds and tastants, and simultaneously measured resulting variations in the perceived intensity of flavor. In this manner, interactions between senses were observed. The second group of studies (predictive studies) combining instrumental-sensory measurements to study flavor perception includes research which attempts to build mathematical models from correlations between instrumental and sensory data so that they can be used later for predictions of the sensory properties of food such as texture or flavor. For example, de Wijk et al. (2003) demonstrated a good prediction of creaminess of vanilla custard desserts based on flavor/taste and tactile oral sensations (e.g., fattiness, thickness and astringency). In another study, the sensory perception of sulfur attributes in soft cheeses was successfully correlated with flavor release parameters using principal component analysis (Salles et al. 2003). Of note, most of these studies included only volatiles in the mathematical models. And if non-volatiles measurements were included, they were limited to "traditional" physicochemical measurements (pH, titrable acidity, °Brix, etc.). Other researchers used the predictive models as classification tools (i.e., to predict the sample's class); for instance, to determine the authenticity of a food product (e.g., "A.O.C" controlled designation of origin).

One challenge of combining instrumental-sensory data is that the type of results obtained from sensory and instruments have very different formats. Consequently, appropriate mathematical and statistical techniques are needed to interpret the data with a biological meaning. These techniques include linear regression, correlation and multivariate analysis (Williams 1994; Lindinger et al. 2008). Modeling is another means to relate sensory and instrumental data. In the past, some mathematical models have been developed to predict flavor release from foods (Harrison et al. 1997) but they become limited with complex foods. Indeed, foods used for the development of those models were relatively simple in comparison with most real food systems and ignored interactions between food matrix components. Finally, the expansion of chemometrics (presented with more details later in

the text) has provided new techniques that are able to deal with numerous variables simultaneously and data with different formats.

Chemometrics (multivariate analyses) is being used increasingly in flavor research either for QC or research purposes. There are mainly three categories of flavor studies that use chemometrics: pattern association, prediction and classification (Marsili 2007). The first category (pattern association) is used to reveal groupings in the data based on patterns seen in the instrumental data. Hierarchical cluster analysis, factor analysis and principal component analysis (PCA) are the most commonly employed. Other alternatives are also being developed: recently, correlation maps and cluster analysis were used to identify a compound responsible for the fermented off-flavor in coffee by investigating patterns and associations between instrumental measurements off coffee headspace and the sensory description (Lindinger et al. 2009). The second category (prediction) uses regression techniques such as partial least squares (PLS) to develop models from measurements that can be used to predict specific food properties. PLS has been successfully used in food science to correlate odor profile and sensory attributes of cheese (Biasioli et al. 2006; Cabezas et al. 2006), to predict sensory quality of watermelon (Tarachiwin et al. 2008) and aroma properties of wine (Aznar et al. 2003; Cozzolino et al. 2008), among others. The last category (classification) uses chemometrics as a tool to assign samples to predefined classes responding to a given criterion. For instance, to determine if a sample is genuine or adulterated (olive oil, wine). The reader is referred to Aishima and Nakai (1991) and Marsili (2007), and references therein, for a more detailed discussion.

2.1.2.4. Other types of studies on flavor perception

Instrumental and sensory analyses provide only insights on the chemical and physical attributes of a food and subsequent sensory properties. Studies on brain imaging, molecular biology and genetics have also contributed to the understanding of the biology of flavor perception (Lindemann 2001; Max et al. 2001; Shepherd 2006). However, we are still far from understanding the mechanisms involved in the integration and transduction of the stimuli responsible for flavor perception. As well, psychophysical and behavior studies (e.g., influence of culture on flavor perception) are needed.

2.2. COMPREHENSIVE EXPLORATORY RESEARCH

The technological development of instruments (i.e., mass spectrometers) over the last several decades has lead to a significant increase in the amount and complexity of data generated when conducting research on biological systems. However, most researchers are often looking at a one-to-one relationship between a cause and a response. Such approaches become limiting when studying complex systems such as food as they neglect the influence/ interaction between different components of the system; and discarding excessive data based on *a priori* hypotheses (targeted approach) ultimately leads to major loss of information.

Instead, comprehensive research (also referred to as non-targeted) is more exploratory in a sense since they consider all of the data as valuable. Rather than targeting a few compounds, variations in the relative amounts of "all" compounds are recorded and analyzed thereby giving an unbiased view of the system under study (e.g., food product). When using appropriate mathematical tools to explore the data (correlations, groupings), there is a great potential to extract new information which otherwise would have been ignored. The following section introduces the concept of exploratory research within the frame of comprehensive research and is followed by a presentation of metabolomics research as an example. Then, analytical and mathematical tools (chemometrics) commonly used in comprehensive research are exposed.

2.2.1. What is exploratory research?

Exploratory research is a very controversial topic among the scientific community because of the question of how to approach a scientific question (see Kell and Oliver 2004). Exploratory research, sometimes referred as principled data-driven research, differs in essence from classical "hypothesis-driven" research. It does not intent to approve or disprove *a priori* hypotheses but instead it generates hypotheses as an outcome of the study (Kell and Oliver 2004). One definition of exploratory research was given by Grove and Andreasen (1982): «It [exploratory research] examines a large number of variables in an

attempt to discover whether any of them have a meaningful relationship to a particular problem». Exploratory research is a powerful tool to explore new research paths without being limited by previous knowledge, which often times drives the line of research in a given direction. This approach provides a more comprehensive view to a research question in the sense that it looks simultaneously at the systematic variations of many variables (e.g., all low molecular weight metabolites) and tries to relate them to a particular problem without prior expectation. In this way, many variables can be examined and potentially new explanatory variables highlighted.

Sample collection & preparation

Comprehensive analysis

Non-targeted data collection

Data

Experiments

Interpretation, biological understanding

Hypotheses
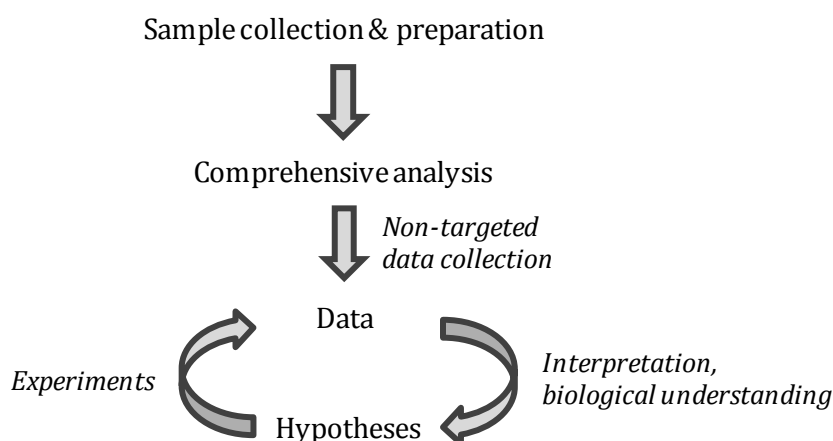
Figure 2.4: Schematic overview of comprehensive research strategy.

Exploratory research was first used in epidemiology, behavioral science and psychiatry (Grove and Andreasen 1982; Weihs 1993) but rapidly found applications in other areas; for instance, analytical chemistry where many chemical variables are measured simultaneously. It is important to stress that exploratory research is only the first stage of the research and always needs a confirmatory step (Figure 2.4). Scientific knowledge is gained through an iterative process, where the scientist refines at each step the rationality between the data and the ideas. As commented by Kell and Oliver (2004), data-driven and hypothesis-driven researches are complementary rather than competitive. Indeed, correlations between variables (e.g., chemical composition and sensory properties) are

established but they do not imply causality. So when interpreting the correlations, their relevance/significance always needs to be examined in term of biological meaning. The validity of the results and hypotheses generated has to be verified and confirmed through a classical approach (i.e., hypothesis testing).

2.2.2. Example of comprehensive research: Metabolomics

In the last years, the number of "-omics" works have been increasing in the scientific community (Omics.org 2008) and their emergence has been possible only because of concurrent advances in analytical chemistry, computer science and chemometrics. The enthusiasm for "-omics" is mainly because this type of research offers a holistic (comprehensive) point of view of complex biological systems, rather than being focused on a particular set of inputs. Metabolomics is one the most recent examples of comprehensive research and is briefly described in the following section since some of the ideas, techniques and tools were used as a model for this thesis.

2.2.2.1. Presentation of metabolomics

The term and concept of metabolome research were introduced a decade ago by Oliver et al. (1998) as an additional tool to understand biological systems in a more comprehensive way. The metabolome represents the highest level of organization at the cellular level since metabolites are the end products of cellular activity. Metabolomic research is considered part of the evolution of science and the end-point of the "-omics" cascade (Figure 2.5) because it complements the information given by genomic, transcriptomic and proteomic studies. Originally, genomics constituted the initial step to understand organisms by means of gene sequencing. Later, the research was directed to the study of gene expression (transcriptomics), which was then expanded to the study of proteins (proteomics) and finally to the study of the metabolome (metabolomics).

Several research approaches have been developed for the study of the metabolism of biological organisms and they are classified upon their scope: targeted analysis, metabolite (or metabolic) profiling, metabolic fingerprinting and metabolomics (Fiehn 2002). Targeted

analysis focuses on a few specific compounds; for example, this strategy can be used for screening of specific markers of genetic alterations. Metabolite profiling limits its identification and quantification to a selected group of compounds (e.g., polyphenols, amino- acids, compounds sharing pathways); this strategy is used in drug research to study the fate of drugs. Metabolomic fingerprinting does not intend to separate and identify all metabolites. Spectra from samples are compared and screened for elements that differentiate classes of samples; then, only discriminating compounds are identified and interpreted in terms of biological meaning. Finally, metabolomics is defined as "a comprehensive analysis in which all the metabolites of a biological system are identified and quantified" (Fiehn 2002). This approach is unbiased (in theory) and gives an overall vision of the biological system. As opposed to traditional research, it monitors a wide range of compounds instead of a particular group of compounds.

DNA	Genomics	What can happen?
	⇩	
RNA	Transcriptomics	What appears to be happening?
	⇩	
PROTEINS	Proteomics	What makes it happen?
	⇩	
METABOLITES	Metabolomics	What has happened and is happening?
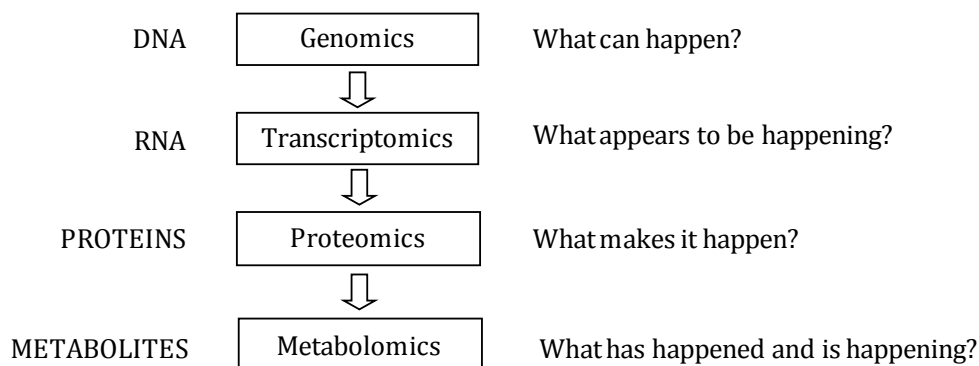
Figure 2.5: The "-omics" cascade (modified from Dettmer et al. 2007). (Copyright 2007 Wiley. Used with permission.)

2.2.2.2. Different fields of application

Metabolomics has been used in many fields of research to characterize (e.g., what are the metabolites responsible for…?), to diagnose (e.g., are there metabolite differences between the control and treatment samples?), to study changes caused by external factors (e.g., which metabolites are different between the two treatments?), and to understand the

biochemistry (e.g., how is it formed?) of biological samples. Current examples of these applications are briefly presented below; these are not limiting and are only given as examples of the breadth of metabolomic applications.

The medical field has been among the first to benefit from metabolomics. For instance, the metabolomic approach has been used to diagnose diseases like cancer by tracking biomarkers (Woo et al. 2009); or to understand the effect of mutations on metabolism (Schirra et al. 2008). Pharmaceutical research also uses metabolomics to follow metabolic response to a drug treatment, to examine the fate of drugs (catabolic degradation) or their toxicity (Wishart 2008a). Another field is microbiology, where comprehensive studies have been achieved to understand and characterize the metabolism of yeasts and bacterial strains (Mori and Begley 2008). Also, work has been conducted in plant research. For example, plant metabolomics has been used to examine differences in metabolic composition between cultivars, to study the influence of growing/stress conditions on the levels of specific compounds (e.g., to improve quality of crops) and also to investigate biological pathways (Schauer and Fernie 2006). Finally, food science and nutrition research recently started using metabolomic approaches (Wishart 2008b). To give some examples, some of the objectives for human nutrition are to identify bioactive substances that are responsible for health benefits, to evaluate the impact of diet on the metabolites formed, and to find metabolic markers of diseases like diabetes. Following these lines, in 2006 the EU research division launched a multi-national program (META-PHOR) focused on the identification of bio-active compounds in foods (i.e., broccoli, melon and rice) (Hall 2007). One example for food science is to understand the influence of parameters such as raw materials, handling/processing conditions on the final composition of a food or on key food components. This is relevant for the food industry if one wants to promote or decrease the formation of specific compounds (e.g., to enhance the levels of flavor precursors in coffee beans through breeding selection, to retain high levels of polyphenols in berries through modification of handling and storage conditions).

2.2.3. **Methodologies used in comprehensive research**

One major challenge of comprehensive research is one's ability to study all chemical compounds present in the system. The methodologies used need to cover the broad range of compounds, in terms of physicochemical properties and ranges of concentration, and yet provide a high throughput. The choice of sample preparation technique and analysis method is usually a compromise between the number of compounds extracted, the sensitivity of the method and the preparation time required. There are currently no standard methods established for metabolomic investigations because of the diversity in composition of biological samples studied (e.g., urine, saliva, blood, plasma, and tissue); similarly, food systems are very diverse in composition and the choice of analytical methods should be considered on a case-by-case basis. An overview of selected current analytical methodologies used in comprehensive research, as well as data handling follows.

2.2.3.1. Sample preparation techniques

The number of sample preparation steps should be kept to a minimum as each is a potential source of changes in the original sample and a limit in sample analysis throughput. One important consideration when preparing biological samples is to ensure their biological stability; that is, to prevent any enzymatic deterioration or oxidation which might transform or form metabolites. Freezing in liquid nitrogen, cold shock in methanol, or acid treatment are some of the techniques used to stop reactions in the sample (Fiehn 2002). Once the sample is "stabilized", another step is the extraction of the analytes from the matrix. The methodologies used for that purpose differ based on whether the compounds reside primarily in the non-gaseous or gaseous phase of the food. High throughput methods are generally preferred since they allow sample replication in analysis, an important requirement to overcome biological variability (i.e., to have sufficient statistical power to draw conclusions).

Compounds in the liquid phase (non-volatiles). When analyzing metabolites in the liquid phase (e.g., blood, urine), the removal of matrix components that can interfere with the analyses is often a required step. However, in order to be as comprehensive as possible, only salts and macromolecules (proteins, larger peptides) should be removed from the

matrix. Removing any other components would narrow the scope of investigation. Liquid-liquid extraction (LLE), solid phase extraction (SPE), protein precipitation, dialysis and ultracentrifugation are examples of sample preparation techniques. Their use, advantages and limitations have been presented elsewhere (Dettmer et al. 2007).

Liquid-liquid extraction is a commonly used technique to extract compounds from liquid and solid (tissue) samples. It is a fairly simple technique which consists of mixing the sample with a solvent; metabolites will migrate from the sample into the solvent according to their polarities. However, this technique is time-consuming and easily results in metabolite losses during the extraction steps. Also, the choice of solvent determines the compounds (proportions) being extracted. Some researchers have used a mixture of solvents to create a bi-phasic system allowing the separate analysis of both hydrophilic and lipophilic fractions (Dettmer et al. 2007).

Solid phase extraction is another sample preparative technique commonly used and is based on the partitioning of the compounds between two phases (liquid and solid). Briefly, the sample is loaded onto a cartridge that contains a solid sorbent phase where the metabolites are adsorbed while flushing away undesired compounds (or vice-versa). The retained metabolites are then washed out of the cartridge using an appropriate solvent. The diversity of stationary phases available and the ability to conduct multiple SPE simultaneously have contributed to SPE popularity in comprehensive research, particularly in metabolomics. Solid phase extraction was chosen for the research presented in this thesis.

Compounds in the gaseous phase (volatiles). There are many techniques available for the analysis of compounds in the gaseous phase, especially when looking at flavor investigation tools. One of them is solid phase microextraction (SPME). This technique has been used for many years in flavor analysis for the analysis of volatiles in the headspace (Yang and Peppard 1994; Steffen and Pawliszyn 1996) and is also gaining popularity in metabolomic investigations for profiling volatiles from plants or fruits (Tikunov et al. 2005) as well as for residue analyses (e.g., drug, pesticides, organometallics) in biological fluids such as blood, saliva and urine (Mills and Walker 2000).

Briefly, SPME is a quick (little sample preparation), easily automated, sensitive and a solvent-less method that uses a reusable coated fiber onto which volatiles are retained. The amount of volatiles extracted by the fiber depends on the equilibrium between sample/headspace and headspace/fiber; this equilibrium is compound-dependent based on vapor pressure and matrix interactions and can be reached with proper method optimization (e.g., temperature, extraction time, agitation, salt addition). Despite its common usage, this method has some weaknesses which must be acknowledged when developing methods (Nongonierma et al. 2006). Major disadvantages are related to its limited quantification abilities due to the selectivity of the fiber coating toward some compounds, the low volume of coating material that easily leads to saturation of the fiber (absorbent fibers) or competition between compounds for the binding sites on the fiber (adsorbent fibers), and less than optimum sensitivity (Li and Weber 1999; Shirey and Mindrup 1999; Nongonierma et al. 2006). Despite these limitations, we feel the preserved sample composition (no major preparation steps) and the ability to automate (and therefore, analyze large numbers of samples) override the limitations of this method, and therefore this technique was chosen for this research.

2.2.3.2. Sample analyses

Similar to sample preparation, special attention should be paid when analyzing samples so that the separation and detection of compounds is as comprehensive as possible. As stated by de Vos et al. (2006) there is « no single analytical method […] currently capable of extracting and detecting all metabolites ». There is no gold standard method and multiple methods should be used whenever possible to get the most comprehensive view of the sample being analyzed. Numerous analytical platforms have been used for metabolomic studies but only instruments relevant to this thesis study will be briefly presented: gas-chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS). Other common methods used are nuclear magnetic resonance spectroscopy (NMR), Fourier transform-infrared spectroscopy (FTIR) and capillary electrophoresis-mass spectrometry (CE-MS). These methods have been reviewed elsewhere (Dunn and Ellis 2005; Dettmer et al. 2007).

Mass spectrometry (MS) has been the method of choice in most metabolomic studies because of its ability to identify and to quantify compounds with high selectivity and sensitivity. The basic principles of MS are 1) the ionization of the molecules which leads to fragments of various mass-to-charge ratios (m/z), 2) the detection of the different m/z in the mass analyzer, and 3) the identification of the compounds based on a unique fragmentation pattern by comparison with spectral libraries, if available. There are different types of ionization available (electron impact ionization (EI), electrospray ionization (ESI), chemical ionization (CI)) and each of them result in different fragmentation patterns of the analytes. Several mass analyzers have been developed; they include (are not limited to): quadrupole, time-of-flight and ion-trap (for further details see Downard 2004).

MS can be either interfaced with a separation technique such as gas (GC) or liquid chromatography (LC) or used directly (direct injection). Direct injection analyses have been successfully used in metabolomics, for example, to obtain the metabolite fingerprinting of broccoli (Luthria et al. 2008) and beers (Araujo et al. 2005). However, interfacing a MS with a separation technique is still preferred. The main advantage of a separation technique is an easier spectral interpretation and consequently, easier compound identification. For both GC and LC, the choice of mobile and stationary phase will dictate the chromatographic resolution, and unquestionably bias the compound profile obtained. The use of GC is preferred for volatiles or compounds that can be easily derivatized into volatiles; note that derivatization implies additional sample preparation steps and the possibility of modifying the original sample. LC is chosen for all other compounds (higher molecular weights, non-volatiles); however, it is not the best for polar compounds. Normal phase and more recently, hydrophilic interaction chromatography (HILIC) have been used to fractionate highly polar compounds (Tolstikov and Fiehn 2002). LC analyses provide poorer chromatographic resolution than GC but the recent implementation of ultra high performance liquid chromatography (UHPLC) has greatly improved the results. UHPLC columns are packed with smaller particle sizes and work at higher pressures (10,000-15,000 psi). UHPLC runs are shorter, use smaller sample volumes, provide better resolution and consequently detect more peaks than conventional HPLC (high performance liquid chromatography) (Swartz

2005). UHPLC has been used in many metabolomic studies (Grata et al. 2008; Bruce et al. 2009).

2.2.3.3. Data handling

Another major challenge for non-targeted comprehensive studies is data handling due to the amount and complexity of the data generated, and how to extract relevant biological information from it (e.g., from the chromatograms generated from GC-MS and/or LC-MS). Data analysis is as important as the sample analysis itself and great care should be placed when treating the data. Overall, data analysis can be divided in two steps: data pre-processing and data processing; these have been reviewed with greater details in the literature (Katajamaa and Oresic 2007; Sumner et al. 2007).

Pre-processing of the raw chromatographic data is a critical step before comparing samples and is done with specialized software packages. The commercial and free packages currently available were reviewed by Pierce et al. (2008). With non-targeted data analysis, the main objective is to convert the instrumental data into organized matrices (list of all mass spectral signals detected during the entire chromatographic runs across samples and their intensity) in order to be explored with mathematical tools.

Pre-processing of the data includes noise and background reduction, chromatogram alignment, deconvolution of co-eluting peaks, and peak picking (Sumner et al. 2007). The data matrix obtained contains information on retention time (RT), associated m/z values (nominal or accurate) and corresponding intensity. RT-m/z pairs are referred to as variables or features and both terms will be used inter-changeably in this thesis. All RT-m/z detected by the software will be extracted for each chromatogram (sample) and compiled in a table with y rows (samples) and x columns (variables). Following this, some mathematical transformations (scaling, normalization, transformation) may be required so that the biological meaning of the data is improved. It has been shown that the choice of the data treatment will greatly affect the results (and conclusions) and therefore should be carefully decided based on some biological criteria (van den Berg et al. 2006). For instance, differences in concentration between compounds are not necessarily proportional to their biological relevance. Logarithmic transformation of concentrations has been mentioned

when dealing with chemical compounds and sensory perception, arguing that concentrations are not necessarily proportional to their sensory relevance. This suggestion is supported by Steven's power law which states that the perceived intensity of a given stimulus grows as a power function of the intensity of the stimuli (concentration) and, therefore, is not linear in magnitude (Stevens 1957; Dravnieks 1976; Meilgaard et al. 2007a; Lindinger et al. 2008).

Following data pre-processing, the data matrix is analyzed using chemometrics tools (see next section for a presentation of the mathematical tools used). At this point, the chemical identity of the variables is unknown and unnecessary. It is only in later stages of the research that efforts are made to identify compounds of interest.

2.2.3.4. Identification of markers

After performing mathematical and statistical analyses, the last step is the chemical identification of the (mass spectrometric) signals that were determined statistically different between sample groups/treatments or that contributed the most to the mathematical models. Mass spectral libraries, retention time indices, metabolite databases and knowledge about samples are tools typically used for the tentative chemical identification of an unknown compound (Brown et al. 2009). Ideally, the definitive identification is confirmed by analyzing authentic standards (if available) or by NMR spectroscopy. Note that libraries for GC-MS are more developed than those for LC-MS because EI-MS spectra are more reproducible and have been collected for many years (Halket et al. 2005). One strategy to overcome this limitation has been the use of high mass-accuracy mass spectrometers. Accurate mass measurement (5 ppm) and associated isotope ratios are used to calculate the elemental composition of the unknowns, from which a list of probable compounds is generated. In addition, tandem-mass spectrometers can be used to provide additional fragmentation (ESI does not provide enough fragments with structural information) and thereby to facilitate the structural elucidation. They are also used for quantification purposes because they provide high specificity and sensitivity (Dettmer et al. 2007). Finally, it is important to note that it is most often not possible to have simultaneous identification and absolute quantification of the compounds of interest. In fact, the sample

preparation and characterization techniques were initially designed to be as comprehensive as possible. For subsequent experiments, they need to be revised and directed toward the specific analysis of the compounds of interest (targeted analysis).

2.3.   CHEMOMETRICS: A KEY TOOL IN COMPREHENSIVE RESEARCHES

The word "chemometrics" derives from the latin *chemo-*, meaning chemistry, and the greek *-metrics*, meaning measure, and was first coined by Wold in 1974 as « the art of extracting chemically relevant information from data produced in chemical experiments […]; in chemometrics the main issue is to structure the chemical problem to a form that can be expressed as a mathematical relation » (Wold 1995). By extension, the term "chemometrics" is used to refer to the set of mathematical and statistical tools used to interpret chemical measurements made on a system (e.g., biological) (Frank and Kowalski 1982). The emergence of chemometrics in the 1970's was favored by the increased usage of computers and automation of data collection in scientific research. As the instruments were developed (e.g., mass spectrometer), the amount and complexity of data generated scaled up. A very simple example would be: if a GC-MS is set to scan m/z 40 through 280, this is 240 points collected per spectra recorded, which is 240*5=1200 points per second if there is 1 spectra recorded every 0.2 sec; for a GC run of 25 min, this is the equivalent of 1200*60*25=1,800,000 points collected per sample! Data became rapidly overwhelming and the limited "human" ability to process them pushed toward the use of chemometrics for data visualization and interpretation. Their ability to identify systematic patterns of variation among measured variables make chemometrics (most particularly multivariate analyses) a perfect fit for the evaluation of data such as the ones presented in this thesis.

2.3.1. Methods commonly used: PCA and PLS

Multivariate techniques are part of the chemometrics toolbox and constitute primarily a graphical tool to visualize the entire data set and to highlight any correlations between variables. They are commonly used in comprehensive research because of their ability to

extract the most relevant information by reducing the massive amount of instrumental data (NIRS, UV spectroscopy, NMR, MS) into visual graphics, thereby making their interpretation easier in terms of biological meaning (Trygg et al. 2007). In fact, data generated by comprehensive research cannot be analyzed through conventional univariate statistics due to the nature of the data. In contrast, multivariate methods can handle incomplete, noisy, and collinear data (variables are not independent); further, conventional assumptions (normality and variance homogeneity) are not necessary (Eriksson et al. 2006a). As stated previously, data pretreatment may be necessary before applying multivariate techniques in order to prevent significant variables with low abundance to be neglected in favor of insignificant high abundance variables (Sumner et al. 2007). The use of software packages and computer power are critical for data handling; such calculations would be highly time-consuming and tedious if done otherwise.

The most common multivariate techniques used in comprehensive research are principal component analysis (PCA) and partial least-squares regression (PLSR) (Sumner et al. 2007). The following paragraphs do not intend to provide a statistical presentation of those techniques but rather to provide the reader with basic knowledge.

2.3.1.1. Principal component analysis (PCA)

PCA is the basis of multivariate data analysis, and is often used as a diagnostic tool before applying other techniques. It provides a graphical overview of the variation in a data matrix X with N rows (observations) and K columns (variables). This is useful to understand the structure of the data but also to detect any outliers. PCA is referred as an unsupervised method because no additional information (e.g., sample type) besides the raw data is required for analysis. The relationships between observations and variables, and between variables themselves, are revealed through plots: a score plot and a loading plot. The score plot represents the grouping of observations (samples) while the loading plot assists in identifying the influential variables (e.g., signals measured by the instrument) responsible for the patterns seen in the score plot (Eriksson et al. 2006d; Trygg et al. 2007). Examining the score plot reveals how observations are related to each other (similar samples will be nearby) and also allows detecting any outliers. One advantage of PCA is its ability to work

well with any type of matrix (more rows than columns, more columns than rows, etc.) which makes it a technique of choice for any multivariate study (Eriksson et al. 2006d).

2.3.1.2. Partial least-squares regression (PLSR)

PLSR, also known as projection to latent structures, combines features from PCA and multiple regression. PLSR is a modeling method that finds a linear multivariate model to link two data matrices, X (predictor variables) and Y (dependent variables), to each other. For instance in this thesis, it was used to search for a quantitative relationship between the data matrix X (intensities of recorded mass spectral signals of the set of samples) and the data matrix Y (associated scores of sensory attributes). Once the mathematical relationship is established, the model can be used for predictions of new samples (i.e., to predict sensory scores using instrumental inputs). For example, PLSR has been successfully used in food science to correlate odor profile and sensory attributes of cheese (Biasioli et al. 2006), to predict sensory quality of watermelon (Tarachiwin et al. 2008) and aroma properties of Riesling wine (Cozzolino et al. 2008).

An important fact to consider when building the mathematical model (also called calibration in analytical chemistry) is the proper selection of samples used. The selection of representative and diverse samples is crucial and can be achieved either through statistical experimental design (factorial, D-optimal, mixture designs, etc.) if samples can be prepared with desired varying levels of constituents, or proper sampling strategy to cover at least the expected range of variations if the experiment does not allow adequate experimental design (Eriksson et al. 2006c; Trygg et al. 2007). The advantage of the first situation (mixture design) is that a large range of properties can be covered with few samples (15-20), whereas the sampling strategy requires collecting many samples (50-100) to ensure that any of the components that influence the data matrix Y (flavor quality) is spanned over a wide range. Typically, three subsets of samples are considered: calibration, validation and prediction set. The calibration set is used to generate models, the validation set is used to select the best model, and the prediction set is to test the applicability of the selected model (Kalivas 2009). Samples included in the prediction set are not used to form/ select the model.

One advantage of PLSR is its ability to model together several response variables Y rather than having separate models for each response. This offers an easier understanding of the overall system and also a stronger model when the responses are strongly correlated between them (Eriksson et al. 2006e). Another advantage of PLSR is that it works well with matrices with more columns (variables) than rows (samples) and therefore is well suited for the application described in this thesis. A shortcoming of PLS models is that they are negatively affected by systematic variation in the X matrix that is not related to the Y matrix (Trygg et al. 2007); for example, variation that is specific to the instrument (noise, temporal drifts). This affects the interpretation of the models but also the selection of markers. OPLS and O2PLS are two recent modifications of the PLS technique that separate the unrelated systematic variation from the data set; their use have been limited and mostly deal with two class studies (control vs. treatment samples). Thus, they were not considered in this thesis.

2.3.2. Multi-blocks

Multi-block techniques are emerging methods used to combine several blocks of data together, they are mostly extensions of the projection methods PCA and PLS. They are used when the number of variables is large and the dataset can be divided into meaningful blocks. For instance, in comprehensive studies where several instruments are used to characterize the same set of samples: data collected from different instruments constitute different blocks. One example is the study of Kreutzmann et al. (2008) who applied multi-block PLS to predict the sensory quality of carrots based on measurements on the dry matter content, non-volatiles and volatile compounds.

Several algorithms have been developed to process multi-block data; the best known are multi-block PLS, consensus PCA, hierarchical PCA and hierarchical PLS. The differences between the methods (multi-block vs. hierarchical) are related to the way data are normalized and to which data are used for the regression model (for further details see Westerhuis et al. 1998). The main advantage of multi-block regression methods is in the interpretation of the models compared to single block models (all variables as one large block) because relevant information specific to each block is extracted (e.g., nature of the

relation of the block with Y and its contribution to the response Y). However, there is no improvement in the predictive performance if variables are divided in blocks (Westerhuis and Smilde 2001).

Prior to applying multi-block techniques, there are several pre-processing steps to consider. For instance, applying a weighting factor to each block to prevent those with larger variance (because of larger numerical range of values given by the instrument) to dominate the others and to lead to erroneous conclusions. Also, a scaling factor may be used to adjust for the differences in size between blocks (in terms of number of variables) as larger blocks are not synonyms of higher importance for the model (Smilde et al. 2005). Another pre-processing step occasionally used when merging (fusing) blocks of data is variable selection, meaning reducing their number to those most influential. This practice is also referred as "deflation", and is mostly used with hierarchical PCA and PLS, where the most important variables in the initial model (base level) are used to build a second model (top level) thereby simplifying the model to a few explanatory variables (Westerhuis and Smilde 2001). However, deflation of the matrix X (predictor variables) can lead to inferior prediction of Y.

2.3.3. Remarks on variable importance, correlations and model validation

This section contains supplementary information relevant to the analysis of multivariate data that need to be taken into account when dealing with such data.

2.3.3.1. Variable importance

One problem commonly faced when performing multivariate data analyses is whether or not to keep all of the variables and how it impacts model quality (fit and prediction). Two diagnostic tools are most commonly used to decide on the relevance of a variable to the model: the regression coefficients and the Variable Influence on Projection parameter (VIP). PLS regression coefficients are useful to interpret the model especially if there are several responses and give a summary of the X-Y relations of the PLS model. For each Y-variable, there is a coefficient plot showing its relationship with the X-variables. The absolute value

and sign of the coefficients are indicators of the influence of each model term (X-variables). One disadvantage of using regression coefficients over PLS weight plots for interpretation of the model is that information about the relationship between the responses (Y's) is not conveyed (Eriksson et al. 2006e). The Variable Influence on Projection parameter gives a simplified overview of the overall contribution of each X-variable to the PLS model, pooled over all components and Y-variables. There is only one set of VIP values for the entire model (one VIP per X-variable). Variables with a large VIP are the most influential on the model, generally a cut-off value VIP> 0.7-0.8 is used to retain only the important X-variables (Eriksson et al. 2006e).

2.3.3.2. Correlations

When using multivariate techniques, one has to be careful about the interpretation of the results as it is very easy to confuse correlation and causation. It is important to understand that multivariate methods are only mathematical tools that are looking for patterns of variations/associations between variables (correlation) and they do not evaluate in any way the causality of the relationship found (Eriksson et al. 2006a; Kjeldahl and Bro 2010). For instance, a positive correlation between variable 1 and 2 means that when one changes, the other one changes in the same direction. Concluding that variable 1 causes the variation in variable 2 is wrong. Both variables may be linked to a third variable that causes changes seen in 1 and 2. Causation and correlation are fundamentally different (Kjeldahl and Bro 2010). A causal relationship implies that one variable (factor) is responsible for the change in another variable (response). This fact introduces the next paragraph which presents the necessity to validate a model not only from a mathematical point of view, but most important, from a biological point of view.

2.3.3.3. Model validation

Validation (or confirmation) is a required step (often overlooked) when conducting exploratory research, it is done to ensure that correct hypotheses are generated and to prevent erroneous conclusions (Andersson 2000). The validation process includes 1) verification of the predictive validity of the model by applying it to new samples and testing

the fit, and 2) evaluation of the interpretation validity of the model in terms of biological/chemical meaning.

The predictive ability of a model is estimated by the statistic Q^2 which measures how well the data can be predicted for samples not used in the model determination (Eriksson et al. 2006e). Of note, the calibration set (samples used to build the model) must cover all of the variations expected in X-data so that the model is valid over a range of conditions normally encountered. The sum of squares of the differences between the predicted and observed values are used to compute the predictive power of the model PRESS (Predictive Residual Sum of Squares), which is used to express Q^2. External validation is the best way to validate a model and consists of computing predictions for "new" samples (prediction set: not used to build the model) but it is also the most demanding because of extra time/cost needed for the analysis of additional samples. Therefore, internal validation, also referred as cross-validation (CV), is mostly used. CV consists of iteratively fitting the model for only part of the samples (calibration set) and keeping a portion outside (validation set) until all samples have been left out once. Other statistics used to evaluate the predictive ability of the model include RMSEP (Root Mean Squared Error of Prediction) and RMSECV (Root Mean Squared Error of Cross-Validation): RMSEP is calculated from an independent data set (external validation) whereas RMSECV is calculated from samples within the data set (no additional samples are available) (Faber and Rajkó 2007).

After finding the most influential variables in the model and generating hypotheses from the results, the interpretation validity of the model is assessed. The hypotheses resulting from the study need to be evaluated based on prior knowledge and examined for their coherence with previous research. As well, the design of targeted studies is necessary to prove the biological/ chemical significance of the variables for the system under study. For the present thesis, the causality of the correlations established between chemical compounds and sensory attributes will require subsequent sensory validation in order to ensure the causal nature of the relationship.

2.4. CITRUS FLAVOR (ORANGE AND MANDARIN)

This section provides information on citrus from a general aspect and from a flavor chemistry point of view; details about citrus origin, botanical, cultivation and technological practices are not discussed here and can be found elsewhere (Ramana et al. 1981). Although this thesis was focused on mandarin juice flavor, most of the discussion deals with orange juice since it has been the most studied given its greater consumption (and economic importance).

2.4.1. Citrus market

The world citrus market is principally divided between sweet oranges (*Citrus sinensis*), mandarins and tangerines (*Citrus reticulata*), lemons and limes (*Citrus limon and Citrus aurantifolia*), and grapefruits (*Citrus paradisi*) with 51.4, 20.3, 6.3 and 5.2 million metric tons, respectively, of global production in 2010 (USDA Foreign Agricultural Service 2011). In 2008, the three top citrus producing countries were China, Brazil and USA (USDA Foreign Agricultural Service 2008).

Orange juice dominates, by far, the citrus juice industry: last year (2010-2011), over 11 billion liters (~ 3 billion gallons) of orange juice were consumed world-wide, of which 35 % was consumed in the USA (USDA Foreign Agricultural Service 2011). In contrast, mandarin juice consumption is small since mandarins are mostly eaten as a fresh fruit. For instance, less than 6 % of its total production is processed as juice, concentrate, canned in syrup or blended. One of the reasons for the limited mandarin juice importance is its economical viability compared to orange juice: the yield of juice per extractor is about half that from sweet oranges (Ranganna et al. 1983b). Also, the juice is unstable to storage if an excessive amount of peel oil is extracted along with the juice. However, the market of tangerines is growing due to an increase in demand from the EU and Russia over the last few years while grapefruit and lemon/lime production holds steady (USDA Foreign Agricultural Service 2008; USDA Foreign Agricultural Service 2011).

The term "tangerine" has no botanical meaning *per se* and is generally used by horticulturists to refer to a class of mandarins that are easier to peel and have a deeper orange color like Dancy tangerines (Ranganna et al. 1983a), nevertheless people incorrectly use the terms tangerine and mandarin interchangeably. The taxonomic classification and phylogeny of citrus cultivars is relatively complicated and will not be discussed here. Mandarins are divided in four classes: Satsuma mandarins, Mediterranean mandarins, King mandarins and Common mandarins (Fake 2004). Independent of the taxonomic system used, the "common mandarin" is *Citrus reticulata* Blanco and, because of the long history of mandarin cultivation, many cultivars have been developed either through natural cross-pollination or controlled hybridization. Due to their diversity in genetic background and consequently in flavor genesis pathways, hybrids have marked differences in flavor.

Mandarin varieties employed for processing differ between countries, for instance the varieties Dancy, Orlando, Robinson and Nova are used in the US whereas Satsuma, Clementina and Common mandarins are used in Spain (Ranganna et al. 1983b). In the US, mandarin fruits commercially available include the early-maturing mandarins (September-December): Robinson, Fallglo, Fairchild and, medium to late-maturing mandarins (December-March): Dancy tangerine, Orlando tangelo, Minneola tangelo, Ortanique, Temple tangor, Murcott (Honey tangerine), and Kinnow (University of California Riverside 2011; University of Florida IFAS Extension 2011).

2.4.2. Juice composition

Citrus juice composition depends on many variables such as the maturity of the fruits, growing and processing conditions and, of course, fruit variety. The following is an overview of orange juice composition (and mandarin if information was available); more details on the chemical compounds involved in flavor are given in the next section.

Orange juice is composed of two heterogeneous phases. The first phase is a clear aqueous phase, referred to as serum. It represents about 95 % (w/w fresh) of the juice and contains all water soluble compounds such as sugars, acids, soluble vitamins, pectins, salts and soluble flavors. The second phase is called cloud (< 1 % w/w fresh) or pulp (~ 4 % w/w

fresh) and contains fine (< 2 μm) and larger (> 2 μm) insoluble particles, respectively (Brat et al. 2003). These particles are non-wall proteins, cell wall materials (e.g., pectin, cellulose, hemicellulose), lipids, minerals and phenolic compounds (Mizrahi and Berk 1970).

The overall composition of fresh orange juice is given in Table 2.1. The main constituents are water, sugars and acids. The fraction of non-volatiles and volatiles in the juice represents ca. 12.5-13 % and 0.010-0.015 % (w/w fresh weight), respectively (Ahmed et al. 1978a; Brat et al. 2003).

Carbohydrates (mono- and disaccharides) are the most abundant class of soluble solids in the juice: about 8.5-9.8 % and 11% in orange and mandarin juices, respectively. In mandarin juice, sucrose accounts for about 70% of the total sugars while in orange juice it is 50 % (Nagy and Shaw 1990). Organic acids constitute the second most abundant class of soluble solids in the juice, and the total acid value is generally expressed as % citric acid. The quantity of citric acid and malic acid reported in mandarin are 0.86-1.22% and 0.08-0.21%, respectively; and the ascorbic content is 0.03 %. The typical pH of orange juice is 3.3-3.7 and, 3-3.5 for tangerine juice. Sugar/acid ratio is an important quality indicator for flavor and is used to grade the legal maturity of citrus fruits in Florida (Nagy and Shaw 1990). For mandarins, the minimum Brix is 9 and the minimum Brix/acid ratio is 9:1.

Other minor (in concentration) non-volatiles are flavonoids, carotenoids and limonoids. There are three main groups of flavonoids in citrus: flavanone, flavones and anthocyanins (usually restricted to blood citrus) (Ranganna et al. 1983a). Flavanones are the most abundant and generally found as glycosides whose sugar moiety (usually attached at the position 7) will determine its taste properties: the β-rutinose glycoside is tasteless, whereas the β-neohesperidose glycoside is bitter. Mandarins have been reported to contain, among others, flavanone glycosides (isosakuratenin 7- β-rutinoside, hesperidin), flavanone aglycone (citromitin) and flavones aglycones (xanthomicrol, tangeritin, nobiletin).

46

Table 2.1: Global composition of fresh orange juice (calculated and compiled from Ahmed et al. 1978b; Park et al. 1983; Pupin et al. 1999; Gil-Izquierdo et al. 2001; Brat et al. 2003).

Non-volatiles (total ~ 12.5 g/100 g fresh weight)	g/100 g	
Carbohydrates	9.8	
Sucrose		50 %
Glucose		24.5 %
Fructose		26.5 %
Pectin	0.04	
Acids	0.8	
Citric		90.5 %
Malic		8.5 %
Isocitric		< 1 %
Lipids	0.18	
Proteins	0.1	
Vitamin C [a]	0.05	
Phenolics [a]	~ 0.08	
Carotenoids [a]	< 0.01	
Ash [a]	0.03	
Volatiles [b] (total ~ 0.012 g/100 g fresh weight)	mg/ 100 g	
Monoterpene hydrocarbons	10	
Limonene		90-94 %
Sesquiterpene hydrocarbons	0.4	
α-selinene		60-70 %
Valencene		8-10 %
Esters	0.25	
ethyl hexanoate		24%
ethyl octanoate		20%
Aliphatic alcohols	0.22	
Octanol		32%
Butanol		27%
Monoterpene alcohols	0.19	
terpinen-4-ol		37%
Linalool		32%
Aldehydes	~ 0.07	
Hexanal		29%
Ketones	~ 0.07	

Nootkatone	50%
Carboxylic acids	0.1
Water (total qsp. 100 g/100 g fresh weight)	

[a] Calculated with a juice density d=1.023 g/cm^3, measured at room temperature. [b] Only volatiles contributing the most to their chemical group are reported here as a %.

Carotenoids epoxides and ketocarotenoids have been reviewed as the principal compounds responsible for the color of orange and tangerine, respectively. Carotenoids are more abundant in mandarins than in oranges (total carotenoids: 15-23 mg/L orange juice vs. 27 mg/L mandarin juice), and carotenoid composition is strongly influenced by the genetic background of the fruits (Fanciullino et al. 2006). The higher concentrations of β-cryptoxanthin are responsible for the deep color of the peel and pulp in mandarins (Ranganna et al. 1983a). While carotenoids in orange juice were sensitive to pasteurization, no losses were reported in mandarin juice but a clear explanation of the reasons was lacking (Ranganna et al. 1983b). In addition, carotenoids act as flavor precursors for many aroma compounds.

Finally, limonoids are also important (to flavor) non-volatiles in both mandarin and orange juices. Among them, limonin has received particular attention since it is responsible for delayed bitterness in citrus juices (Hasegawa et al. 1996; Dea et al. 2010). Limonin develops after extracting the juice by acid-catalyzed and enzyme-catalyzed lactonization of the ring of the tasteless precursor limonoic acid A-ring lactone (Nagy and Shaw 1990).

2.4.3. Chemistry of orange and mandarin juice flavor

While the previous paragraph provided a general overview of juice composition, the following is a description of the chemical constituents known to participate in citrus (orange and mandarin) flavor.

When referring to citrus juice flavor it is important to distinguish between flavor from thermally processed and fresh-squeezed juices. Typically, a "fresh" juice flavor is the most desired. It is commonly accepted that fresh juice flavor refers to the flavor obtained from

fresh, hand-squeezed fruits immediately after squeezing. However, the flavor of fresh-squeezed juice is highly unstable due to chemical, enzymatic (enzyme release upon cell breakage), and possible microbial reactions (Rouseff et al. 2009). Therefore, the main challenge for juice manufacturers is to preserve this "fresh" flavor. At the industrial level, the best way to stabilize citrus (orange) juice flavor for longer storage periods is through thermal treatment just after juice extraction (i.e., to prevent further enzymatic deterioration and to reduce microbial activity), immediately followed by packaging. These two manufacturing steps are known to strongly affect juice flavor (Perez-Cacho and Rouseff 2008b) resulting in a processed juice flavor very distinct from fresh-squeezed flavor. Thermal processing tends to decrease the low molecular weight "top note" compounds (desirable), and increase compounds resulting from breakdown of carotenoids or increase oxidation of some terpenes (off-flavor).

Differences in raw materials (cultivars, maturity and blend proportions), processing conditions (thermal treatment parameters, extraction method) and storage conditions (packaging material, environmental parameters) have been principally identified as responsible for variations in flavor between juices (Perez-Cacho and Rouseff 2008b). These factors influence the amount and balance between compounds or generate new compounds, which ultimately affect the final flavor profile.

Extensive research has been done to understand how external (processing, storage, and packaging) and internal factors (chemical composition) influence the flavor of orange juice. However, most research has focused on the volatile odor active compounds and their changes in concentration (Rouseff et al. 2009). Little is known about the contribution of non-volatiles to the final flavor perceived. Techniques commonly used for studying orange juice flavor include headspace analysis (direct headspace and SPME), gas-chromatography, gas-olfactometry, and sensory evaluation. Some studies have also examined the physicochemical properties of juices (° Brix, color, pH) but rarely interpreted them directly in terms of sensory meaning (flavor).

2.4.3.1. Volatiles

This summary does not address flavor changes due to heat processing, storage or packaging conditions but rather focuses on presenting volatiles found in fresh unpasteurized orange juice. A recently published review reported that over 300 volatile compounds have been identified in orange juice (Perez-Cacho and Rouseff 2008a). In terms of quantities, the volatile groups present in the greatest quantity are, in decreasing order: monoterpenes (limonene, p-cymene, β-myrcene), sesquiterpenes (α-selinene, valencene) and esters (Table 2.1). However, it is recognized that high concentrations do not reflect proportionally the contribution to the sensory character. In fact, less than 36 compounds appear to be significant contributors to orange juice aroma. The main aroma active compounds reported include 14 aldehydes, 7 esters, 5 terpenes, 6 alcohols and 4 ketones. Some threshold studies have used deodorized orange juice as a matrix to evaluate the contribution of these compounds to the odor (Ahmed et al. 1978c; Plotto et al. 2004; Plotto et al. 2008) while other studies have used model systems or water thereby neglecting potential interactions of aroma compounds with other components (non-volatiles) of the juice matrix (Ahmed et al. 1978b; Buettner and Schieberle 2001).

Volatile compounds are unevenly distributed between the serum (21 % of total volatiles), the pulp (71 %) and cloud (8 %) (Brat et al. 2003). The distribution of volatiles between the different phases depends on the chemical class of the compounds. For example, monoterpene and sesquiterpene hydrocarbons (~ 80 % of the total juice volatiles) are primarily found with the pulp and cloud. In contrast, a large portion of the esters, monoterpene alcohols and aldehydes (number of carbons < 8) are associated with the serum (Radford et al. 1974). The reasons of the association between volatiles and pulp/cloud are still unclear. As summarized by Brat et al. (2003), it is proposed that aroma compounds are associated with solid particles either due to absorption onto oil droplets, to physical entrapment within the cell wall, or to interactions with polysaccharides or glycopeptides.

Table 2.2: Principal odor-active volatiles reported in unpasteurized orange juice and their respective concentrations and thresholds (compiled from Ahmed et al. 1978b; Park et al. 1983; Moshonas and Shaw 1994; Shaw et al. 1999; Buettner and Schieberle 2001; Brat et al. 2003; Plotto et al. 2004; Plotto et al. 2008).

Compound	Conc. (ppm) [a]	Threshold (ppb) [b]	Compound	Conc. (ppm) [a]	Threshold (ppb) [b]
Aldehydes			**Terpene hydrocarbons**		
acetaldehyde	11[c]	187	α-pinene	0.6	1650
hexanal	0.4	151	β-pinene	0.7	37200
octanal	0.1	233	β-myrcene	2.2	773
nonanal	0.1	312	limonene	91.7	13700
decanal	0.1	204	α-terpinolene	0.1	
dodecanal		0.56[d]	**Ketones**		
(Z)-hex-3-enal	0.399[c]	26.9	1-octen-3-one	0.0057	1 [d]
(Z)-non-2-enal			(Z)-octa-1,5-dien-3-one		
(E)- non-2-enal	0.0015	1.52	β-ionone		16.5
(E,E)-2,4-nonadienal			nootkatone	0.7	2240
(E,Z)- 2,6-nonadienal			**Esters**		
(E,E)-2,4-decadienal	0.0012[c]	1.71	methyl butanoate	0.01[f]	116
neral	0.0005[e]	1230	ethyl acetate	0.2	6038
geranial	0.0005[e]		ethyl butanoate	0.3	1.71
Alcohols			ethyl-2-methylpropanoate	0.0027[c]	0.35
1-hexanol	0.2[e]		ethyl-2-methylbutanoate	0.2	0.08
(Z)-3-hexen-1-ol	0.31[e]	348	ethyl hexanoate	0.6	3.3
1-octanol	0.7	190[d]	ethyl octanoate	0.5	
linalool	0.6	113			
geraniol	0.1				
terpinen-4-ol	0.7				

[a] Concentration in unpasteurized mechanical-squeezed orange juice from *Citrus sinensis (L.) Osb, cv. Naveline*. Reported values vary between studies and cultivars, these values are given here for

illustrative purposes and should not be considered as is. [b] Orthonasal threshold in deodorized orange juice matrix. [c] Determined in hand-squeezed juice. [d] Orthonasal threshold in water. [e] In juice from Valencia oranges. [f] In pasteurized juice.

The main aroma compounds contributing to fresh orange juice flavor are summarized in Table 2.2. Even though they are present in very low concentrations (ppm to ppb), their low sensory thresholds make them important to flavor (Ahmed et al. 1978b; Plotto et al. 2004; Plotto et al. 2008). It is important to note that the type of cultivar and fruit maturity significantly impact the levels of these compounds (Buettner and Schieberle 2001; Arena et al. 2006). Also, the juice extraction process plays a significant role in differences in concentrations. For instance, terpene amounts are higher (up to 10-fold for limonene) in mechanically squeezed juices than in manually squeezed juices because of the peel oil extracted along with the juice (Moshonas and Shaw 1994). An additional remark is that there are large discrepancies between the sensory threshold values reported in the literature; differences may be due to any of several factors including whether the compound was evaluated singly or in combination with other compounds, or in water or in a juice matrix. In fact, thresholds of most aroma compounds are higher in juice than in water due to a matrix (pulp) effect (Ahmed et al. 1978a; Roberts et al. 1996).

Aldehydes were reported as the most important compounds (desirable) to the typical aroma of orange juice but they are easily degraded by thermal processing. They impart fruity, green and fatty/metallic odors (Perez-Cacho and Rouseff 2008a). Neral and geranial (isomers of citral coming from the peel oil) are two important contributors to orange flavor and impart a lemon-like, citrus-like aroma. Ethyl butanoate has a fruity aroma and is believed to be one of the most important odorants in orange juice. Terpenes, although being the most abundant group of volatiles in orange juice, contribute little to orange juice flavor because they have high odor thresholds. Limonene is the most abundant volatile in orange juice but its contribution to flavor is controversial; more than an odor-active compound, it has been proposed that limonene acts as a promoter of other aroma active compounds similarly to ethanol in wines (Perez-Cacho and Rouseff 2008a). Among alcohols, linalool is considered a major odorant with sweet and floral odor. β-ionone is described as floral and raspberry-like and is the most potent of the four norisoprenoids identified in orange juice

(Mahattanatawee et al. 2005), no quantitative data have been reported for this compound as far as we are aware.

Although mandarin peel oil and essential oil have been extensively studied for their use in perfumery (e.g., Moshonas and Shaw 1974; Shaw 1979), there is less work available on volatiles in fresh mandarin. More than 200 volatiles belonging to 11 chemical classes have been reported in hand-squeezed juices from various tangerine hybrids (Kerbiriou et al. 2007a; Miyazaki 2009; Miyazaki et al. 2011). Most of the volatiles reported in mandarin fruits are terpene hydrocarbons, aldehydes, esters, alcohols and ketones; they are mainly derived from terpenoids (e.g., carotenoids), fatty acids and amino acids (Yajima et al. 1979; Miyazaki 2009).

As one would expect, the genetic background of mandarin hybrids affects volatile composition and content of fruits since it directly influences flavor biosynthesis pathways and consequently, final volatile profile. For example, Murcott samples were characterized by higher levels of water-soluble compounds, especially C1-C5 alcohols (Moshonas and Shaw 1997) and hybrids with sweet orange (*Citrus sinensis* L. Osb.) in their lineage were higher in esters and sesquiterpenes, and described with more fruity notes (Kerbiriou et al. 2007a; Miyazaki et al. 2011).

Recently, the findings of eight different studies regarding the volatile composition of mandarins were summarized (Tietel et al. 2011). The researchers identified 37 volatiles that were reported in at least four experiments and among those, nine of them were present in seven out of eight experiments. These nine compounds were proposed to be core aroma volatiles of mandarin flavor and included: linalool (floral, green, citrus), α-terpineol (floral, lilac-like), terpinen-4-ol (woody, earthy), carvone (spearmint or caraway), nonanal (piney, floral, citrusy), decanal (beefy, musty, cucumber), limonene (citrus-like, fresh), α-pinene (pine-like, resinous) and myrcene (musty, wet soil).

Mandarin juice (single-strength) is very unstable to storage and thermal treatment, and rapidly develops off-flavors. Off-flavor development is due to precursors located in the peel, flesh and juice sacs that are extracted along with the juice: for instance, due to the oxidation of limonene into carvone and carveol (Ranganna et al. 1983b). However, the complete

53

removal of peel oil from the juice is not a complete solution to decrease the off-flavor problem. In fact, one study reported that excessive removal of peel oil in the juice by vacuum evaporation (remaining average of 0.0075 % (v/v) of juice) was associated with a carrot-like aftertaste (Fishman and Bednyak 1974). This is not totally surprising since the compounds associated with the typical "mandarin flavor" are not present (or are in trace quantities) in the juice itself but are located in the peel oil (Moshonas and Shaw 1997). Thymol and N-methylanthranilate have been identified as key odorants of mandarin peel oil flavor, and the presence of β-pinene and γ-terpinene was also reported necessary to impart the characteristic mandarin peel oil flavor (Yajima et al. 1979; Shaw and Wilson 1980).

Mandarin juice flavor quality is greatly affected by the thermal treatment as previously reported (Pérez-López and Carbonell-Barrachina 2006). Limonene, linalool, α-terpineol, and terpinen-4-ol have been proposed as quality control markers of mandarin juices: the concentration of α-terpineol, and terpinen-4-ol were positively correlated with a decrease in sensory quality of the mandarin juices (Pérez-López et al. 2006). Those two compounds are also markers of oxidation and excessive heating.

2.4.3.2. Non-volatiles

Even though most of the non-volatiles identified in citrus fruits are known taste contributors (Table 2.1), there is little literature available on their contribution to flavor or aroma-taste interactions. The major non-volatiles are sugars (primarily sucrose, glucose and fructose) and acids (citric and malic acids) that represent about 78 % and 6.5 % of the dry matter of orange juice, respectively.

Glucose, fructose and sucrose are responsible for the sweetness of the juices, whereas citric and malic acids, mostly, impart sourness (Ranganna et al. 1983a). As reviewed by Rouseff et al. (2009), the relation between sugar-acid ratio and orange juice flavor has been reported by early studies. It was observed that the sugar-acid ratio correlated to fruit maturity, and indirectly to flavor development since volatiles (secondary plant metabolites) are generated during fruit ripening. Ahmed et al. (1978c) also reported that the sugar-acid ratio affected the sensory ratings of orange juice but a clear explanation of the reasons (e.g., interactions with the matrix or aroma compounds) was lacking. Acids have been reported to

significantly increase the sensory threshold of limonene in water and sugars tended to increase it as well, while opposite trends were observed for pectin (Ahmed et al. 1978a). Unfortunately, these findings were limited to one volatile compound (limonene) and were not evaluated with other volatile compounds. Another study reported that the headspace concentration of α-pinene in model solutions of equal viscosity was further decreased when sucrose was used compared to carboxymethylcellulose, suggesting that sugars have an effect on volatile headspace concentrations (Roberts et al. 1996). The contribution of sugars and acids to flavor have been examined in other food systems also. For instance, Baldwin et al. (2008) found that the intensity and profile of perceived flavor in tomato puree depended on sugar and acid levels.

The contribution of other non-volatiles in the juice matrix has been overlooked as witnessed by the small amount of literature available. Lipids have been shown to influence the perceived flavor in juice: orange juice flavor was described as being more pleasant, with respect to the typical orange-like aroma, when a model solution (water, sugars, acids) contained 0.1 % sunflower oil (Buettner and Schieberle 2001). Other non-volatile compounds, such as limonoids (limonin, nomilin), limonoid β-D-glucopyranosides, and polymethoxylated flavones have received some attention but mostly because of their negative impact on consumer liking of orange juice (Attaway and Carter 1971; Hasegawa et al. 1996; Havekotte et al. 2009) and for their potential health benefits (Huang et al. 2007). These compounds have been associated with bitterness and astringency in orange juice. They can be perceived even when at levels below their individual thresholds because of a synergistic effect between them (Havekotte et al. 2009).

Finally, components such as flavonoids (e.g., flavanones and their derivatives such as naringin and hesperidin) and carotenoids (e.g., α-, β-carotene, lutein) are expected to influence flavor perception of citrus juice. Their effect on flavor has been reported elsewhere (Jung et al. 2000) and found to be compound-dependent. For example, naringin, the bitter compound in grapefruit, decreased the volatility of ethyl benzoate but did not affect the volatility of limonene (King and Solms 1982). Also, carotenoids are known precursors of aroma volatiles (Mahattanatawee et al. 2005).

2.4.3.3. Influence of juice matrix on flavor

As presented earlier, orange juice is a heterogeneous system formed by the combination of serum, pulp and cloud. The suspended solid materials (cloud and pulp) have been reported to affect the perception of orange juice flavor either by causing a change in aroma release (viscosity, volatility), by imparting a mouthfeel sensation or by providing a protective effect to some specific compounds.

The influence of the orange juice matrix on sensory characteristics was reported in an early study by Ahmed et al. (1978c). Researchers noticed that sensory ratings were always higher when aroma compounds were tested in deodorized orange juice rather than in water suggesting aroma-matrix interactions. Similar observations have been made for tomato flavor for which the flavor of model systems was judged less alike to real samples if matrix compounds were absent (Tandon et al. 2003). The presence of pulp in the juice has been shown to modify the solubility of limonene and subsequently its headspace concentration which affects its orthonasal perception (Ahmed et al. 1978a; Braddock et al. 2004). Even though limonene is not a direct aroma contributor, its presence is necessary to the success of flavor recombination studies (Rouseff et al. 2009).

The impact of suspended solids on sensory perception in orange juice and the consequences of their removal were further highlighted in a recent study (Rega et al. 2004b). Besides providing large quantities of aroma compounds and modifying the texture (mouthfeel), the addition of pulp (and cloud) to low-pulp juice increased the "freshly-squeezed" attribute ratings. This is not very surprising given the fact that about 79 % of the volatiles in the juices are associated with the pulp and cloud (Brat et al. 2003).

Pulp was also reported to provide some protection to some important aroma compounds (acetaldehyde and hexanal) against deteriorative losses due to thermal treatment; as well, headspace concentrations of carveol and α-terpineol (two off-flavors) were strongly reduced when pulp was present in the juice prior to pasteurization (Rega et al. 2004a).

2.4.4. Sensory analysis of orange and mandarin juice flavor

Sensory analysis, and more particularly descriptive sensory analysis, has been long used by scientists and industry professionals to characterize the flavor of citrus juices. This section is mostly intended to familiarize the reader with past sensory studies on orange and mandarin juice flavor and to introduce some of the common sensory attributes used. Once again, more literature is available on orange juice.

Even though descriptive sensory analysis is done routinely by juice manufacturers, there is little data available in the literature probably because most of the work is proprietary. Some descriptive studies have focused on characterizing the sensory differences between juices produced using different process (Perez-Cacho et al. 2007). Fresh hand-squeezed juices were described with strong citrus, floral, grassy, and sweet attributes but those attributes were significantly lower or absent for juices that have been extensively thermally processed (i.e., canned juices). New attributes (unpleasant) such as cooked, bitter, grapefruit/tropical fruit, among others, were used instead. Other descriptive sensory studies have been conducted to monitor the flavor profile changes due to storage (Petersen et al. 1998) and also been used to categorize juices based on sensory characteristics instead of chemical measurements (Lotong et al. 2003). Descriptive sensory analyses have been used as well to determine the role of juice components on flavor perception. For example, in reconstitution experiments to verify the contribution of key aroma compounds to orange juice flavor by comparing sensory profiles between reconstituted juice (aroma compounds added to a model solution containing 8.5 % sugars, 1.6 % acids and 0.1% lipids) and hand-squeezed juice (Buettner and Schieberle 2001). This study also demonstrated the influence of lipids on orange-like flavor perception.

During training of a descriptive sensory panel, a list of terms is developed to describe the sensory properties of the food product. Such list is called a lexicon and most often includes taste, aroma and texture (mouthfeel: astringency) descriptors (for examples, see Lotong et al. 2003; Perez-Cacho et al. 2008). Other descriptors such as color or sound may be used but they are not commonly employed for beverage studies. One lexicon was recently published by Perez-Cacho et al. (2008) to describe the flavor of fresh-squeezed and processed orange

juice. The lexicon was developed by two independent trained panels in Spain and the United States. It included over 34 descriptors that can be divided into four categories: odor, aroma, basic tastes and trigeminal/tactile sensations. The odor and aroma descriptors were used to develop an aroma/odor wheel for orange juice and it is presented in Figure 2.6. This aroma/odor wheel was also based on sensory work done by Elston (2005). Overall, aroma descriptors were classified in seven families: fruity, green, earthy, floral, heated, chemical, and others. Such an aroma/odor wheel is valuable for future sensory studies, both in academic and industrial settings, as it is a useful tool to help the judges in developing a vocabulary and in getting trained.
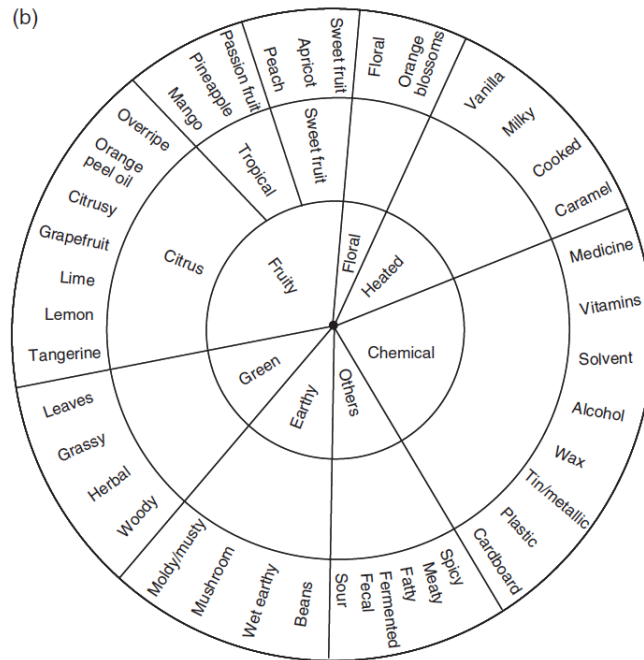
Figure 2.6: Odor/aroma wheel developed for the descriptive sensory analysis of orange juice (Reproduced from Perez-Cacho et al. (2008). Copyright 2008 Sage Publications, with permission).

Few sensory studies have been published on fresh mandarin (fruits and juices). In a recent study, Carbonell et al. (2007) reported on the descriptive sensory analysis of fresh

and processed Spanish mandarin juices and proposed a set of 29 descriptors including appearance, odor, taste and texture descriptors. Another research group described the sensory profile of mandarin hybrids and cultivars over three years using a trained panel (Plotto et al. 2010). This study was done as a part of the University of Florida citrus breeding program to develop new hybrids with high sensory quality. The samples evaluated covered a large range of sensory properties, ranging from pumpkin and fatty flavor profiles to very sweet and floral flavors, as well as sour and grapefruit notes. As one would expect, the sensory quality of the juices was somewhat related to the genetic background of the fruits used in that study, as they tended to cluster together (Plotto et al. 2010).

Sensory analysis was also used to determine the optimum harvesting maturity of the new hybrids by collecting the fruits at different times throughout the season (Kerbiriou et al. 2007b; Miyazaki et al. 2008). Researchers were able to follow the evolution of the sensory profile of the fruits as they were maturing; for instance, decreasing in sourness and increasing sweetness and fruity notes. Finally, very few consumer studies have been published even though it is critical to determine what the expected desirable properties of the juices are, and to assess the acceptance of the fruit juices before releasing them to the market. Carbonell et al. (2008) studied the acceptability of chilled mandarin juices (*Citrus clementina*) and chilled orange juices (included for comparative purposes) by 100 Spanish consumers. Interestingly, most of them preferred mandarin juices even though 10 % of consumers found orange juices more acceptable. The mandarin juices were judged sweeter and fresher than the orange juices. In a follow-up study, the researchers evaluated the sensory profiles and the consumers' acceptability of juices and blends from different mandarin varieties to determine their suitability for commercialization (Carbonell et al. 2009).

2.4.5. Mathematical models of orange juice flavor

The value of relating sensory and analytical data has been discussed earlier in this chapter; a brief overview of published studies that attempted to develop orange juice flavor models follows.

59

There is some information in the literature on studies developing models of orange juice flavor as early as in the 1970s (Attaway and Carter 1971; Attaway et al. 1972; Carter and Buslig 1975). Researchers used simple to multiple linear regressions, as well as multivariate analysis to relate two data sets. One study developed polynomial flavor predictive models from 30 physicochemical quality measurements (e.g., °Brix, acids, oil, viscosity, sucrose, limonin, etc.) measured on a total of 108 processed juices from three cultivars (Carter and Buslig 1975). Multiple regression analysis was applied and resulted in a relatively good data fit (R^2 =0.8 to 0.9) if enough analytical variables were included (at least 4). However, the predictive performance of the model with new samples is lacking in the paper. Also, no data on the volatiles were included in the models probably because of the limited availability of gas chromatographs at that time. It is guessed that the absence of aroma compounds in the models would have caused poor predictions with new samples.

It is only around the 1980s that reports including aroma compounds and sensory analysis started to appear. Pino (1982) was able to establish good predictor models of flavor quality using multiple linear regression analysis between selected GC chromatographic peaks (myrcene, linalool, 2-hexanol, α-terpineol) and sensory data from juices with different thermal processes. Similarly, changes in flavor due to storage conditions have been studied by means of sensory and GC profiling (Velez et al. 1993). In another study, selected sensory attributes were accurately predicted using PLS regression model developed from sensory and aroma compound measurements of stored juices (Petersen et al. 1998). The PLS model was further refined and included only GC data that explained most of the sensory differences between normally stored and accelerated stored juices; however, no external validation was done. Analytical measurements (GC, °Brix, acid values) and sensory characteristics combined with multivariate analysis have been also used to categorize orange juices as to flavor quality, cultivar, fruit maturity and/or processing technology (Shaw et al. 1993; Burgard 1995; Shaw et al. 1999; Lotong et al. 2003) suggesting that the models can be used as a classification tool for quality control purposes.

More recently, Elston (2005) used a sensory-directed multivariate approach to evaluate the sensory quality of orange juices. The approach focused on primarily using only instrumental variables with known aroma or taste activity. The calculated equation was

tested on a set of six external juices and the model was able to predict relatively well the overall sensory quality score (even though the predicted scores were off by up to 4.5 points on a scale 0-15). Of importance, the percentage of misclassification observed in the discriminant analysis was higher for juices of fair and poor sensory quality and the predictions more accurate for good quality juices. Elston proposed that other differentiating compounds were not being measured which accounted for the miscalculations made.

In summary, none of these works combined both volatile and non-volatile measurements with sensory characteristics in the predictive models. Elston's work (2005) did include measurements related to non-volatiles but they were limited to gross measurements of traditional taste stimuli (°Brix, % oil, % acid, sugar-acid ratio). Also, the analytical measurements were confined to known individual flavor contributors which limited the opportunities to include in the model other relevant chemical stimuli and/or interactions that contribute as well to the <u>total</u> flavor. The inclusion of volatile and non-volatile compounds (not restricted to known flavor contributors) in the flavor models is expected to improve the predictive performance and this was tested in the present thesis. This approach acknowledges that flavor results from the combination of all chemical stimuli as a mixture and not only from individual compounds having sensory properties on their own. Therefore, a model that integrates all chemical stimuli should depict more accurately the total flavor perceived as both individual contributions and hidden interactions between compounds are accounted for. The disadvantage of this approach is that the correlations observed between compounds are not guaranteed to be causative and they will need a mandatory sensory validation.

Chapter 3 : Development of instrumental methods for the untargeted analysis of orange juice flavor

Note 1: Sections of this chapter have been published in Flavour and Fragrance Journal (Charve et al. 2011) and are being reproduced here with permission from the editor (Copyright 2011 Wiley).

Note 2: The methods presented below were developed for the analysis of orange juice as the research was initially focused on orange juice flavor before being changed to the study of mandarin juice flavor. We decided that the methods developed were valid for the analysis mandarin juice since both fruits belong to the genus Citrus, therefore having close chemical composition.

Abstract. The objective was to develop methods for the untargeted analysis of chemical stimuli related to orange juice flavor. The sample preparation and analytical techniques chosen were a compromise between maximizing the number of compounds extracted and detected, method throughput, and reproducibility. A headspace solid-phase micro extraction method coupled with a gas chromatograph time-of-flight mass spectrometer (GC-TOF-MS) was used to determine volatiles. Of the SPME fibers screened, CAR-PDMS fiber was found to be best suited for this purpose. The extraction parameters were optimized using a response surface methodology. Non-volatiles were extracted by solid phase extraction (SPE) before analysis by ultra high performance liquid chromatography coupled with a TOF-MS (UHPLC-TOF-MS). The analysis of all SPE fractions with two columns (reversed-phase and hydrophilic interaction liquid chromatography) and the use of electrospray ionization in negative and positive modes increased the number of compounds analyzed and detected.

3.1. INTRODUCTION

As presented in the literature review, flavor research has traditionally minimized the contribution of non-volatiles to flavor perception while focusing primarily on volatile compounds. The work presented in this thesis, "Flavoromics", is an alternative approach for flavor research that may address some of the issues being neglected in past studies. This research adapts concepts (untargeted analysis) and tools (chemometrics, ultra high performance liquid chromatography UHPLC) from metabolomic investigations. Flavoromics considers for study all (ideally) low molecular weight compounds in food systems (volatiles AND non-volatiles) as candidate chemical stimuli in human flavor perception instead of focusing only on compounds that are already known to influence the flavor quality (biased).

Using an untargeted methodology to study flavor opens the possibility to identify new flavor contributors and also to have a better prediction of flavor as it includes inputs from more chemical compounds. However, this approach is challenging from a sample preparation and analytical point of view because of the expected diversity of compounds studied (physicochemical properties and ranges of concentration). As stated by de Vos et al. (2006) when referring to the use of metabolomics in coffee research, there is « no single analytical method [...] currently capable of extracting and detecting all metabolites ». Consequently, several methods should be used to get the most comprehensive view of the sample being analyzed. Also, one should limit the number of preparation steps as each step is a potential source of changes in the original sample and a limit in sample analysis throughput. Numerous analytical platforms have been used for metabolomics however MS-based methods combined with chromatography were chosen for flavoromics because of their abilities to separate, to identify and to quantify compounds with high selectivity and sensitivity compared to other techniques (e.g., nuclear magnetic resonance (NMR) spectroscopy, Fourier transform-infrared spectroscopy) (Dunn and Ellis 2005; Dettmer et al. 2007). Other reasons were the availability of MS instrumentation in most laboratories, the familiarity with MS spectra interpretation of most scientists and the ability to automate mass spectra data processing for subsequent multivariate analyses.

The objective of the present study was to develop instrumental methods capable of providing the largest unbiased pool of compounds for analysis. Four commercial juices and their blend were analyzed by headspace solid-phase micro extraction gas chromatography (GC) and ultra high performance liquid chromatography (UHPLC) – time of flight mass spectrometry for volatiles and non-volatiles, respectively. Sample preparation and analytical techniques chosen were a compromise between the number of compounds extracted and detected, throughput, and repeatability. Next chapter (Chapter 4) examines the ability of the methods to distinguish between orange juices based on mass spectral information using chemometrics.

3.2. MATERIALS AND METHODS

Orange juice samples. Four commercial premium orange juices (referred to as brand A, B, C and D), all pulp-free, were purchased at a local grocery store. The juices were chosen as a sample of commercial premium orange juices available and only for method development purposes. All juices were transferred from their original packaging (either gable-top paperboard carton or PET bottle) to glass Erlenmeyer flasks and stored at 4 °C until sample preparation. A reference sample was prepared by blending equal proportions of all four juices (v/v) and was used to develop the instrumental methods ensuring the broad coverage of the compounds analyzed. Five replicates of the reference (blended juices) were prepared for each analysis.

Chemicals. Sodium chloride (>99.5%), acetic acid anhydride (99.7%) and acetonitrile (99.9 % Optima LC-MS grade) were purchased from Fisher Scientific (Pittsburg, PA). Ammonium acetate (97 %), formic acid (98-100 %) and leucine enkephalin acetate salt hydrate (≥ 95 %) were purchased from Sigma-Aldrich (St Louis, MO). Nanopure water (18.2 MΩ.cm) was obtained using a NANOpureDiamond system (ThermoScientific Barnstead, Waltham, MA).

Sample preparation for the analysis of orange juice volatiles. Samples used to develop and optimize the headspace solid phase micro-extraction (SPME) method were

prepared by adding 10 mL of juice blend into 20 mL headspace vials sealed with a butyl/PTFE septum and a magnetic crimp cap, and adding salt (NaCl) if directed by the experimental design (method optimization). All samples were prepared in five replicates the day prior to analysis, stored overnight at 4°C and then placed at room temperature at least 2h prior analysis. Blank vials (empty) were prepared to allow checking for trace residual volatiles on the SPME fiber or artifact that may confound the data and were analyzed randomly throughout the experiment. Prior to analyses, the SPME fibers were conditioned according to manufacturer's recommendations. Samples were loaded onto a Gerstel MPS2 autosampler equipped with a SPME option. The Maestro software (Gerstel, Baltimore, MD) was used as cycle composer with the co-current sample preparation feature enabled, which allowed higher throughput.

   Sample preparation for the analysis of orange juice non-volatiles. Compounds not analyzed by headspace SPME (i.e., non-volatiles) were extracted from the juice matrix either directly or by solid phase extraction (SPE) before analysis by ultra high performance liquid chromatography (UHPLC). For the direct analysis of the juice (no SPE), the blended juice was diluted (1:4) in water before centrifugation. The clarified juice obtained after centrifugation (6800 g/ 5 min) was used immediately for analysis by UHPLC or for the SPE extraction. The C18-SPE cartridges (Discovery DSC18, 500 mg sorbent, 3 mL tubes, Supelco, Bellefonte, PA) were conditioned with 3 mL acetonitrile and equilibrated with 3 mL water/acetonitrile (95/5) before loading 2 mL of clarified juice onto the cartridge. After passing the sample through the cartridge, the SPE cartridge was washed with 3 mL water/acetonitrile (95/5) and the fraction collected (SPE-wash fraction). Compounds retained on the cartridge were eluted using 3 mL acetonitrile and the fraction collected (SPE-eluate fraction). SPE blanks (cartridge loaded with nanopure water) were prepared similarly to check for potential artifacts due to sample preparation. A SPE Visiprep 12-port vacuum manifold (Sigma-Aldrich) fitted with disposable liners was used to increase sample preparation throughput and repeatability. Before analysis by UHPLC, samples were diluted (1/4) using nanopure water and acetonitrile as diluents for reversed-phase (RP) and hydrophilic interaction liquid chromatography (HILIC) separations, respectively. Samples were then filtered through a 0.2 µm nylon membrane (Millex, Millipore Corp., Bedford, MA)

before transfer to 2 mL autosampler vials. All samples were prepared in five replicates and analyzed under each tested condition (no SPE vs. SPE, column separation and ionization mode).

   Analysis of orange juice volatiles. The first step of the study was to select a SPME fiber for the analysis of the volatiles from orange juice headspace. Three 1-cm SPME fibers (Supelco, Bellefonte, PA), 100 μm polydimethylsiloxane (PDMS), 75 μm carboxen/ polydimethylsiloxane (CAR-PDMS) and Stableflex 50/30 μm divinylbenzene/carboxen-polydimethylsiloxane (DVB/CAR-PDMS), were compared for their ability to extract volatiles from the headspace of the juice blend. The different characteristics of the SPME fibers are summarized in Table 3.1. The three fibers were exposed to the headspace of the samples for 10 min at room temperature and then were desorbed into the injection port of a gas chromatograph (260 °C) for 10 min. As a criterion for method repeatability, only peaks detected in at least 80 % of the replicate samples were counted. The fiber giving the highest total number of detected peaks was selected for the remainder of the study and used to optimize the headspace SPME method.

Table 3.1: Characteristics of the SPME fibers tested (Shirey and Mindrup 1999).

Fiber coating	Type	Polarity
100 μm PDMS	Absorbent	Non-polar
75 μm CAR-PDMS	Adsorbent	Bi-polar
50/30 μm DVB/CAR- PDMS, Stableflex	Adsorbent	Bi-polar

   Following fiber selection, a response surface methodology (RSM) with a central composite face-centered experimental design (three center points) was used to optimize the headspace SPME extraction conditions. Incubation temperature (40 to 50°C), extraction time (10 to 30 min) and addition of salt (NaCl) (0 to 0.4 g/mL) to samples were optimized since they are known to generally affect the recovery of compounds by SPME (Nongonierma et al. 2006; Risticevic et al. 2010). The experimental runs are summarized in Table 3.2 and

were conducted in random order using the Maestro software. Prior to the SPME extraction, samples were incubated at 50 °C for 5 min under constant stirring (500 rpm).

Analyses were performed on an Agilent GC 7890A equipped with a HP-5MS capillary column (30 m length, 0.25 mm i.d., 0.25 µm film thickness) (J&W Scientific, Folsom, CA) and coupled with an electron impact time-of-flight mass-spectrometer (EI-TOF-MS) (GCT, Waters, Milford, MA). The operating conditions were as follows: GC: split 5:1; initial oven temperature: 40 °C increased at 5 °C/min to 150 °C, then 40 °C/min to 250 °C; TOF-MS: 40-280 amu, centroid mode, scan time: 0.1 sec, inter-scan time: 0.02 sec, 2750 volts, 70 eV. An inlet liner (78.5 mm x O.D. 6.5 mm x I.D. 0.75 mm) designed to achieve sharper peaks with SPME/GC analyses was used.

Table 3.2: Experimental runs for the optimization of the SPME method by response surface methodology using a central composite face-centered experimental design.

	Experimental runs																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Incubation temp. (°C)	40	50	40	50	40	50	40	50	40	50	45	45	45	45	45	45	45
Extraction time (min)	10	10	30	30	10	10	30	30	20	20	10	30	20	20	20	20	20
Salt (g/mL)	0	0	0	0	0.4	0.4	0.4	0.4	0.2	0.2	0.2	0.2	0	0.4	0.2	0.2	0.2

The GC chromatograms were processed using the Peak Detect option from Masslynx software (Masslynx v.4.1., Waters, Milford, MA) with the following parameters: automatic noise measurement, smoothing by moving mean (3 scans, 2 smooths), peak detect: default values, absolute peak height threshold: 100. The number of peaks detected for each run (no efforts were made to detect co-eluting compounds) was used as response variable in the RSM experimental design and a quadratic model was built to describe the observed data. An analysis of variance (ANOVA) was conducted to identify first- and second-order terms that should be included in the model and a lack-of-fit test was run to verify the adequacy of the model to describe the response in the region of experimentation. The fitted model was then

used to draw a response surface plot from which optimum extraction parameters were determined. The statistical analysis was done with Statgraphics statistical software (StatPoint Inc, 2006, v.15.1.02.) on a personal computer and the level of significance was set at α=0.05.

Analysis of orange juice non-volatiles. Extracts prepared either by simple dilution with water (no SPE) or by SPE were analyzed by UHPLC (Acquity UPLC, Waters, Milford, MA). To increase the separation and the detection of compounds, complementary stationary phases and ionization modes were used. The washed fraction (highly hydrophilic) was injected onto a HILIC column (BEH amide column, 2.1 mm x 100 mm, 1.7 μm, Waters, Milford, MA), the eluted fraction (relative hydrophobic) onto a RP column (BEH C18 column, 2.1 mm x 50 mm, 1.7 μm, Waters, Milford, MA) and the juice diluted with water was analyzed by both HILIC and RP.

The gradients used for UHPLC analyses are summarized in Table 3.3. The mobile phases were: A: ammonium acetate (10 mM, adjusted to pH 5 with acetic acid), B: 95/5 acetonitrile/ ammonium acetate (10 mM, adjusted to pH 5 with acetic acid) for HILIC separation; and A: formic acid (0.1 %); B: acetonitrile + 0.1 % formic acid for RP separation. The injection volume was 5 μL (partial loop overfill mode), the columns were maintained at 40 °C and the autosampler chamber at 12 °C. The UHPLC column effluent was directed into a quadrupole-TOF-MS (Synapt MS, Waters, Milford, MA) equipped with an electrospray ionization (ESI) source operating first in positive and then, in a separate analysis, in negative mode. The MS operating conditions were as follows: source temperature: 120 °C; desolvation temperature: 350 °C; desolvation gas: 700 L/h; cone gas: 50 L/h; capillary voltage: 3 kV (ESI-), 3.2 kV (ESI+); cone voltage: 35 V (ESI-), 30 V (ESI+); extraction cone: 4 V (ESI-), 4 V (ESI+); TOF-MS: 50-1500 amu, centroid mode, scan time: 0.3 sec. Leucine enkephalin (m/z 556.2771 and 554.2615 in positive and negative mode, respectively) at a concentration of 0.5 ng/ μL was used as a lockmass for mass accuracy and infused at a flow of 40 μL/min during the experiment.

Table 3.3: Gradients used for UHPLC-TOF-MS analyses of the SPE-eluate and SPE-wash fractions by reversed-phase (RP) and hydrophilic interaction liquid chromatography (HILIC), respectively.

	Time (min)	Flow (mL/min)	% B	curve
RP	initial	0.5	0.5	/
	4	0.5	20	6
	7.5	0.5	100	8
	8.5	0.5	100	6
	8.6	0.5	0.5	1
	10	0.5	0.5	6
HILIC	initial	0.3	95	/
	2	0.3	95	6
	6	0.3	65	6
	6.5	0.3	50	6
	6.6	0.3	95	1
	10	0.5	95	6

The base peak ion (BPI) chromatograms were processed using ApexTrack peak integration to detect chromatographic peaks (Masslynx software). The parameters were set as follows: smoothing disabled, automatic peak-to-peak baseline noise, automatic peak width at 5% height, baseline start threshold 0%, baseline end threshold 5%. Also, chromatograms were inspected visually for peak width and signal intensity. The repeatability of the method was assessed by the coefficient of variation calculated from the variations between replicates in the number of chromatographic peaks detected and variations in the signal intensity of randomly selected peaks. As well, chromatograms were evaluated for potential retention time shifts. The experimental conditions (sample preparation, column and ionization mode) chosen for the remainder of the study were a compromise between the number of chromatographic peaks detected and chromatographic quality.

3.3. RESULTS AND DISCUSSION

The objective of this study was to develop instrumental methods able to gather information on as many compounds as possible that are likely to contribute to flavor perception in orange juices. Due to the nature of untargeted analyses, we needed to maximize the number of compounds extracted and analyzed from orange juice with minimum bias, with an overriding need to be rapid: certainly conflicting or mutually exclusive objectives requiring compromise. To overcome potential changes in sample composition due to sample preparation steps and meet the need for a short analysis time, techniques involving little sample preparation were chosen. A headspace SPME-GC-TOF-MS method was developed for the analysis of orange juices volatiles, and a SPE-UHPLC-TOF-MS method was used for compounds that were not extracted from the headspace (i.e., non-volatiles). The overall workflow of the study is outlined in Figure 3.1.
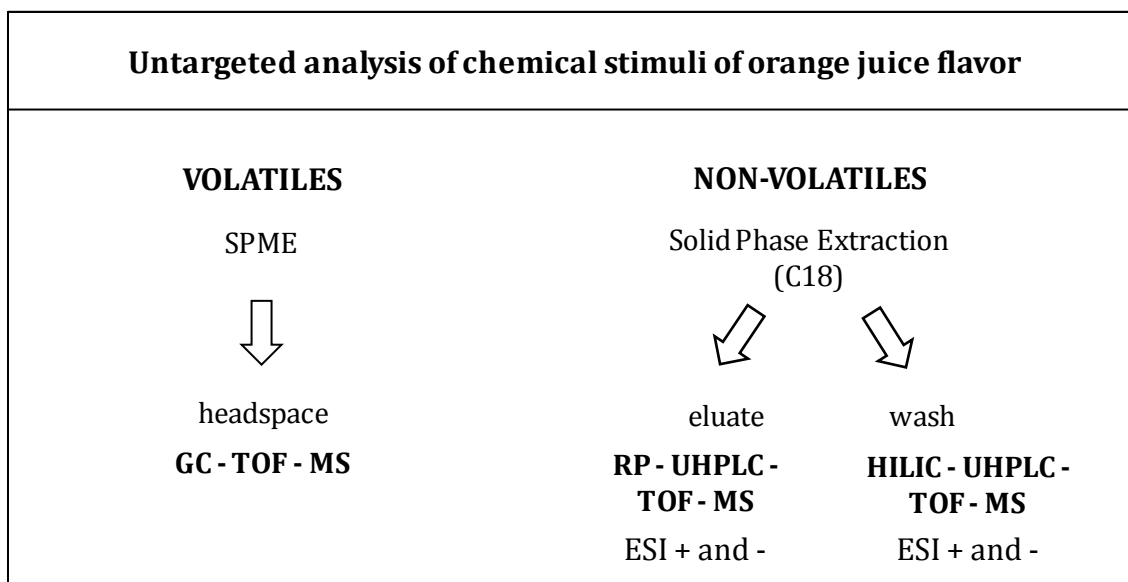
Figure 3.1: Summary of the analytical methods used for the untargeted analysis of orange juice.

3.3.1. Untargeted analysis of orange juice volatiles

SPME is commonly used in flavor analysis for the analysis of volatiles in the headspace (Yang and Peppard 1994; Steffen and Pawliszyn 1996) and is also gaining popularity in metabolomic investigations for profiling volatiles from plants or fruits (Tikunov et al. 2005) as well as for residues analyses in biological fluids such as blood, saliva and urine (Mills and Walker 2000). It is a quick (little sample preparation), easily automated and solvent-less method that uses a coated fiber onto which volatiles are retained. However, SPME has some weaknesses which must be acknowledged when developing methods. Major disadvantages are related to its limited quantification abilities due to the selectivity of the fiber coating toward some compounds, the low volume of coating material that easily leads to saturation of the fiber (absorbent fibers) or competition between compounds for the binding sites on the fiber (adsorbent fibers), and less than optimum sensitivity (Li and Weber 1999; Nongonierma et al. 2006). Also, SPME fibers have a limited lifetime which becomes problematic in storage studies or research involving large sample sets (Li and Weber 1999; Shirey and Mindrup 1999; Nongonierma et al. 2006). Despite these limitations, we feel that the advantages of no major sample preparation steps to produce artifacts or extended analysis time and the ability to automate (and therefore, analyze large numbers of samples) override the limitations of this method.

3.3.1.1. SPME fiber selection

The first step for the development of the headspace SPME-GC method was the selection of the fiber to use. Three SPME fibers of different binding mechanisms and polarities (Table 3.1) were compared for their ability to extract volatiles from the headspace of orange juice. The goals were to maximize the number of compounds being extracted by the fiber and also to have the broadest coverage in terms of chemical classes. At this point of method development, no attempt was made to identify volatiles or to examine potential co-eluting peaks.

As expected, differences in type and polarity of the fibers led to differences in extraction efficiency (i.e., number of peaks detected in the chromatograms after extraction by each fiber) although the number of peaks detected was within comparable ranges. The most

effective fiber at extracting volatiles from the headspace was the bi-polar CAR/PDMS fiber which gave 59 peaks present in at least 80 % of replicates, then the PDMS fiber (55 peaks) and finally the DVB/CAR/PDMS fiber (51 peaks). We wish to comment here that even though the number of chromatographic peaks detected appears low compared to the number of volatiles reported in orange juices (Perez-Cacho and Rouseff 2008a), we need to recall that for the purposes of comparing fibers and developing the SPME method, only peaks (not individual compounds) automatically detected by Markerlynx in at least 80 % of the replicates were counted while ignoring compounds with a low signal or those that would co-elute. We will provide data later (Chapter 4) on individual "features" (unique RT-m/z) that gives a more accurate view of the compounds being measured.

Esters, aldehydes,
terpene derivatives

CAR-PDMS
8

2   4

45

PDMS   DVB-
6   2   CAR/PDMS
0

Terpenes and
derivatives
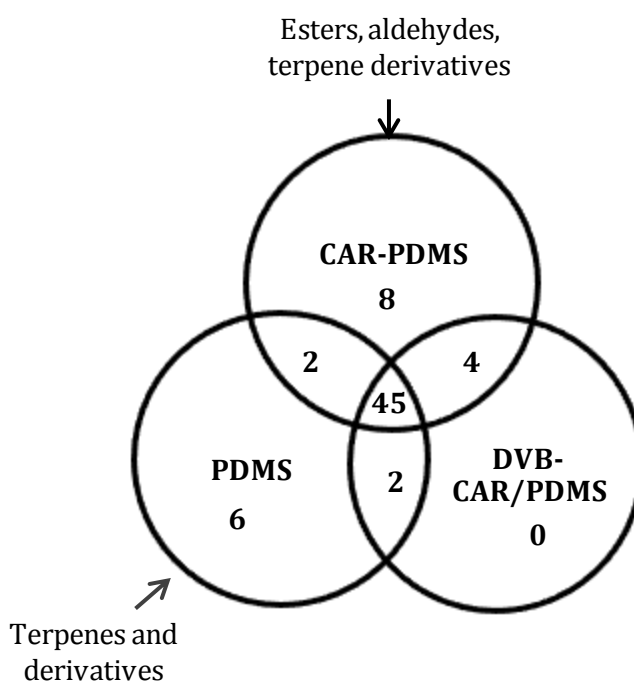
Figure 3.2: Venn's diagram showing relations between information collected by three different SPME fibers. Numbers represent the total of chromatographic peaks detected; peaks that were specific to one type of fiber are tagged (tentative identification).

Before making a final choice on the fiber to use, a Venn diagram was established to examine the similarities in compounds extracted by each fiber (Figure 3.2). Since our goal was to have an untargeted analysis of volatiles, it would have been erroneous to choose a fiber that was able to extract only a particular group of compounds. Compounds that were specific to a fiber were tentatively identified using mass spectral library matches. A first observation is that there were no peaks detected by the DVB/CAR-PDMS fiber only. This is not totally surprising since the dual coated fiber is composed of the same coating materials as the other two fibers, with the addition of porous DVB. Apparently the DVB material, which can adsorb larger hydrophilic analytes because of its larger pore size compared to Carboxen (Shirey and Mindrup 1999), did not help in extracting other compounds from the orange juice headspace. A second observation is that the CAR-PDMS and the PDMS fiber shared a majority of the extracted peaks (47 peaks shared). However, the CAR-PDMS fiber was more selective toward polar compounds such as esters and aldehydes whereas the non-polar PDMS fiber extracted better larger terpenes and their derivatives (Figure 3.2). These differences in selectivity are likely due to the nature of the coating material of each fiber which defines the mechanism of binding onto the fiber and the compounds retained (as influenced by their polarity, volatility and molecular size). Our observations are in agreement with the theory since the CAR-PDMS fiber was able to retain well smaller analytes (MW<90) by adsorption most likely because of the porous Carboxen coating, and the liquid phase of the PDMS fiber favored more the absorption of non-polar analytes like terpenes (Shirey and Mindrup 1999).

The CAR-PDMS fiber was selected as the fiber of choice as it gave the best repeatability, broadest extraction of diverse compounds and the largest number of volatiles extracted from the headspace (esters, aldehydes, terpenes and their derivatives) compared to other fibers. Even though terpenes and their derivatives were better extracted by the PDMS fiber, we were confident that, under optimized extraction conditions, the selected fiber would extract this class of compounds since the fiber is composed as well of a liquid phase PDMS. One drawback of the chosen fiber (Carboxen-PDMS: adsorbent fiber) is the potential competition for sites when volatiles are in high concentrations compared to absorbent fibers (e.g., PDMS), and therefore, accuracy in quantification may suffer.

3.3.1.2. Optimization of the headspace SPME method

The number of volatiles extracted by a SPME fiber depends on the equilibrium between sample/headspace and headspace/fiber (Shirey and Mindrup 1999; Nongonierma et al. 2006). These equilibriums are compound-dependent based on vapor pressure, matrix interactions and affinity for the fiber polymer vs. the sample matrix. Equilibrium can be reached (in theory) with proper method optimization. In the present experiment, the incubation temperature, the extraction time and the addition of salt (NaCl) to orange juice samples were optimized so that the extraction of volatiles by SPME was maximized.

A RSM methodology was conducted to optimize the extraction parameters yielding to the best volatile recovery from the headspace and 17 experimental runs were performed according to the experimental design (Table 3.2). For each run, the number of detected peaks was recorded. A quadratic model was built to describe the observed data and evaluated by ANOVA. Only first- and second-order terms that were significant were included in the model. The regression equation (coded variables) obtained was:

$$\text{Peaks detected} = 70.2 + 3.5*T + 4.2*t + 7.2*S + 5*T^2 - 5.5*t^2 - 7.5*S^2$$

where T: temperature, t: extraction time and S: salt content. The model was statistically adequate to approximate the response in the region of experimentation as indicated by the lack-of-fit test (p value = 0.9), and to explain about 93 % of the variability in responses.

Results from the ANOVA indicated that the addition of salt significantly influenced (p value = 0.03) the number of detected peaks. Likewise, Steffen and Pawliszyn (1996) and Siegmund et al. (2003) reported that addition of salt to fruit juices samples had a positive influence on the extraction of the volatile compounds by a polyacrylate and CAR-PDMS fiber, respectively. Increasing the ionic strength of the sample favored the displacement of water soluble volatiles to the headspace (salting-out effect) (Nongonierma et al. 2006). On the other hand, the extraction time (p value = 0.09) and temperature (p value = 0.13) did not influence the extraction recovery although they tend to have a positive effect. Generally, increasing the extraction time allows more compounds to reach their equilibrium (liquid/gas) and consequently to be adsorbed onto the fiber. However, long extraction times can result in saturation of the fiber and/or competition for adsorption sites leading to

biased volatile profiles (Shirey and Mindrup 1999). Sample stirring and temperature increase are often used to speed up the equilibration time. Heating samples generally helps in extracting volatiles from the headspace, as seen in our results, by increasing the rate of mass transfer onto the fiber coating. Nonetheless, the retention of compounds with high volatility on the coating material can be adversely affected if too high temperatures are used due to a reduction of the partition coefficient fiber/headspace. No such detrimental effect was observed in our experiment.
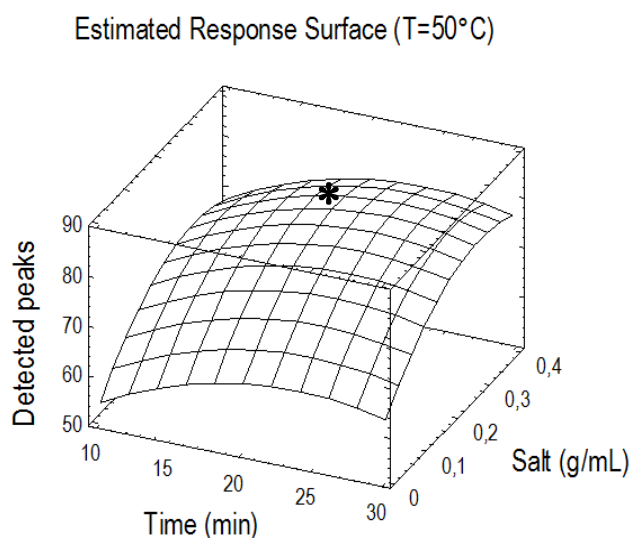
Figure 3.3: Response surface plot showing the variation in the number of peaks detected in orange juice headspace by a CAR-PDMS fiber at 50 °C relative to salt content and extraction time. * indicates the optimum conditions chosen for the headspace SPME-GC method.

The optimum extraction conditions for the SPME analysis were determined from the response surface plot obtained from the fitted model (Figure 3.3) so that the number of detected peaks was maximized. The curvature observed in the response surface indicated that intermediate salt levels and extraction times should be used when sampling at 50 °C (anticipated maximum that could be used with little artifact generation). The maximum response predicted was 81 peaks ([77; 86] ± 95% confidence interval) if samples were

incubated at 50 ℃ with 0.296 g/mL salt added and the fiber exposed for 23.8 min. The optimum conditions chosen for the rest of the study, in regards to ease of sample preparation (e.g., weighing salt) and throughput (e.g., 20 min vs. 23.8 min), are indicated with an asterisk in Figure 3.3. These conditions were: 50 ℃, 0.3 g/mL salt added, and 20 min fiber exposure.

The repeatability of the optimized SPME method to extract volatiles from the orange juice headspace was evaluated by the analysis of five replicates. Using the optimized conditions, a total of 78 peaks were detected in all five replicates which is in agreement with the predicted values from the fitted model (80 peaks detected [76; 85] ± 95% confidence interval). The averaged coefficient of variation (% CV) of the peak area across all detected peaks was 9.4 % so we considered the method was adequately repeatable.

3.3.2. Untargeted analysis of orange juice non-volatiles

A UHPLC-TOF-MS approach was developed to complement the coverage of potential flavor stimuli in the juice (non-volatiles). Once more, the challenge was to choose experimental conditions that did not affect the diversity of compounds studied. The influence of sample preparation, column stationary phase and ionization mode on compound coverage, chromatographic quality and repeatability was evaluated.

3.3.2.1. Influence of sample preparation

When studying the entire pool of compounds present in a system, minimal or no sample preparation is always appealing to the researcher in terms of integrity of the sample, ease of preparation, and limited variability between samples (no or few steps before injection). In this application, centrifugation and filtration were necessary steps to prevent column plugging due to the presence of large (> 2 μm) and fine (< 2 μm) suspended particles in the aqueous phase (serum)(Brat et al. 2003). The effect of further sample preparation on compound coverage and chromatographic quality was examined by comparing chromatograms obtained from the injection of orange juice diluted in water or SPE fractions. Only chromatograms obtained with the negative ionization mode are presented in

Figure 3.4 for comparison purposes (the effect of ionization mode is discussed in a following paragraph).

A first observation is that chromatograms obtained from the RP separation of the juice diluted in water and SPE-eluate were fairly similar, as well as for the HILIC separations. The large peak seen at the beginning of the RP run (< 1 min) for the juice diluted in water is likely due to the presence of highly hydrophilic compounds that were not retained on the C18 column; on the other hand, those compounds were not present in the SPE-eluate as they were collected in the SPE-wash fraction. Regarding the chromatograms obtained by HILIC separation, the signal was slightly higher (and peaks wider) for the juice samples diluted in water (1.15×10^5 counts) than for the SPE-wash samples (9.14×10^4 counts). The difference in signal intensity (i.e., concentration) is likely due to the sample preparation: recall that the SPE-wash fraction was obtained by washing-off hydrophilic compounds from the cartridge with 3 mL of water/ acetonitrile but was originally coming from 2 mL of juice (and therefore diluting the compounds). Finally, HILIC separation for the juice diluted with water (no SPE) gave few more peaks (e.g., 4.2 min) than the SPE-wash fraction even though they were not well separated. These additional peaks were from compounds of middle hydrophilicity as indicated by their elution time between 3.5 and 4.5 min which translated to about 85 and 75% of organic solvent in the mobile phase. They were not present in the SPE-wash fraction but most probably in the SPE-eluate fraction indicating the SPE technique roughly divided the sample into two fractions (highly hydrophilic and relative hydrophobic) so they were detected only in one of the chromatographic separation. Another example, two unknown compounds (m/z 649.2575 and 693.2859) were present in both RP (3.80 and 4.98 min, respectively) and HILIC (4.22 and 3.74 min, respectively) traces from the juice diluted in water while they were detected significantly only in one of the SPE fraction. In other words, the SPE fractionation minimized the redundancy of data generated (more particularly for compounds of mid-polarity) when both RP and HILIC separations are used on the same sample (Idborg et al. 2005a). If no SPE fractionation was used, one possible way to get around the issue would be to inspect the datasets and remove redundant data. However, this task is cumbersome when large amounts of data are

collected and also because the compounds do not have the same retention times on both stationary phases.

Consequently, we decided to use a SPE procedure for the remainder of the study and results obtained from juice diluted in water will not be discussed further. Sample fractionation by SPE minimized the redundancy of the data but it is also a good practice to increase column life (Dettmer et al. 2007) even if it involves more preparation time compared to a dilution. At this point of method development, we decided to analyze both SPE fractions as discarding one could have been a potential loss of information. Whether or not their analyses provided a sufficient gain in information was assessed during the evaluation of the methods (Chapter 4).
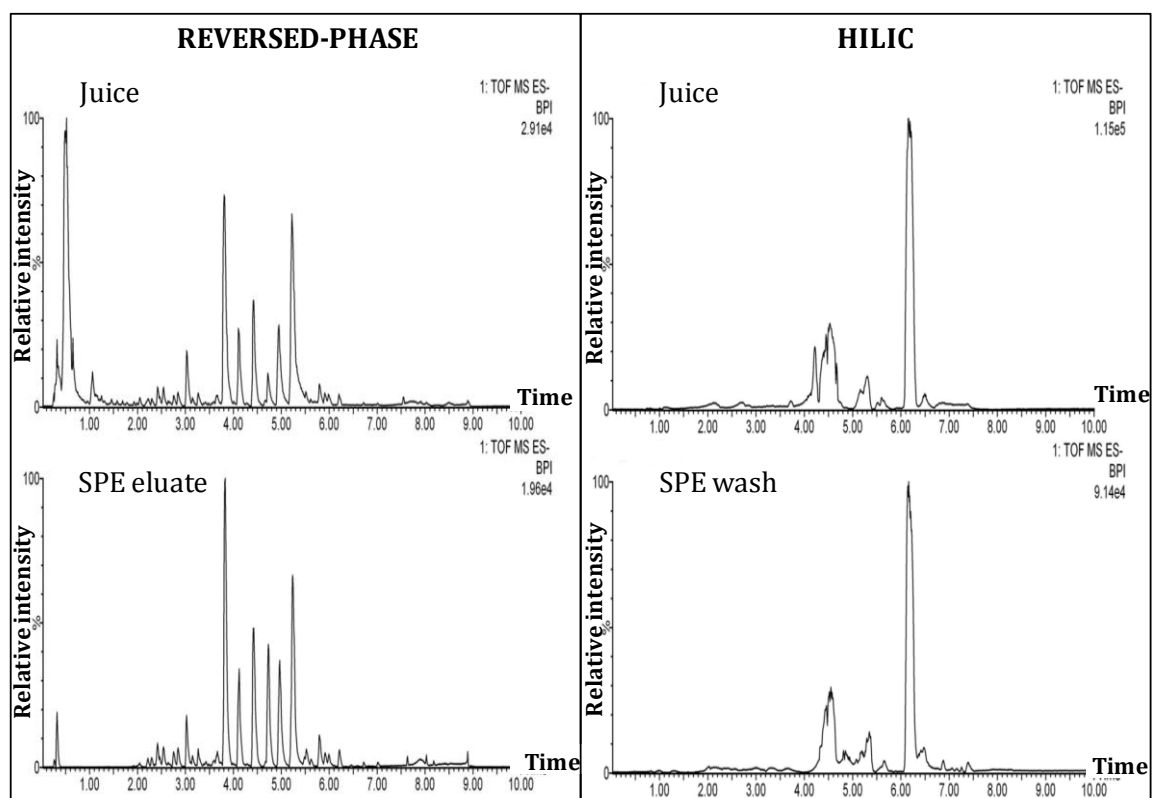
Figure 3.4: Comparison of BPI traces of orange juice samples prepared either by dilution with water or by SPE and analyzed by ESI (-) UHPLC-TOF-MS. Each BPI chromatogram is normalized to its most intense peak.

3.3.2.2. Analysis of SPE fractions by RP and HILIC

The objective behind sample preparation by SPE was to simplify the sample by dividing it into two fractions (highly hydrophilic and relative hydrophobic compounds) thereby making analysis by LC-MS easier. Due to the difference in polarity of the SPE fractions, two complementary column phases were used for their analysis. The eluted fraction was separated by RP (C18) and the washed fraction was injected onto an HILIC column (amide stationary phase). Indeed, highly polar and/or ionic compounds are not retained by C18 stationary phase and generally elute in the void volume of the column without being analyzed. HILIC columns have been successfully used - among others - for the analysis of polar metabolites in rat urine (Idborg et al. 2005a; Cubbon et al. 2007) and plants (Tolstikov and Fiehn 2002). In preliminary trials (not reported), the redundancy of the fractions (compounds present in both fractions) was examined by injecting some washed fractions on the C18 column and some eluted fractions on the amide column: no major redundancy between the results was observed. The gain in information from using complementary stationary phases for the untargeted analysis of orange juice non-volatiles was evaluated by comparing the chromatographic profiles obtained for SPE-eluate (RP) and SPE-wash (HILIC) (Figure 3.6). We have later examined (Chapter 4) if using all of the MS information gathered for both SPE fractions and both ionization modes contributed efficiently to differentiate the juices.

The total number of chromatographic peaks (in BPI chromatograms) detected in each fraction is summarized in Table 3.4; no efforts were made at this stage of method development to detect co-eluting compounds. At first the number of peaks detected overall may seem low but we need to recall that only chromatographic peaks (not compounds) automatically detected by Markerlynx ApexTrack peak integration tool were counted, while ignoring compounds with a low signal or those co-eluting. As noted earlier in the SPME section, we will provide data in Chapter 4 on individual "features" (unique RT-m/z) that gives a more accurate view of the compounds being measured.

Table 3.4: Number of chromatographic peaks (see text for explanation) detected in orange juice SPE fractions by UHPLC-TOF-MS in positive (+) and negative (-) electrospray ionization modes. Data are presented as the average of five replicates and associated % CV.

	SPE-eluate (RP)		SPE-wash (HILIC)	
	+	-	+	-
mean	26	31	16	26
% CV	3	4	5	2

Overall, there is an unequal contribution from washed and eluted fractions in terms of the number of chromatographic peaks obtained, independent of the ionization mode. The number of peaks detected in the SPE-wash fraction (16 and 26 peaks in ESI positive and negative, respectively) was lower than in SPE-eluate fraction (26 and 31 peaks in ESI positive and negative, respectively). However, analyzing both fractions allowed us to collect additional information (potentially, different chemical stimuli that could be related to the sensory attributes) about the samples that would have been lost if the washed fraction was discarded. This is important, more particularly, for highly hydrophilic compounds like carbohydrates and organic acids which are known to affect the sensory perception of juices and fruits (Ahmed et al. 1978c; Baldwin et al. 2008). Very few studies have reported on analyzing both SPE fractions (Idborg et al. 2005a): most metabolomic studies using SPE fractionation generally consider only one fraction and ignored compounds present in the other. In addition, most research is based on RP separation only (Dettmer et al. 2007), and often highly polar and/or ionic compounds are ignored. The benefit of using HILIC separation, in combination with RP, has been shown to increase the number of metabolites investigated (Cubbon et al. 2007) and to allow the discovery of additional markers for the differentiation of samples (Gika et al. 2008).

The chromatographic quality (peak width and peak shape) obtained for both columns was also examined. UHPLC columns have been reported to give better chromatography than conventional LC columns due to smaller particle sizes packed in shorter columns, higher flow rates and higher pressures; consequently, greater column plate number and a greater peak capacity are achieved (Swartz 2005). The elution gradients (Table 3.3) were chosen to

have an acceptable separation over a 10-min run (our objective for high-throughput) in both positive and negative ionization modes. We recognize that the elution conditions may have been further optimized, especially for HILIC separations for which chromatograms appear to be poorly resolved (Figure 3.4, right vs. left). However, one of the advantages of using MS detection is the lack of need to obtain well resolved chromatographic peaks: using individual extracted ion chromatograms (EIC) gave overall good peak separation which facilitated the subsequent data processing steps.

Two extracted ion chromatograms (randomly selected) are displayed as an example of column efficiencies (i.e., the ability of the column to furnish narrow peaks) and peaks shape obtained using RP and HILIC (Figure 3.5). Both RP and HILIC columns gave overall very narrow peaks (average half-height peak width for EIC = 0.042 min and 0.041 min with RP and HILIC separations, respectively) and reasonable Gaussian-shaped peaks, with no major fronting or tailing. Few peaks were poorly-resolved on the HILIC column (amide phase), most likely due to column overloading or inappropriate mobile phase.
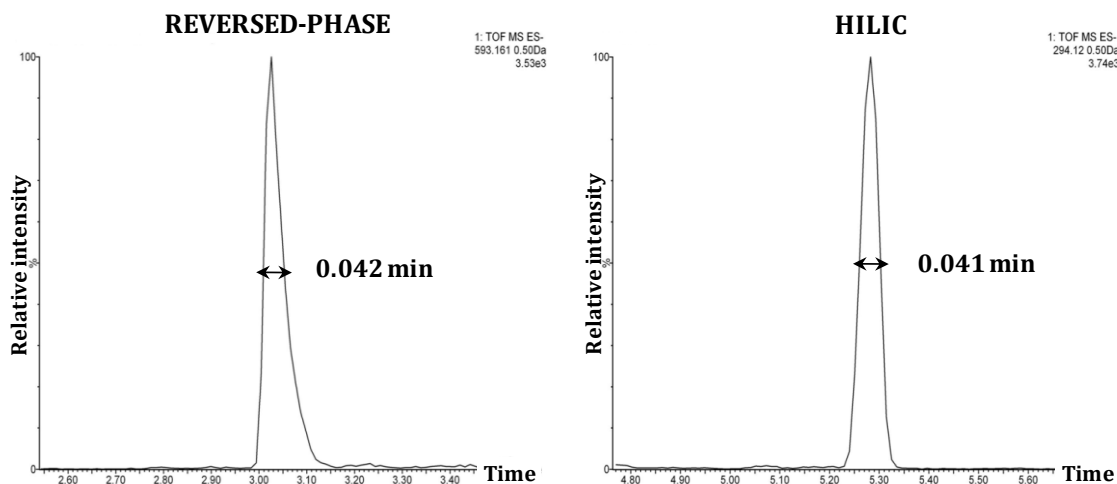
Figure 3.5: Representative chromatographic peak width of an extracted ion chromatogram obtained by UHPLC RP (m/z 593.161) and HILIC (m/z 294.120) separation using ESI (-) TOF-MS detection.

These parameters (peak width and peak shape) are critical to get good peak separation and increased resolution which both govern the quality of peak-picking used in subsequent multivariate analyses (Chapter 4). Also, the maximum drift in retention times observed for randomly selected peaks across five replicates was about 0.0088 min for RP, and 0.0138 min for HILIC separations (Table 3.5). These very small retention time shifts gave us confidence for using automatic chromatogram alignment across samples in further processing of the data.

Table 3.5: Variation in retention time and signal intensity of randomly selected peaks showing the precision of the LC-MS method. Data are presented as the average of five replicates.

		RP			HILIC	
	m/z	RT (± min)	Intensity (% CV)	m/z	RT (± min)	Intensity (% CV)
ESI (+)	227.1741	2.2412 (0.0004)	1192 (18)	226.1260	2.2092 (0.0004)	4615 (20)
	595.1675	3.0200 (0.0060)	1384 (9)	144.0990	4.8114 (0.0047)	52080 (5)
	303.0864	4.4094 (0.0060)	1930 (7)	136.9392	5.5532 (0.0074)	1496 (25)
	533.2399	5.2296 (0.0005)	1350 (7)	314.1609	5.9698 (0.0049)	7173 (11)
	728.3986	6.2004 (0.0047)	761 (8)	175.116	7.3338 (0.0093)	1888 (12)
ESI (-)	443.1971	2.419 (0.0000)	1532 (6)	351.1336	2.4866 (0.0094)	1144 (18)
	625.1630	3.5550 (0.0055)	754 (10)	205.0356	4.3858 (0.0138)	2040 (11)
	693.2831	4.9714 (0.0088)	6802 (12)	179.0556	5.1260 (0.0000)	1710 (20)
	593.1942	5.8026 (0.0048)	2006 (5)	341.1107	6.1456 (0.0060)	110369 (5)
	793.4729	7.0164 (0.0005)	342 (9)	503.1662	6.8560 (0.0055)	1502 (13)

3.3.2.3. Influence of ionization mode

ESI is the most commonly used ionization technique for LC-MS based metabolomic studies, and most frequently is operated only in the positive ionization mode. However, due to the variety of compounds present in orange juice (e.g., polyphenols (flavones), sugars, organic acids) potentially having very different ionization capabilities, it was difficult to predict which ionization mode would give the most comprehensive information regarding the sample composition. Consequently, separate UHPLC-MS analyses of SPE fractions were conducted using ESI in both the positive and negative ionization mode. Some researchers have suggested the use of other ionization techniques alongside ESI to expand compound coverage and favor the detection of compounds that are not ionized well by ESI. For example, the use of atmospheric pressure chemical ionization in addition to ESI resulted in 20 % more unique ions being detected in human serum samples (Nordstrom et al. 2007). Despite this recognition, no other ionization techniques were attempted in the present study.

The base peak ion (BPI) chromatograms of the SPE fractions analyzed by ESI – and + modes are presented in Figure 3.6. As expected, very different profiles were generated depending on the polarity of the ionization source. One example of the differences in ionization abilities of the compounds is demonstrated with the SPE-wash fraction (Figure 3.6, right). Using the negative ionization mode, one major peak at 6 min dominated the signal (and minimized the contribution to total signal from other peaks) whereas using the positive mode allowed the detection of other peaks before 6 min. Similarly in the SPE-eluate chromatograms, late peaks (> 6.5 min) were detected in the positive mode while their signal was lower or not recorded in the negative mode. The total number of peaks detected in each fraction with each ionization mode is summarized in Table 3.4. Overall, more peaks were detected using the negative ionization mode for both fractions (31 vs. 26, and 26 vs. 16 peaks in RP and HILIC, respectively) and both ionization modes offered little variation in terms of peaks detected ($\leq$ 5 % CV). Also, the signal intensity was higher in the negative mode (1.96×10^4 vs. 1.28×10^4 counts for SPE-eluate and 9.14×10^4 vs. 7.59×10^4 counts for SPE-wash in negative and positive mode, respectively) offering better sensitivity and indicating that there are probably more acidic compounds or acidic functional groups in the

extracts. The variation in signal intensity of randomly selected peaks obtained from the analysis of five SPE fraction replicates was between 5 and 25 %, and was higher for peaks separated by HILIC than by RP, and peaks detected using positive ionization (Table 3.5).
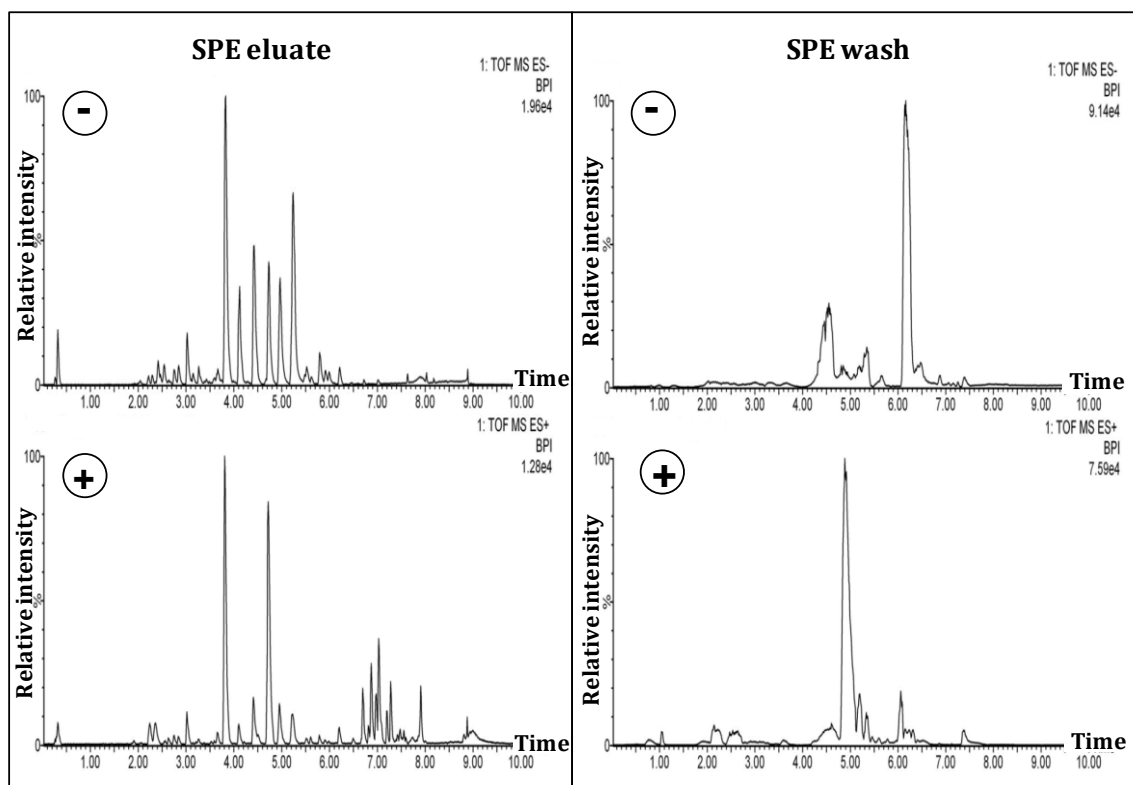
Figure 3.6: Influence of MS ionization mode on BPI chromatographic profiles of SPE fractions from orange juice. Each BPI chromatogram is normalized to its most intense peak.

These results confirmed the benefit of using both ionization modes to increase the number of compounds detected over single-ionization mode, as other studies have pointed out (Waybright et al. 2006; Cubbon et al. 2007; Nordstrom et al. 2007). Yet, a potential downside of using dual ionization (besides increasing the analysis time if dual ionization cannot be done simultaneously) is the redundancy of information generated. Some compounds have the ability to be ionized both in the positive and negative modes and thus

they may be included twice in the results with possibly a different fragmentation pattern, which makes it difficult to determine that the same compound is included in the data set twice. For example, hesperidin (an O-disaccharide-substituted flavanone from orange juice) has been reported to generate fragments m/z 301, 463, 609 in the negative mode and m/z 303, 449, 465, 611 fragments in the positive mode (Gattuso et al. 2007). If a compound is reported twice, its contribution in the consecutive multivariate analyses may be artificially inflated if the data sets (from positive and negative ionization) are merged. However, the double positive/negative detection would be an advantage if the data sets are analyzed separately and the compound is highlighted as a "flavor marker" by both data analysis. Currently, we do not have a way to identify and automatically reduce the redundancy between data sets.

3.4. CONCLUSIONS

In summary, this chapter reports on the development of analytical methods for the untargeted analysis of compounds in orange juice. The coverage of compounds responsible for the chemical characterization of orange juices was comprehensive since it included volatiles and non-volatiles. Even though it was not possible to be totally unbiased in terms of the compounds being analyzed, a careful selection and optimization of the sample preparative techniques and instrumental methods employed helped in maximizing compound extraction and detection with a satisfying throughput, and reproducibility. An automated headspace SPME extraction coupled with GC-TOF-MS analysis was used for the analysis of volatiles. Of the SPME fibers screened, CAR-PDMS fiber was found to be best suited for our purpose. The extraction parameters were optimized using a response surface methodology and the addition of salt to the juice significantly influenced the recovery of volatiles. A UHPLC-TOF-MS method was developed to complement the coverage of potential flavor chemical stimuli in the juice (non-volatiles). The use of a SPE fractionation simplified the analysis of non-volatiles present in the juice with no loss of information when compared to untreated samples (no SPE). The analysis of all SPE fractions with two columns (reversed-phase and hydrophilic interaction liquid chromatography) decreased the

redundancy of data generated and the use of electrospray ionization in negative and positive modes increased the number of compounds analyzed and detected. The stability of the methods over long time periods was not evaluated here. In future studies, the use of the quality control samples and the addition of internal standards are recommended to adjust for variations in recovery between successive SPME fibers (replacement) and differences in MS ionization efficiency between days of measurements (and to prevent MS data clustering based on day effect).

Chapter 4 : Evaluation of instrumental methods for the classification of orange juices using mass spectral information and multivariate data analyses

---

Abstract. The ability to discriminate commercial orange juices using untargeted mass spectral-based methods (Chapter 3) was examined. Four commercial juices were analyzed by gas chromatography (GC) and liquid chromatography (LC) - mass spectrometry (MS) for volatiles and non-volatiles, respectively. These data were successfully used for the classification of the juices using principal component analysis and partial least squares - discriminant analysis. Both methods allowed the categorization of the juices: the models obtained were robust and the samples were clustered by brands, with little overlap for hydrophilic interaction liquid chromatography (HILIC) data. The tight grouping of replicates suggested the good precision of the methods, especially for GC and reversed-phase (RP) analyses. The LC analyses appeared slightly biased toward one juice but still provided adequate information on chemical differences between the brands to differentiate the juices. The markers originating from HILIC analyses were not specific enough to be used reliably, as opposed to those from RP separation and GC analyses. Using both ionization modes increased the number of markers considered for juice differentiation. Fusion of GC-MS and RP-UHPLC-MS data sets gave similar classification models compared to that of using only data from volatiles or non-volatiles but with the advantage of finding potential interactions between volatiles and non-volatiles.

4.1. INTRODUCTION

In the previous chapters, the concept and reasons for using flavoromics to study flavor perception were presented. It is hypothesized that using an untargeted instrumental methodology which includes both volatiles and non-volatiles can provide better correlations between the chemical composition of a food (i.e., orange juice) and its flavor attributes than the traditional approach of using only volatile compounds in this correlation. However, this approach presents substantial analytical (discussed in Chapter 3) and data processing challenges (discussed in this chapter). The data processing challenges relate to handling the vast amount of data gathered, and extracting from it chemical/biological information relevant to a problem.

In this study, we used a mass spectrometry (MS) based profiling approach to characterize instrumentally orange juices. When using MS-based approaches, the amount of data generated is very large given that each time point of a chromatogram is associated with a mass spectrum (whose range depends on the instrument and user's choice). The information produced rapidly becomes overwhelming and data handling becomes a hurdle. Chemometrics is often used to graphically summarize analytical data so that it can be interpreted further in terms of biological meaning. The most common multivariate data analyses used in metabolomics studies are principal component analysis (PCA) and partial least squares (PLS) (Eriksson et al. 2006e; Sumner et al. 2007). Briefly, PCA is an unsupervised method in the sense that no additional knowledge (e.g., sample class) besides raw data is required to describe the data set. PCA gives an overview of the data which is useful to detect outliers, common groupings between samples and to evaluate the relationships between samples and variables (i.e., compounds measured), and between variables themselves. Partial least squares-discriminant analysis (PLS-DA) is a supervised method used to maximize the separation between classes of observations in the multivariate space according to their class membership (e.g., juice brand, type of cultivar, etc.), and to understand which variables contributed to the separation (Eriksson et al. 2006e).

The objectives of the present study were to test the handling and chemometric analysis of the massive amounts of data coming from different instruments (GC- and UHPLC-TOF-MS), and to evaluate if the mass spectral information collected from the developed instrumental methods (Chapter 3) were adequate to detect chemical differences between commercial orange juices even though they are technically very similar products (all are premium juices, which have received a mild thermal treatment). For this purpose, we analyzed four commercial premium orange juices. Our theory was that if we could not detect chemical differences between these juices, there would be little probability that we would be able to make use of the instrumental data to predict flavor quality or to discover "flavor markers" associated with specific sensory attributes. Data resulting from the analyses of the commercial juices and a blend thereof were processed using chemometric tools (PCA and PLS-DA). The results of the multivariate data analyses of each data set were compared for their quality and ability to classify samples in order to determine the value of information generated by each instrumental analysis. We would like to emphasize that the juices used in the present study were chosen for method development/evaluation purposes rather than to investigate a specific flavor-related question (e.g., juices made from healthy vs. infected orange fruits in the case of "citrus greening" disease).

4.2. MATERIALS AND METHODS

Orange juice samples. Four commercial premium orange juices (referred to as brand A, B, C and D), all pulp-free, were purchased at a local grocery store. The juices were chosen as a sample of commercial premium orange juices available and only for method development purposes. All juices were transferred from their original packaging (either gable-top paperboard carton or PET bottle) to glass Erlenmeyer flasks and stored at 4 °C until sample preparation in the following days. A reference sample was prepared by blending equal proportions of all four juices (v/v) and was used to evaluate the discriminative performance of the proposed methods. Five replicates of each juice and blend were prepared for each analysis and analyzed in random order.

Chemicals and supplies. Purchase details and information related to the materials used are described in the Materials and Methods section in Chapter 3.

Analysis of orange juice volatiles. All samples were prepared and analyzed according to the optimized SPME parameters; the complete description of the method and instrument settings used is given in Chapter 3. Briefly, samples were incubated at 50 °C for 5 min under constant stirring (500 rpm) before the SPME extraction. Following, a 75 μm carboxen/ polydimethylsiloxane fiber was exposed for 20 min to the headspace of 10 mL orange juice containing 0.3 g/mL salt (NaCl) in a 20 mL vial and maintained at 50 °C. The fiber was then desorbed into the injection port of a gas chromatograph coupled with a time of flight-mass spectrometer (GC-TOF-MS).
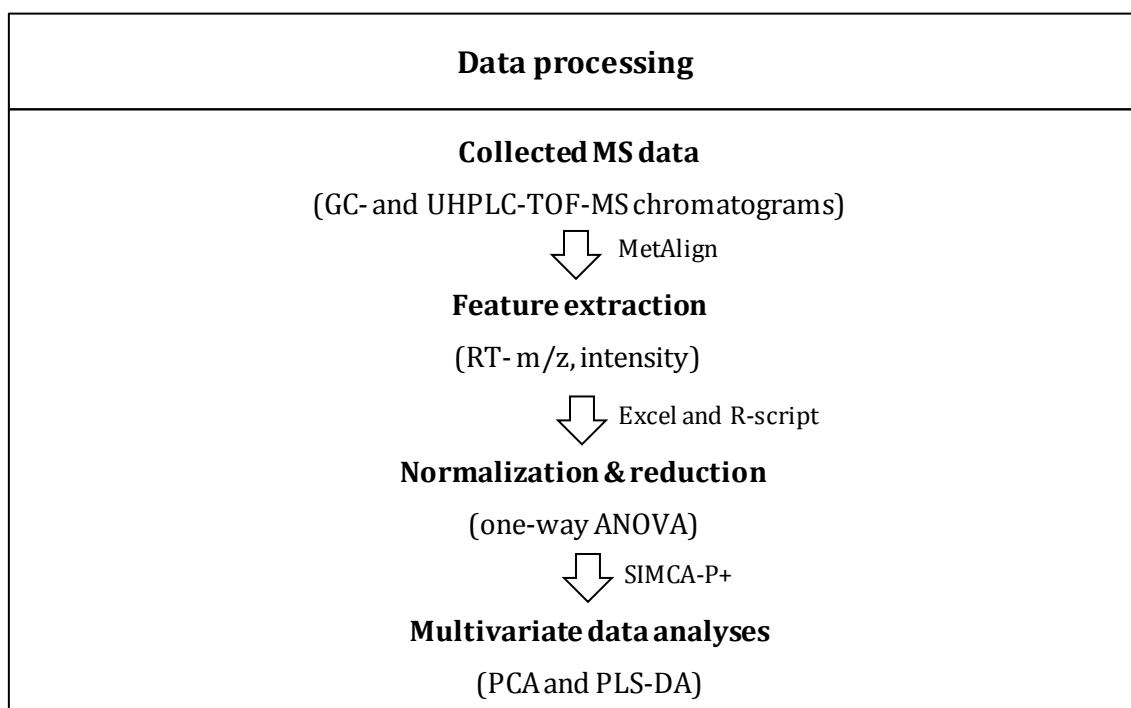
Figure 4.1: Overview of the steps used for processing the MS data collected from the analysis of orange juices.

Analysis of orange juice non-volatiles. Compounds not analyzed by headspace SPME were isolated from the juice matrix by solid phase extraction (SPE) before analysis by ultra high performance liquid chromatography (UHPLC); a detailed description is given in Chapter 3. Briefly, the clarified juice obtained after centrifugation was immediately used for the SPE. The two collected fractions, SPE-wash (highly hydrophilic) and SPE-eluate (relative hydrophobic), were injected onto a HILIC column and RP column, respectively. UHPLC column effluents were directed into a quadrupole-TOF-MS equipped with an electrospray ionization (ESI) source operating first in positive and then, in a separate analysis, in negative mode.

Data processing and multivariate analyses. The data processing steps are summarized in Figure 4.1. The chromatograms generated from GC- and UHPLC-TOF-MS analyses were used to extract chemical information about each sample. The MS raw data files were converted to *.cdf format and pre-processed using the freeware Metalign (http://www.metalign.nl). The program has been described elsewhere (de Vos et al. 2007; Lommen 2009) and successfully used in metabolomic research. In short, the program extracts mass spectral information from the chromatograms after baseline and noise correction, and aligns chromatograms across samples. The output is a table containing a list of the features (x columns) detected across samples (y rows) and their corresponding intensities; a feature being defined as a unique combination of a mass-to-charge ratio (m/z) detected at given retention time (RT) (i.e., pair RT-m/z). After processing with MetAlign, the output table (*.csv format) was manipulated in Excel (Microsoft Office 2007) and the intensities of the features (also referred to as variables) detected in each sample were normalized (intensity of the feature/ sum of intensities of all features in the sample $\times$ 10,000). The normalized data sets were reduced using a in-house script developed in R (http://www.r-project.org) which discarded features that did not differ significantly between sample classes (one-way analysis of variance ANOVA, $P < 0.05$); meaning, retaining only features that contribute the most to the variance of the data set. The reduced data sets were then imported into Simca-P+ statistical software (v. 12.0, Umetrics, Umea, Sweden) for the multivariate data analyses (PCA and PLS-DA). Five data outputs were used (one per type of instrumental analysis): GC, SPE-eluate separated by RP with ESI in positive (RP pos.) and

negative (RP neg.) ionization mode, and SPE-wash separated by HILIC with ESI in positive (HILIC pos.) and negative (HILIC neg.) ionization mode. After Pareto scaling ($1/\sqrt{SD}$, where SD is the standard deviation) and mean-centering for each variable, data sets were first inspected for outlier samples by PCA. Samples outside of the 95 % tolerance ellipse given by Hotelling's T^2 were not considered for the remainder of the data processing (outliers). The PCA and PLS-DA models obtained were examined for their quality (statistics R^2 and Q^2) and their ability to distinguish between juices. Data sets (GC and UHPLC) were initially processed separately and then fused into a single data set using low-level data fusion (i.e., no preliminary variable selection) (Smilde et al. 2005), only for the instrumental methods judged adequate to reveal the differences in chemical composition of the juices and thereby being used for flavoromics.

Tentative compound identification. The PLS-DA weight plots were inspected to locate the features (variables) responsible for the separation between the juices (found on the periphery of the plots). No efforts were made to identify all compounds or to confirm their identity as this task was not relevant to the present objectives and would have required considerable additional work. For volatiles (GC data set), the mass spectrum associated with the selected feature (RT-m/z) was retrieved from the chromatogram and compared to Wiley Registry with NIST 08 mass spectral library. Compounds were tentatively identified based on the similarity of their mass spectrum with the library hit and sample knowledge. For non-volatiles, the accurate mass of the feature was used to search the online Scripps center Metlin database (http://metlin.scripps.edu/metabo_search.php). The use of MS/MS or NMR spectroscopy in future stages of the research will offer a means for further structure elucidation and confirmation of compound identities.

4.3. RESULTS AND DISCUSSION

The objective of the study was to evaluate if the developed methods were powerful enough to distinguish samples, meaning powerful enough to detect chemical differences between different but yet similar orange juices in terms of processing (i.e., mild thermal treatment using mass spectral information and multivariate data analyses. The results of

the multivariate data analyses (PCA and PLS-DA) of each data set were compared for their quality and ability to classify samples in order to determine the value of information generated by each instrumental analysis (Table 4.1). We wish to add a note here that the classification of the samples was not the goal of the study but rather an indirect way to test the instrumental methods, the handling and chemometric analysis of these data. Consequently, we did not invest much effort in the identification of the markers responsible for groupings between juices.

4.3.1. Sample classification based on volatiles

Once again we would like to remind that the purpose was not the classification of orange juices by headspace SPME-GC-MS data and chemometrics, as it has been already reported in the literature (Shaw and Moshonas 1997; Shaw et al. 1999; Lotong et al. 2003), but rather to use those instrumental measurements as inputs for predictive models of flavor (Chapter 6). Therefore, there is a need to confirm that the method is able to detect differences in volatiles between sample types.

GC-MS chromatograms were processed with MetAlign and resulted in 3471 features (or variables) detected across the 25 samples (observations). The dimensionality of the data matrix was reduced to 1902 features using a one-way ANOVA, so that only variables that changed between sample classes ($P<0.05$) were used for differentiation. One should remember that the number of detected features in the GC chromatograms does not reflect the actual number of compounds detected in the entire set of samples; indeed, electron-impact, which is a hard ionization technique, generated multiple fragments for the same compound. We are currently evaluating a way to filter out fragments from the same compound and to select only the most representative fragment.

PCA was applied to the GC data set and a three-component model was obtained with an explained variation (goodness of fit) $R^2X= 0.77$ and a predicted variation (goodness of prediction) $Q^2X= 0.66$ (Table 4.1). These results are regarded as good because R^2X is high, $Q^2X > 0.5$ and that the difference between R^2X and Q^2X does not exceed 0.2 (Eriksson et al. 2006e). The first three principal components (PC) explained 40.5 %, 25 % and 11.5 % of the

total variation of the data set, respectively. The PCA score plot (not presented) was examined to see how samples related to each other, recalling that samples being close to each other in the score plot have similar volatile profiles. We observed a clear demarcation of the juice samples by brand (denoted A-D) with no overlap, indicating that the GC-MS method was able to reveal the differences in the volatile composition of the juices and thereby being used for flavoromics. This was further confirmed by the PLS-DA results (R^2X=0.77, R^2Y=0.72, Q^2Y=0.68).

Table 4.1: Statistics (R^2Xcum, R^2Ycum and Q^2cum) and number of variables used for the first three principal components of PCA and PLS-DA models obtained from GC-TOF-MS and UHPLC-TOF-MS analyses of orange juices.

	Number of variables [a]	PCA		PLS-DA		
		R^2Xcum [b]	Q^2cum [d]	R^2Xcum [b]	R^2Ycum [c]	Q^2cum [d]
GC	1902	0.77	0.66	0.77	0.72	0.68
RP neg.	399	0.72	0.56	0.71	0.72	0.56
RP pos.	179	0.76	0.62	0.74	0.70	0.63
HILIC neg.	12	0.78	0.40	0.77	0.61	0.45
HILIC pos.	25	0.86	0.72	0.86	0.53	0.36
Fused data (GC & RP)	2480	0.71	0.61	0.71	0.73	0.64

[a] Number of variables after reduction using one-way ANOVA; [b] R^2Xcum represents the cumulative fraction of the variation of the data set (X) explained by the 3 PCs; [c] R^2Ycum represents the cumulative fraction of the variation of the sample classes (Y) explained by the 3 PCs; [d] Q^2cum represents the cumulative predicted fraction, according to cross-validation, of the variation of the variables explained by the 3 PCs (goodness of prediction of PCA and PLS-DA models).

The grouping obtained with the PLS-DA was very similar to the PCA: the score plot (Figure 4.2) showed that the four commercial juices were clearly separated from each other and that each class was tightly clustered demonstrating good repeatability of the method. Brand D samples were located in the upper right-hand corner suggesting that both PC 1 and PC 2 governed its differentiation from other juices, whereas brands A and C were mostly separated according to PC 2; and brand B according to PC 1. Further, the GC-MS method was not biased toward one type of juice since the reference samples were clustered at the center

of the PCA score plot; recalling that the reference samples were prepared by pooling equal quantities of all four brands.
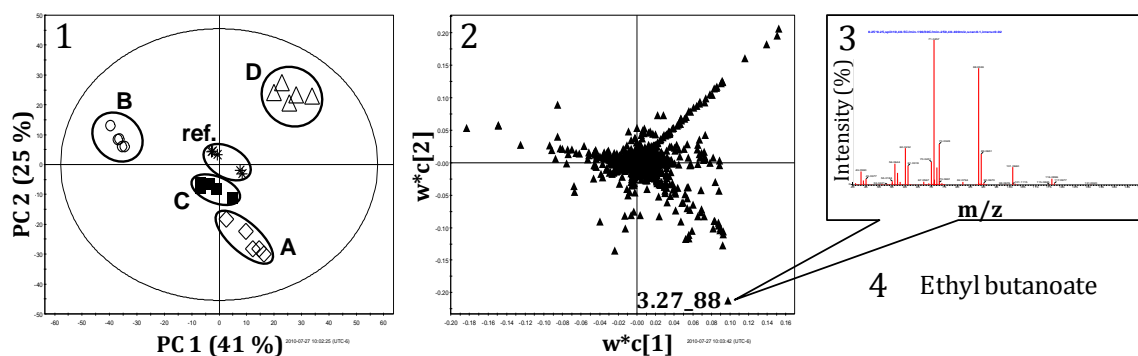
Figure 4.2: PLS-DA score plot (1) and weight plot (2) showing the separation of the four commercial premium orange juices and their blend based on the volatiles present in the headspace (GC-TOF-MS data). The tentative identification of a marker (3.27 min - m/z 88) is done by retrieval of corresponding mass spectrum (3) from the chromatogram and compared to the spectral library for identification (4).

As it is done in metabolomic studies using volatile patterns for the discovery of markers (genotype groups (Tikunov et al. 2005), disease (Prithiviraj et al. 2004), etc.), the inspection of the weight plots (Figure 4.2) from the multivariate models revealed which compounds (given as RT-m/z) contributed the most to the categorization of the juices. Ethyl butanoate (CAS 105-54-4) and valencene (CAS 4630-07-3) were positively associated with juices A and C, whereas β-myrcene (CAS 123-35-3) and α-terpineol (CAS 98-55-5) were positively associated with juices B and D, respectively. The sensory character and origin of these volatiles have been reported elsewhere (Perez-Cacho and Rouseff 2008a) and will not be discussed here. Such information is valuable to a juice manufacturer since it points out the chemical differences (aroma compounds) between two juices; investigating the sensory impact and origin of these volatiles can provide direction on altering the sensory profile of the product. Although the information was not available to us, we anticipate that the observed differences in volatile profiles are explained either by variations in raw materials,

or processing and storing conditions (Perez-Cacho and Rouseff 2008a; Perez-Cacho and Rouseff 2008b; Rouseff et al. 2009). These results are not discussed further in terms of flavor meaning. Again, the juices were not selected to address a research question so a discussion about their flavor differences was not considered relevant here.

4.3.2. Sample classification based on non-volatiles

Once more, the goal was to verify if the LC analyses collected enough information to distinguish variations in non-volatiles between samples. If successful, non-volatiles will be included as part of the chemical inputs in developing predictive models of flavor. As stated in the introduction and evidenced by the literature (although most of food flavor studies do not include them), there is potential value in considering non-volatiles as contributors to flavor perception.

Similarly to GC data, UHPLC-TOF-MS chromatograms were processed with MetAlign and filtered out with a one-way ANOVA. Four outputs were generated upon processing: RP neg., RP pos., HILIC neg. and HILIC pos. Data gathered for SPE-eluate resulted in 609 features in the positive ionization mode and 750 features in the negative mode; and when analyzing the SPE-wash, 117 were found when using the positive ionization mode and 135 features in negative mode (before applying ANOVA). Even though ESI is a soft ionization technique, we should mention that not all of the detected features represented individual compounds. Indeed, some features may have resulted from adduct and isotope formation or in-source fragmentation/ dimerization. Removal of isotope fragments, dimers and adducts might be necessary to reduce the noise in the data set but it was not considered in this work. The dimensionality of the data matrices was then reduced to 399, 179, 12 and 25 features for RP neg., RP pos., HILIC neg. and HILIC pos., respectively.
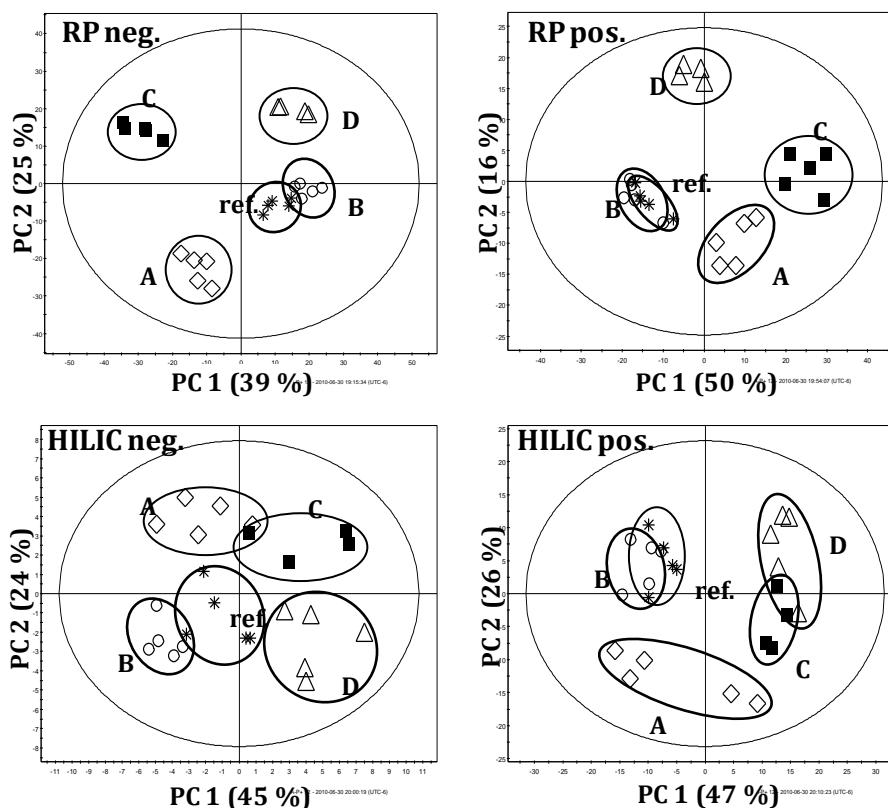
Figure 4.3: PCA score plots (t_1/t_2) showing the separation of the four commercial premium orange juices and their blend based on the non-volatiles (UHPLC-TOF-MS data). The plots show that the juices are clustered by brands suggesting that each method provided enough information for the classification of the juices.

PCA was applied to each data set and gave four score plots (Figure 4.3). Overall, the score plots showed a good separation of the orange juices according to their brand, and with a little overlap between juices C and D for HILIC data sets. Similar observations were made using PLS-DA; the quality of the classification models obtained is summarized in Table 4.1. The juice blend (reference sample) was slightly out-centered alongside PC 1 and tended to overlap with juice B for both RP and HILIC data, but more so with ESI pos. This suggested that the LC method may have a minor tendency to better extract/detect compounds similar to those present in juice B, but it still provided enough diversity in terms of chemical

compounds analyzed to differentiate the other juices. HILIC multivariate data analyses offered higher variability than RP as seen by the greater dispersion of the replicate samples on the score plot (Figure 4.3). This is probably because the number of relevant variables used for the PCA and PLS-DA models was much smaller than the other instrumental techniques (see Table 4.1) and therefore, the models were less robust to variations. Similar groupings were observed for each ionization mode, independent of the separation column used. Of note, samples appeared to be divided according to the origin of the oranges used in the juice manufacture (as declared on their labels) along PC 1 and PC 2 for HILIC and RP data sets, respectively: juices A and B (made with oranges from Florida) were separated from juices C and D (made with oranges from U.S.A. and Brazil). We do not have sufficient information on the products analyzed (e.g., heat treatment, cultivar) to explain the differentiation of the juices according to the direction of the other PCs.

One application of flavoromics is to uncover/identify compounds (markers) resulting from changes (in processes, raw materials, etc) and evaluate their relation to differences in flavor attributes. If this is the research goal (vs. prediction of flavor) then it is critical to verify if the variables (features) governing the classification in PLS-DA models obtained from UHPLC-TOF-MS analyses are reliable markers of the differences between samples. In other words, if the feature's normalized intensity was much higher/lower for a given class of samples and therefore indicating its specificity for that class. Overall, the examination of the trends plot (intensity of a feature across all samples) revealed that markers from the RP data sets were very discriminative of sample classes whereas markers from the HILIC data sets were not as specific. As an example, Figure 4.4 illustrates the differences in marker's class exclusivity between RP and HILIC data. The marker's trend across samples for RP pos. showed a very clear "stair step" indicating that the marker (6.98 min - 433.1493 m/z) was present at higher levels for some brands and absent from the others, and therefore specific to these classes. On the other hand, no such clear demarcation was observed for the marker from HILIC pos. (2.20 min - 226.1262 m/z) and presented more intra- and inter- sample classes variability. Unlike in the study of Idborg et al. (2005b), the analysis of SPE-wash in addition to the SPE-eluate was not successful in detecting additional markers for the categorization of the juices.
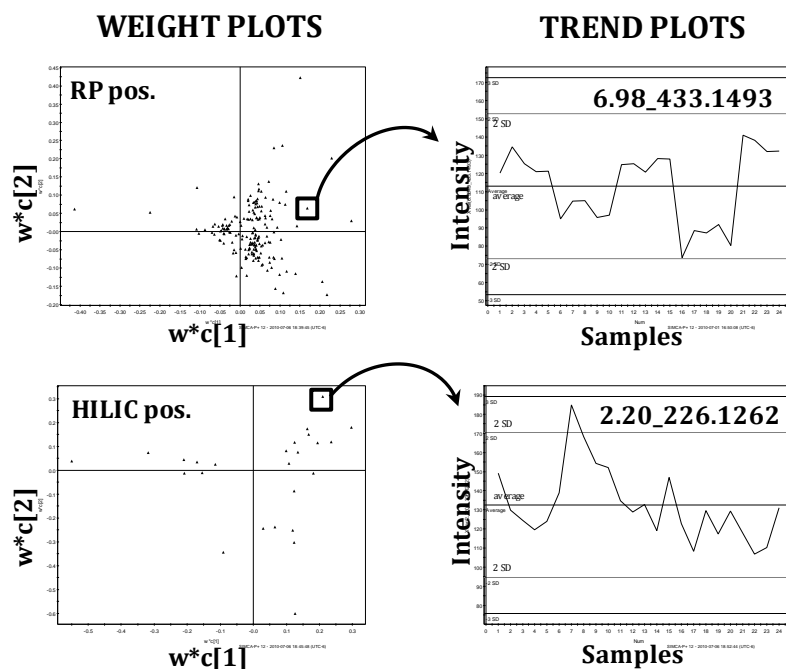
Figure 4.4: PLS weight plot (w^*c_1/w^*c_2) corresponding to the PLS-DA of commercial orange juices analyzed by RP pos. (upper left) and HILIC pos. (lower left) and two trend plots of selected markers of each data set. For clarity, sample classes are not displayed on the weight plot. The absence of a clear demarcation between the sample classes on the trend plot of HILIC pos. suggests the lack of specificity of the marker for one class in contrast with the marker from RP pos.

Our results suggested that only RP separation (SPE-eluate) provided reliable markers of the class differences for UHPLC-TOF-MS analyses; identifying these compounds could provide information on their origin and subsequently on the parameters to modify to enhance or decrease their formation if they are associated with desirable or undesirable flavor attributes, respectively. We would like to stress that this is a case-by-case result and that HILIC separation should not be dropped systematically in studies working with other food systems. We also want to note that if the final goal of the research is to predict flavor attributes from the chemical composition (which is the other application of flavoromics - explored in this thesis work), then all data need to be included to build the predictive

models. In this way, the approach is unbiased and potential contributors of flavor perception are not omitted.

Finally, we checked if using both ionization modes for RP analyses increased the number of markers considered for juice differentiation. Table 4.2 summarizes the main markers (RT-m/z) found to govern the classification of the orange juices. For the negative mode, all markers were detected before 6 min whereas the positive mode led to the detection of several markers after 6.5 min. This is in agreement with observations made earlier (Chapter 3), where the positive mode allowed the detection of late peaks while their signal was lower or absent in the negative mode. Also, we noticed, for the negative ionization mode, that some of the markers (denoted with a zigzag border in Table 4.2) were from the same compound. Indeed, after observation of their associated mass spectra it appeared that the features selected as two different markers were in fact dimers. Currently, we do not have a way (other than manual removal) to filter out of the data redundant signals belonging to the same molecule. Finally, markers selected from the RP neg. and RP pos. data sets appeared to be unique to each data set, as indicated by the different retention times and fragments (m/z). Only few markers had similar RT (i.e., 3.81, 4.12 and 4.96 min) but displayed different fragments; this can suggest two things: either the ionization mode led to differences in fragmentation for the same compound (e.g., glycoside loss) (Gattuso et al. 2007) or simply it confirmed the benefit of using both ionization modes in increasing compounds detection over single-ionization mode. Consequently, we decided to keep both ionization modes for the remainder of the study as it increased the number of potential markers. A similar observation was made by Cubbon et al. (2007) who found that using both ionization modes revealed different compounds that could be used for the discrimination of urine samples.

Table 4.2: Markers (features that contributed the most to the observed clusters) selected from RP analyses with ESI in negative and positive ionization mode of commercial orange juices.

RP neg.				RP pos.			
RT (min)	m/z	Tentative ID (CAS #)[a]	Δppm	RT (min)	m/z	Tentative ID (CAS #)[a]	Δppm
2.55	293.1266			2.36	265.1509		
3.81	497.3407			3.81	453.3438		
3.82	649.2575 [b]			4.12	273.0758	Naringenin (480-41-1)	0
3.82	1299.5439 [b]			4.95	515.2283		
4.12	579.1797	Naringin (10236-47-2)	13	6.70	373.1277	Tangeritin (481-53-8)	1
4.42	609.1907			6.88	403.1382	Nobiletin (478-01-3)	1
4.73	723.5129			6.98	433.1493		
4.96	693.2859 [b]			7.01	795.4836		
4.96	1387.605 [b]			7.03	373.1283		
5.96	517.2354						

[a] Accurate mass search performed online using the Scripps center Metlin database, no MS/MS experimentation was performed in this work; [b] Zigzag borders indicate markers derived from the same compound (i.e., dimer).

4.3.3. Fusion of volatiles and non-volatiles data

Since GC and LC analyses are partly complementary in terms of compounds analyzed, the merging of the data sets can be valuable and can provide a better understanding of the system investigated; for instance, reveal possible associations between volatiles and non-volatiles. We evaluated if combining data from the analysis of volatiles and non-volatiles improved sample differentiation. Similar approaches have been suggested by Smilde et al. (2005) and Tikunov et al. (2010). Hence, a low-level data fusion (i.e., no prior variable selection) was performed: GC-TOF-MS and UHPLC-TOF-MS normalized and reduced data sets were fused in a unique matrix in such a way that samples were in the shared mode (Smilde et al. 2005). The fusion of the MS-data resulted in a matrix of 25 samples (rows) × 2480 features (columns).
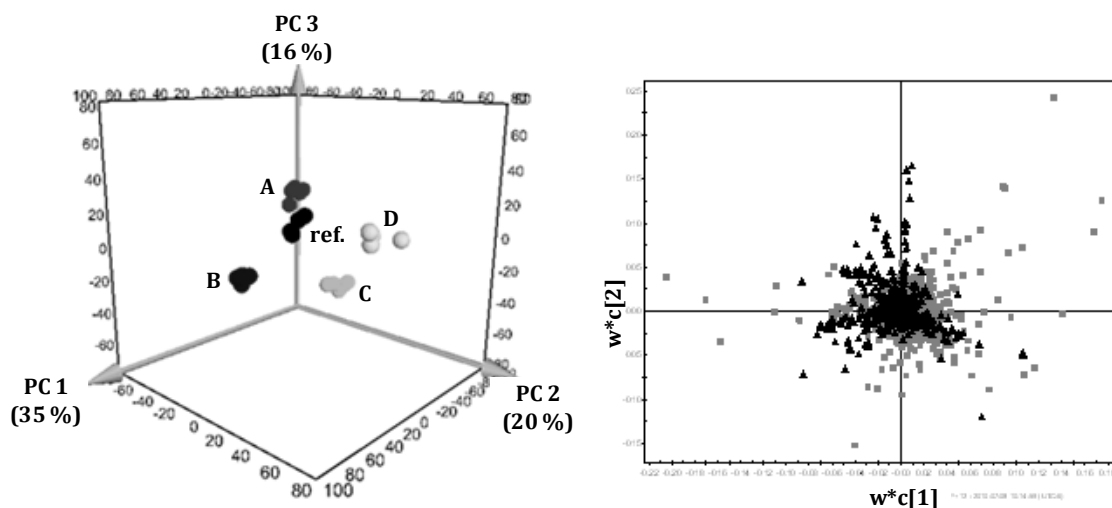
Figure 4.5: PLS-DA 3D score plot (upper) and its corresponding weight plot (lower) showing the class separation between four commercial premium orange juices and their blend. PLS-DA model was obtained from the fusion of the GC-TOF-MS (volatiles) and RP-UHPLC-TOF-MS (non-volatiles) data sets. Black triangles and gray boxes in the weight plot represented the features (RT-m/z) from the volatiles and non-volatiles data sets, respectively. For clarity, sample classes are not displayed on the weight plot.

Similarly to other data sets, a PCA was performed on the combined data set which gave a robust three-components model with $R^2X=0.71$ and $Q^2cum=0.61$. The clustering observed (not presented) was similar to that observed using volatiles or non-volatiles only: samples A and C were opposed to B and D along PC 1 (35 %) and, A and B opposed to C and D along PC 2 (20 %). This was expected since the classification obtained from GC-TOF-MS and RP-UHPLC-TOF-MS data sets were similar, even if the orientation of the PC axes was different. Of note, the contribution of PC 3 (16 %) was greater in the model obtained from the fused data sets than for the individual models obtained from volatiles (12 %) and non-volatiles (8 and 11 % for RP neg. and RP pos., respectively). The grouping and quality of the model given by PLS-DA modeling (Table 4.1) was also fairly similar to that of using only data for volatiles or non-volatiles; the four commercial juices and their blend were clearly separated from each other and their replicates were tightly clustered (Figure 4.5). The PLS weight plot

will be examined with more attention for any potential relationships between volatiles and non-volatiles. Correlation between signals from GC and LC can indicate that the compounds are chemically related, for instance they are formed from the same precursor or they share the same chemical pathway (Smilde et al. 2005; Tikunov et al. 2010). Also, the relation between volatiles and non-volatiles to flavor attributes will be explored in Chapter 6 using PLS regression (Eriksson et al. 2006e). The correlations established between chemical compounds and flavor attributes will require sensory validation (not conducted in this thesis research) in order to ensure the causal nature of the relationship.

4.4. CONCLUSIONS

In summary, this chapter reports on the evaluation of instrumental methods for the untargeted analysis of compounds in orange juice. The MS-information collected was adequate to pick out chemical differences in orange juices using multivariate tools like PCA and PLS-DA, supporting the proposal that the instrumental methods developed are suitable for the untargeted analysis of chemical stimuli of orange juice flavor.

The obtained classification models allowed the categorization of the samples by brands with little overlapping, and the tight clustering of the replicates indicated a good repeatability of the methods, especially for GC and RP-UHPLC. Despite the robustness of the models obtained from HILIC data, the markers from the SPE-wash analysis were not sufficiently specific to one class of samples. The fusion of GC- and RP-UHPLC-MS data sets gave similar classification models compared to that of using only data from volatiles or non-volatiles but can offer the advantage of finding potential correlations between chemical compounds and increased accuracy in flavor predictions as it includes inputs from more compounds.

Given that the developed instrumental methods were verified for their comprehensiveness and ability to discriminate between different orange juices, the research strategy (flavoromics) was applied to the study of a larger set of juices offering different flavor qualities (Chapter 5 and 6). The relationships between chemical compounds

(given as features RT-m/z) and flavor attributes (provided by means of descriptive sensory analysis) were investigated by means of partial least squares regression and used to predict the intensities of specific flavor attributes.

Chapter 5 : Sensory analysis of mandarin and hybrids juices

Note: The sensory analysis described in this chapter was not conducted at the University of Minnesota. The data were collected at the University of Florida Citrus Research and Education Center and at the USDA/ARS, Citrus and Subtropical Products Laboratory (Winter Haven, FL) as part of a three-year sensory study. Only part of the data collected by the study aforementioned was used for this thesis and is presented below.

Abstract. The descriptive sensory analysis of 61 mandarin juices made from commercial mandarin cultivars and new hybrids developed by the University of Florida Citrus Research and Education Center breeding program was obtained by a trained panel over three consecutive harvest seasons. A set of 19 descriptors including aroma (orthonasal perception), flavor (retronasal perception after swallowing sample) and basic taste descriptors was developed. Multivariate statistical tools such as principal component analysis (PCA) and hierarchical cluster analysis on principal components (HCPC) were used to reveal the differences/similarities in flavor profiles and examine the influence of the parentage on the sensory quality. Overall, there was a relatively good consensus between the sensory space collected on different years. The sensory space was divided (in the first two dimensions of the PCA) in four groups of descriptors: ["Sweet", "Fruity non-citrus" and "Pumpkin/fatty"], ["Sulfury", "Tangerine" and "Bitter"], ["Sour", "Green/fresh" and "Grapefruit"] and ["Orange", "Floral"]. Aroma and flavor descriptors corresponding to the same term were very well associated. The flavor profiles were very different between and within hybrids and cultivars, and juices made from fruits with common genetic background (parents and hybrids) tended to share some sensory characteristics. Differences in flavor profiles between harvesting seasons were noticed for some fruits; there were most likely due to differences in maturity and/or seasonal effects.

5.1. INTRODUCTION

Mandarins (tangerines) (*Citrus reticulata*) represent the second world's largest citrus crop after sweet oranges (*Citrus sinensis*). Over the last years, their market share has been growing continuously due to an increase in demand from EU and Russia (USDA Foreign Agricultural Service 2008; USDA Foreign Agricultural Service 2011).

Producing fruits with highly desirable sensory qualities (e.g. flavor) is one of the goals of citrus breeding programs. Traditionally, the development of fruits with improved sensory characteristics passes through crossing and selection which are both highly time-consuming and expensive. Though, more efficient approaches are now available to breeders; for instance, via marker-assisted selection (MAS) (Gmitter et al. 2007). This method allows to select fruits having the desired characteristics at earlier stages, as well as to discard fruits having poor characteristics.

Thus, finding molecular markers associated with specific sensory quality attributes (e.g., flavor) will enable a faster progress in breeding fruits with enhanced flavor characteristics. The description of the flavor profile of mandarin cultivars and their hybrids plays a key role in helping the breeder to select fruits with improved sensory characteristics, and assists in finding DNA-based molecular markers. As one would expect, the sensory quality of the fruits is somewhat related to their genetic background. For example, relationships between volatile composition of tangerine hybrids (analyzed by GC-MS) and their genetic background by comparison of inter-varietal similarities/differences in aroma compounds have been reported (Kerbiriou et al. 2007a; Miyazaki et al. 2011).

Sensory analysis, and more particularly descriptive sensory analysis, has been long used by breeding programs to characterize the flavor of newly developed hybrids. However, few sensory studies have been published. Sensory analysis has been used to determine the optimum harvesting maturity of the new hybrids by collecting the fruits at different times throughout the season (Kerbiriou et al. 2007b). Researchers were able to follow the evolution of the sensory profile of the fruits as they were maturing, for example, decreasing in sourness and increasing sweetness and fruity notes

Flavoromics may be a useful approach to aid in plant breeding programs if it can successfully predict the sensory quality of the product from instrumental data. At this time, researchers utilize expert sensory panels to characterize the flavor of the fruit and juices. However, panels are expensive and time-consuming and it is not possible to use them extensively over a long time period or for large sample sets. An instrumental methodology would definitely accelerate the sensory quality characterization of newly developed hybrids, and consequently, help the breeder in screening more mandarin hybrids. To develop the predictive models, both instrumental and sensory data must be initially collected on a common set of samples with the ultimate goal of establishing a quantitative relationship between the chemical composition of mandarin juices and their sensory attributes. Chapters 3 and 4 introduced the methodology and provided information on the instrumental characterization of orange juice samples; the present chapter describes the sensory analysis of another group of citrus, mandarins.

The objective of the present study was to evaluate the flavor of fresh mandarin cultivars and new hybrids by means of descriptive sensory analysis and Principal Component Analysis (PCA). In order to compare fruits harvested at different times, mandarins were juiced and frozen. A set of 61 juices with different flavor profiles was characterized over a three-year period. The data presented herein were sourced from a larger study of the University of Florida citrus breeding program (Plotto et al. 2010). The samples evaluated covered a large range of sensory properties, ranging from pumpkin and fatty flavor profiles to very sweet and floral flavors, as well as sour and grapefruit notes.

5.2. MATERIALS AND METHODS

Mandarin fruits. The mandarin cultivars and hybrids evaluated in this study were from the University of Florida Citrus Research and Education Center breeding program. They were selected by the breeder to represent a large diversity in terms of sensory properties. The growing conditions and the fruits have been described elsewhere (Kerbiriou et al. 2007a; Miyazaki et al. 2011). Briefly, mandarin trees were grown under similar conditions of soil, irrigation and illumination, and fruits were harvested for three consecutive years

(2006-2009) between the months of November and April depending on fruit maturity (as determined by the breeder and his assistants). Fifteen samples from fruits harvested in 2006-2007, 25 samples from fruits harvested in 2007-2008 and 21 samples from fruits harvested in 2008-2009 were selected. Table 5.1 summarizes the cultivars and hybrids used for the study and Figure 5.1 gives information on the pedigree of the fruit samples.
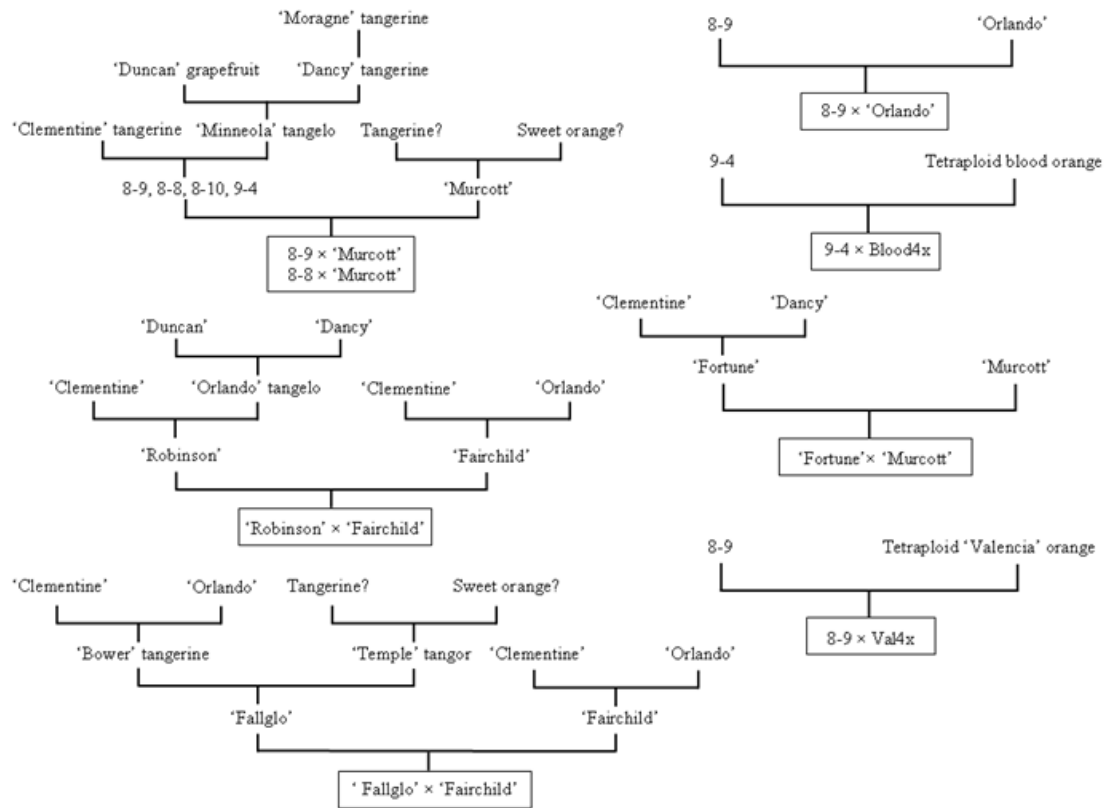
Figure 5.1: Overview of the mandarin parentage (Copyright 2011 Wiley. Used with permission from Miyazaki et al. (2011)).

Table 5.1: Mandarin cultivars and hybrids used for the study.

Cultivar or Hybrid	Coded tree [a]	Harvest date 2006-2007	2007-2008	2008-2009
8-8		1/26/2007		1/26/2009
8-9 × Murcott	a	11/21/2006	11/16/2007	
8-9 × Murcott	b	1/5/2007	11/16/2007	
8-9 × Murcott	c		11/16/2007	12/18/2008
8-9 × Murcott	d			1/26/2009
8-9 × Murcott	e		3/26/2008	
8-9 × Orlando			1/11/2008	1/26/2009
8-9 × VAL4x		3/23/2007	3/26/2008	4/2/2009
9-4 × Blood4x			1/11/2008	
Anna				1/26/2009
Clementine × Minneola LB8-10			12/14/2007	12/18/2008
Clementine × Minneola LB8-9		12/14/2006	12/14/2007	1/26/2009
Fallglo			11/16/2007	12/4/2008
Fallglo × Fairchild	a	9/11/2006	12/14/2007	12/18/2008
Fallglo × Fairchild	b		12/14/2007	
Fallglo × Fairchild	c		12/14/2007	
Fortune × Murcott	a	1/5/2007		1/26/2009
Fortune × Murcott	b		2/14/2008	
Kinnow		2/23/2007		1/26/2009
LS-Murcott [b]		2/9/2007	2/14/2008	2/24/2009
Minneola		12/22/2006		1/26/2009
Ortanique		3/23/2007	3/26/2008	4/2/2009
Robinson × Fairchild	a	11/21/2006	12/14/2007	12/4/2008
Robinson × Fairchild	b		11/16/2007	
Robinson × Fairchild	c		11/16/2007	12/18/2008
Robinson × Fairchild	d		12/14/2007	
Sanguinelli [c]		2/9/2007	2/14/2008	2/24/2009
Temple		2/9/2007	2/14/2008	2/24/2009
Unknown parent 1 [d]		12/14/2006	1/11/2008	12/18/2008
Unknown parent 2 [d]			1/11/2008	12/18/2008
Unknown parent 3 [d]			1/11/2008	

[a] For some crosses, samples were harvested from different progenies (i.e., different trees); [b] Murcott mutant with low seed (LS) number; [c] Sanguinelli sample harvested on 2/9/2007 was not analyzed by

instruments in Chapter 6 (lost sample); [d] Hybrid with unknown parentage but kept in the study for its interesting flavor.

Sample preparation. All samples were prepared on the day of harvest at the USDA/ARS Citrus and Subtropical Products Laboratory (for more details see Miyazaki et al. 2011). Briefly, each sample was prepared from 30 to 70 fruits collected from the same tree. Prior to juicing, fruits were washed and sanitized. The juices were obtained by hand-squeezing with an electric juicer (Oster Model 3183, Rye, NY) for 3 sec (to limit the amount of peel oil in the juice). Juices were kept in glass containers at -20 °C until sensory evaluation. On the day of the analysis, juices were thawed at room temperature and prepared in 3-digit coded 120 mL cups with lids (SOLO® Cup Company, Urbana, IL). Samples were presented in 45 mL portions and served at 14°C ±2.

Descriptive sensory analysis. The quantitative descriptive sensory analysis of the samples was performed at the USDA/ARS Citrus and Subtropical Products Laboratory by 13-15 panelists familiar with citrus juices. Panelists were trained according to standard guidelines to describe the aroma (odor) and flavor (taste) of mandarin juices (Meilgaard et al. 2007b). Upon training, a list of descriptors and their associated definitions and reference standards (Table 5.2) were established as well as the procedure to evaluate the samples. Some descriptors were implemented (or discarded) over the course of the three-year study as panelists were becoming more accurate in their sample characterization. For coding purposes, the descriptors corresponding to the aroma of the samples are followed by the letter "A" and the descriptors referring to their flavor are followed by the letter "F".

A maximum of four (2007-2008, 2008-2009) or five (2006-2007) samples were presented at one sitting, and the sensory evaluation lasted over one month each year of the study. In 2007-2008 and 2008-2009, samples were evaluated in duplicate. Sample presentation was monadic and randomized (Williams design); the randomization pattern was only for the samples presented at one sitting, not for all samples in the study to avoid thawing and freezing of samples multiple times. The judges evaluated the samples in booths under red lighting and were asked to rate the intensity of the descriptors using a 16-point category scale (1=low, 7-8=medium, 15=high); reference standards (Table 5.2), fresh water

and crackers were provided at each session. The odor descriptors were rated before drinking the juice, and then the taste descriptors were evaluated (sample in the mouth). During initial training, it was noted that some of the samples had a strong sulfury odor upon opening the lids, which dissipated quickly. The following procedure to rate samples was therefore adopted: the juice was swirled 3 times, lid open for 3 sec, then closed and the juice swirled one more time, then the evaluation began.

Panel and panelist performances. The panel and individual panelist's performances were evaluated for each attribute by examining the ability of the panel and each panelist to discriminate the juices, and the ability of each panelist to be in agreement with the rest of the panel and to be repeatable. The free SensomineR package (Lê and Husson 2008), available at http://sensominer.free.fr/index.html, was used in the R environment to perform these analyses; the procedures done are summarized in Table 5.3. For each panelist, the number of descriptors for which there was no discrimination ability, no consensus with the rest of the panel, and no repeatability was counted. Panelists with the highest count were eliminated from the study. (Note: the individual panelist's performances were not evaluated for the 2006-2007 sensory data as there was no replicate per sample in the data set).

Selection of descriptors. After selection of the panelists, the descriptors to describe the sensory profile of mandarin juices were chosen. First, descriptors that were not common to all three years were discarded so that data sets can be later merged. Second, the descriptors that had similar definitions but were designated differently (between studies) were assigned a common term for purpose of simplicity (see Table 5.2). Finally, the statistical significance of the descriptors to differentiate juices was tested using a two-way ANOVA (Product, Panelist). Only descriptors that were different between the juice samples (P<0.2) were selected for the Principal Component Analyses.

Table 5.2: Sensory attributes and their associated definition and reference standards used for the descriptive sensory analysis.

	2006-07	2007-08	2008-09	Definition	Reference standard
Aroma (odor)	Tangerine	Tangerine	Tangerine	Odor associated with tangerines	56 ppm tangerine oil in sugar/acid mix [a]
	Orange	Orange	Orange	Odor associated with oranges	Fresh orange juice (frozen and thawed)
	Grapefruit	Grapefruit	Grapefruit	Odor associated with grapefruits	Fresh grapefruit juice (frozen and thawed)
	Fruity non-citrus	Fruity non-citrus	Fruity non-citrus	Odor associated with a mixture of various fruits / sweet aromatics: mango, passion fruit, guava, apples, pineapple, berries	43% passion fruit juice (Welch's) , 21% mango juice (Looza), 12% guava nectar (Petit), 17% peach nectar (Sunchy), 7% pineapple juice (Dole)
	Floral	Floral	Floral	Odor associated with flowers	10 ppm linalool in sugar/acid mix [a]
	Green	Fresh	Green/ Fresh	Odor associated with grass, lemon/lime, freshly picked fruit	1.9 ppm *cis*-3-hexenal + 7.0 ppm *cis*-3-hexenol + 15 ppm lime oil in sugar/acid mix [a]
	Sulfury	Sulfury	Sulfury	Odor associated with rotten eggs, burnt tires, cooked cabbage	0.3 ppm dimethyl sulfide in reconstituted orange juice concentrate (12° Brix)
	Other	Pumpkin/ fatty	Pumpkin	Odor associated with cooked or fresh vegetables/ oil	Tropical pumpkin puree (enzymatic activity stooped)
			Fatty	Odor associated with oil	3.5 ppm *cis*-3-hexenal in water
		Cooked	Cooked	Odor associated with cooked or overripe oranges	Commercial pasteurized tangerine juice (Noble)
		Woody/ spicy		Odor associated with cinnamon, pepper, bark, pine	

Flavor (taste)

Tangerine	Tangerine	Tangerine	Flavor associated with tangerines	Same as aroma
Orange	Orange	Orange	Flavor associated with oranges	Same as aroma
Grapefruit	Grapefruit	Grapefruit	Flavor associated with grapefruits	Same as aroma
Fruity non-citrus	Fruity non-citrus	Fruity non-citrus	Flavor associated with a mixture of various fruits: mango, passion fruit, guava, apples, pineapple, berries	Same as aroma
Floral	Floral	Floral	Flavor associated with flowers	Same as aroma
Green	Fresh	Green/ Fresh	Flavor associated with grass, lemon/lime, freshly picked fruit	Same as aroma
Aftertaste	Sulfury	Sulfury	Lingering flavor in the mouth after you swallow	Same as aroma
Other	Pumpkin/ fatty	Pumpkin	Flavor associated with cooked or fresh vegetables, oil	Same as aroma
		Fatty	Flavor associated with oil	Same as aroma
	Cooked	Cooked	Flavor associated with cooked or overripe oranges	Same as aroma
	Woody/ spicy		Flavor associated with cinnamon, pepper, bark, pine	Same as aroma
Sweet	Sweet	Sweet	Taste associated with sucrose solutions	8 % sucrose solution
Sour	Sour	Sour	Taste associated with citric acid solutions	0.25 % citric acid solution
Bitter	Bitter	Bitter	Taste associated with caffeine	0.1 % caffeine solution

[a] Sugar/acid mix (base for tangerine, floral and green/fresh standards): 8% sucrose and 0.2% citric acid.

Table 5.3: Overview of the methodology used to assess panel and panelist performances.

Ability evaluated	ANOVA model and criterion for decision making [a]
Panel performance (evaluated for each descriptor)	
	Descriptor score ~ Panelist + Product + Session + Panelist × Session + Product × Session + Panelist × Product (panelist as random effect)
Discrimination	Examine the significance of the product effect (P<0.05). A significant effect means that panelists can discriminate between products.
Agreement	Examine the significance of the Panelist × Product effect (P<0.05). A significant effect means that there is no consensus among the panel to evaluate each product.
Repeatability	Examine the significance of the Session, Panelist × Session and Product × Session effect (P<0.05). A significant effect means that, between sessions, one or more panelists are not repeatable in their scores and products are judged differently.
Panelist performance (evaluated for each descriptor)	
	Descriptor score ~ Product + Session
Discrimination	Examine the significance of the product effect (P<0.05). A significant effect means that the panelist can discriminate between products.
Agreement	Examine the correlation coefficients between the panel's means and the panelists'ones (R^2>0.6, agrees with the panel).
Repeatability	Examine the interaction Product × Session effect through the standard deviation of the ANOVA model residuals for each judge.

[a] Based on and adapted from SensoMineR's website.

   Juice sensory characterization. After selection of the panelists and sensory descriptors, the sensory data for each year (2006-2007, 2007-2008, 2008-2009) was organized in a matrix of I rows (observations or juice samples) and J columns (variables or sensory descriptors) and their corresponding mean sensory score. The mean score of each descriptor for each juice was obtained by averaging the scores across panelists. The sensory data was then analyzed using Principal Component Analysis on each set of data (2006-2007, 2007-2008, 2008-2009). The variables (descriptors) were mean-centered and scaled to unit

variance (each column is multiplied by 1/SD, where SD is the standard deviation of the variable), no adjustment was performed for the individuals (panelists). Next, an agglomerative hierarchical clustering analysis (HCPC) was performed on the results from the PCA to better visualize the similarities between samples; the optimal number of clusters was chosen automatically between 4 and 10 (Ward's method). The free FactoMineR package (Lê et al. 2008), available at http://factominer.free.fr/, was used in the R environment to perform the analyses.

5.3. RESULTS AND DISCUSSION

The objective of the study was to obtain the sensory profile of 61 mandarin and hybrids juices. These data were later correlated with instrumental data (Chapter 6) in order to develop predictive models of the sensory quality of mandarin juices. The sensory data were collected over a three-year period and examined using principal component analysis and hierarchical clustering on principal components.

5.3.1. Panel and Panelist's performance

The performance of the panel and the panelists in achieving the sensory analysis of the juices were evaluated using ANOVA models. A panel and a panelist were considered efficient if they were able to discriminate between products, to be in agreement with the other panelists and to be repeatable in their judgments.

5.3.1.1. Sensory data 2008-2009

Twenty-one mandarin juices were evaluated by 13 trained panelists for 23 sensory descriptors (10 aroma descriptors, 10 flavor descriptors and 3 basic taste descriptors).

Panel performance. A three-way ANOVA model was used to evaluate panel performance, the results are summarized in Table 5.4. All of the descriptors had a significant Product

effect which indicated that the juices had different sensory profiles. There was no consensus in the panel to evaluate each product for all descriptors except for "SulfuryA"; the most significant disagreement in the panel was for the descriptor "Green/freshF". The Panelist × Product interaction was significant because panelists disagreed in the ratings of the products for this descriptor; this was confirmed by looking at the scores given by each panelist across samples (graphic not shown). Regarding the repeatability of the panel, there was no significant effect for Session and Product × Session (except for "FattyA") indicating that overall the products were judged similarly between sessions. However, there was a significant effect of Panelist × Session for most of the descriptors indicating that one or more panelists were not repeatable between sessions (which will be discussed when assessing the individual panelist's performance).

Panelists' performance. The ability of a panelist to discriminate products was evaluated by the number descriptors for which the panelist was able to detect significant differences between juices. Panelists 13, 2 and 10 were not good at discriminating juices for 19, 18 and 16 descriptors out of a total of 23, respectively. Examination of the coefficient of correlations between the panel's mean scores for the descriptors and the panelist's confirmed the lack of agreement in the panel, except for the descriptors "Sour", "PumpkinF" and "Bitter" for which 70 to 100 % consensus was obtained. Finally, the repeatability of the panelists between sessions was examined. Panelists 3 and 10 were the least repeatable whereas 11, 9, 13, 4 and 5 were very consistent in their scores; nonetheless panelist 9, 13, and 5 were poor discriminators.

Conclusions for 2008-2009 panel. Table 5.5 summarizes, for each panelist, the number of descriptors for which there was no discrimination ability, no consensus with the rest of the panel, and no repeatability. After reviewing the individual performances of the panelists, it was decided to exclude scores given by panelists 3, 10 and 2. These panelists were not good at discriminating products, conflicted with the panel results, and were poorly repeatable.

116

Table 5.4: Summary of the ANOVA results for the panel performance assessment from the three sensory data sets (P-values are shown).

	2008-2009						2007-2008						2006-2007	
Descriptor	Prod.	Session	Product × Panelist	Product × Session	Panelist × Session	Descriptor	Prod.	Session	Product × Panelist	Product × Session	Panelist × Session	Descriptor	Prod.
TangerineA	**5E-11**	9E-01	**8E-05**	5E-01	**1E-02**	TangerineA	**1E-09**	3E-01	**2E-03**	**3E-02**	**2E-03**	TangerineA	**9E-05**
OrangeA	**3E-25**	2E-01	**8E-07**	6E-01	3E-01	OrangeA	**7E-16**	6E-01	**9E-07**	1E-01	2E-01	OrangeA	**5E-03**
GrapefruitA	**4E-10**	6E-01	**6E-03**	6E-01	**2E-03**	GrapefruitA	**2E-12**	2E-01	**3E-07**	1E-01	**1E-05**	GrapefruitA	**7E-07**
FruityA	**3E-04**	9E-01	**3E-02**	1E+00	5E-01	FruityA	**3E-02**	7E-01	**3E-03**	8E-01	**2E-02**	FruityA	**3E-02**
FloralA	**8E-05**	9E-01	**6E-03**	8E-01	9E-01	FloralA	**9E-04**	7E-01	**2E-03**	7E-01	**5E-03**	FloralA	5E-01
Green/freshA	**1E-04**	3E-01	**6E-03**	4E-01	6E-02	Green/freshA	**9E-03**	8E-01	**4E-03**	8E-01	**5E-03**	Green/freshA	**8E-04**
SulfuryA	**7E-11**	8E-01	2E-01	9E-01	7E-02	SulfuryA	**1E-07**	6E-01	**7E-04**	5E-01	**5E-08**	SulfuryA	**2E-05**
PumpkinA	**4E-20**	3E-01	**2E-07**	6E-01	**4E-02**	Pumpkin/fattyA	**2E-16**	4E-01	**3E-02**	**5E-02**	**2E-02**	Pumpkin/fattyA	**2E-06**
FattyA	**4E-07**	4E-01	**4E-02**	**3E-02**	6E-02								
CookedA	**9E-03**	4E-01	**2E-02**	7E-01	**4E-04**	CookedA	**1E-05**	4E-01	2E-01	**3E-02**	7E-01		
						Woody/spicyA	**5E-03**	8E-01	**2E-02**	1E+00	5E-01		
TangerineF	**7E-17**	7E-01	**2E-08**	6E-02	**7E-03**	TangerineF	**2E-07**	6E-01	**1E-02**	3E-01	**5E-04**	TangerineF	**2E-08**
OrangeF	**3E-15**	6E-01	**5E-08**	1E-01	9E-02	OrangeF	**6E-05**	6E-02	**1E-09**	1E+00	**1E-02**	OrangeF	**2E-03**
GrapefruitF	**5E-17**	6E-01	**1E-04**	5E-01	**9E-03**	GrapefruitF	**6E-28**	9E-01	**2E-15**	**3E-02**	4E-01	GrapefruitF	**5E-14**
FruityF	**2E-13**	5E-02	**9E-03**	1E-01	1E-01	FruityF	**8E-19**	**5E-02**	**4E-06**	3E-01	5E-01	FruityF	**2E-11**
FloralF	**8E-04**	5E-01	**1E-04**	9E-01	**5E-02**	FloralF	**5E-02**	5E-01	**7E-10**	4E-01	**5E-05**	FloralF	2E-01
Green/freshF	**6E-06**	5E-01	**1E-14**	2E-01	**1E-05**	Green/freshF	**5E-10**	2E-01	**5E-06**	2E-01	**4E-02**	Green/freshF	**3E-03**

117

SulfuryF	**7E-06**	6E-01	**3E-07**	3E-01	**2E-02**	SulfuryF	**2E-04**	5E-01	**1E-07**	5E-01	**2E-05**	SulfuryF	**8E-03**
PumpkinF	**2E-25**	8E-01	**2E-07**	6E-01	**1E-02**	Pumpkin/fattyF	**4E-10**	**5E-02**	**9E-03**	5E-01	6E-01	Pumpkin/fattyF	**2E-08**
FattyF	**1E-03**	9E-01	**1E-03**	8E-01	**2E-03**								
CookedF	**3E-04**	7E-01	**4E-05**	4E-01	**5E-04**	CookedF	**2E-07**	1E-01	**6E-05**	2E-01	8E-01		
						Woody/spicyF	**1E-04**	2E-01	**3E-12**	1E-01	**2E-05**		
Sweet	**3E-27**	6E-01	**1E-02**	1E-01	**2E-05**	Sweet	**1E-54**	7E-02	**4E-07**	5E-01	**1E-02**	Sweet	**3E-11**
Sour	**3E-54**	7E-01	**3E-09**	2E-01	**5E-03**	Sour	**6E-69**	5E-01	**7E-06**	1E+00	9E-01	Sour	**1E-25**
Bitter	**4E-14**	2E-01	**2E-08**	8E-01	5E-01	Bitter	**1E-28**	5E-01	**5E-15**	1E-01	**2E-05**	Bitter	**7E-10**

Table 5.5: Summary of the panel and panelist's performances of 2008-2009 sensory data.

Panelist ID [a]	**Number of descriptors (out of 23) for which there is:**			
	No discrimination	No agreement with panel	No repeatability	Total
3	**11**	**18**	**18**	**47**
10	**16**	**17**	**14**	**47**
2	**18**	**20**	**7**	**45**
13	19	21	0	40
8	14	18	6	38
1	10	17	8	35
5	14	19	0	33
6	11	15	5	31
4	11	19	0	30
7	11	17	2	30
9	12	17	0	29
12	7	18	1	26
11	7	15	0	22

[a] Panelists excluded from the multivariate analyses are in bold.

5.3.1.2. Sensory data 2007-2008

Twenty five mandarin juices were evaluated by 15 trained panelists for 23 sensory descriptors (10 aroma descriptors, 10 flavor descriptors and 3 basic taste descriptors).

Panel performance. A three-way ANOVA model was used to evaluate panel performance, the results are summarized in Table 5.4. All of the descriptors had a significant Product effect which indicated that the juices had different sensory profiles. There was no consensus in the panel to evaluate each product for all descriptors except for "CookedA"; the most significant disparity in scoring for the panel was for the descriptor "Bitter". This is not surprising as the sensitivity to bitterness is known to be very different among the population: even though the panelists ranked the products in the same way (see panelist's agreement in next paragraph), they used different levels of the scale to judge the differences between products. Regarding the repeatability of the panel, there was no significant effect

119

for Session and Product $\times$ Session except for few descriptors indicating that overall the products were judged similarly between sessions. However, there was a significant effect of Panelist $\times$ Session for most of the descriptors indicating that one or more panelists were not repeatable between sessions.

Table 5.6: Summary of the panel and panelist's performances of 2007-2008 sensory data.

Panelist ID [a]	Number of descriptors (out of 23) for which there is:			
	No discrimination	No agreement with panel	No repeatability	Total
13	**15**	**16**	**14**	**45**
4	**17**	**17**	**6**	**40**
8	**17**	**19**	**3**	**39**
5	12	17	6	35
3	15	17	2	34
7	16	17	1	34
14	14	19	1	34
2	11	20	2	33
12	14	19	0	33
10	14	17	1	32
11	9	18	5	32
6	10	15	5	30
1	8	17	3	28
9	12	15	0	27
15	7	19	0	26

[a] Panelists excluded from the subsequent multivariate analyses are in bold.

   Panelists' performance. The ability of a panelist to discriminate products was evaluated by the number descriptors for which the panelist was able to pick up significant differences between juices. Panelists 4, 8 and 7 were not good at discriminating juices for 17, 17 and 16 descriptors out of a total of 23, respectively. Examination of the coefficients of correlation between the panel's mean scores for the descriptors and the panelist's scores confirmed the lack of agreement in the panel, except for the descriptors "Sour", "Sweet", "Bitter" and

"GrapefruitF" for which 60 to 100 % consensus was obtained. "SulfuryF" was the descriptor with the least agreement between panelists. Finally, the repeatability of the panelists between sessions was examined. Panelists 4, 5 and 13 were the least repeatable whereas 9, 12 and 15 were very consistent in their scores; nonetheless panelist 9, 12 were poor discriminators.

Conclusions for 2007-2008 panel. Table 5.6 summarizes, for each panelist, the number of descriptors for which there was no discrimination ability, no consensus with the rest of the panel, and no repeatability. After reviewing the individual performances of the panelists, it was decided to exclude scores given by panelists 13, 4 and 8. These panelists were not good at discriminating products, conflicted with the panel results, and were poorly repeatable.

5.3.1.3. Sensory data 2006-2007

Fifteen mandarin juices were evaluated by 13 trained panelists for 19 sensory descriptors (8 aroma descriptors, 8 flavor descriptors and 3 basic taste descriptors).

Panel performance. A two-way ANOVA model was used to evaluate panel performance (recalling that there were no duplicate samples), the results are summarized in Table 5.4. All of the descriptors had a significant Product effect, except for "FloralA" and "FloralF", indicating that the panel was able to discriminate between juices for most descriptors.

Panelists' performance. The individual panelist's performance could not be evaluated as there must be at least two replicates per sample in the data set.

Conclusions for 2006-2007 panel. Based on their performance in the following years, Panelists 4 (a.k.a. Panelist 2 for 2008-2009) and 8 (a.k.a. Panelist 8 for 2007-2008) were removed as they were not efficient judges. Besides, for the 2006-2007 data set, they were poor discriminators (they scored many products for many descriptors with a zero-score). Panelist 5, who did not participate in the following years, was also left-out from the analyses as he was also a poor discriminator.

5.3.2. Mandarin and hybrid juices sensory characterization

Principal Component Analysis (PCA) and Hierarchical Clustering on Principal Component (HCPC) were used to summarize the sensory data sets by looking at the similarities/differences between juice samples (referred as individuals in PCA terminology), the correlations between sensory descriptors (referred as variables in PCA terminology) and the relationships between groups of samples and descriptors.

5.3.2.1. Selection of the descriptors

In order to be able to pool together the sensory results from all three years in the sensory-instrumental analyses (Chapter 6), only common descriptors were retained. In addition, descriptors that had similar definitions but were designated differently (between studies) were assigned a common term for purposes of simplicity. In the 2006-2007 data set, the descriptors "Other" and "Aftertaste" were further identified by the panel as "PumpkinA" and "SulfuryF", respectively. In the 2008-2009 data set, the scores for descriptors "Pumpkin" and "Fatty" were averaged as "Pumpkin/fatty" since they had similar behavior across samples. The following 19 descriptors were finally selected for the remainder of the analyses: TangerineA, OrangeA, GrapefruitA, Fruity non-citrusA, FloralA, Green/freshA, SulfuryA, Pumpkin/fattyA, TangerineF, OrangeF, GrapefruitF, Fruity non-citrusF, FloralF, Green/freshF, SulfuryF, Pumpkin/fattyF, Sweet, Sour and Bitter.

5.3.2.2. Sensory data 2008-2009

The 2008-2009 sensory data set, which included all of the 19 descriptors, was analyzed by means of PCA followed by HCPC (Figure 5.2). The first two principal components (PC) accounted for about 63 % of the total variance of the data set (and 75 % with 3 PCs), indicating that the data are well summarized with a 2-D projection.

The PCA loadings plot was first examined to highlight any correlations between sensory descriptors. In the first two dimensions, the sensory profile space was divided roughly in four groups of descriptors: ["Sweet", "Fruity non-citrus" and "Pumpkin/fatty"], ["Sulfury", "Tangerine" and "Bitter"], ["Sour", "Green/fresh" and "Grapefruit"] and ["Orange", "Floral"].
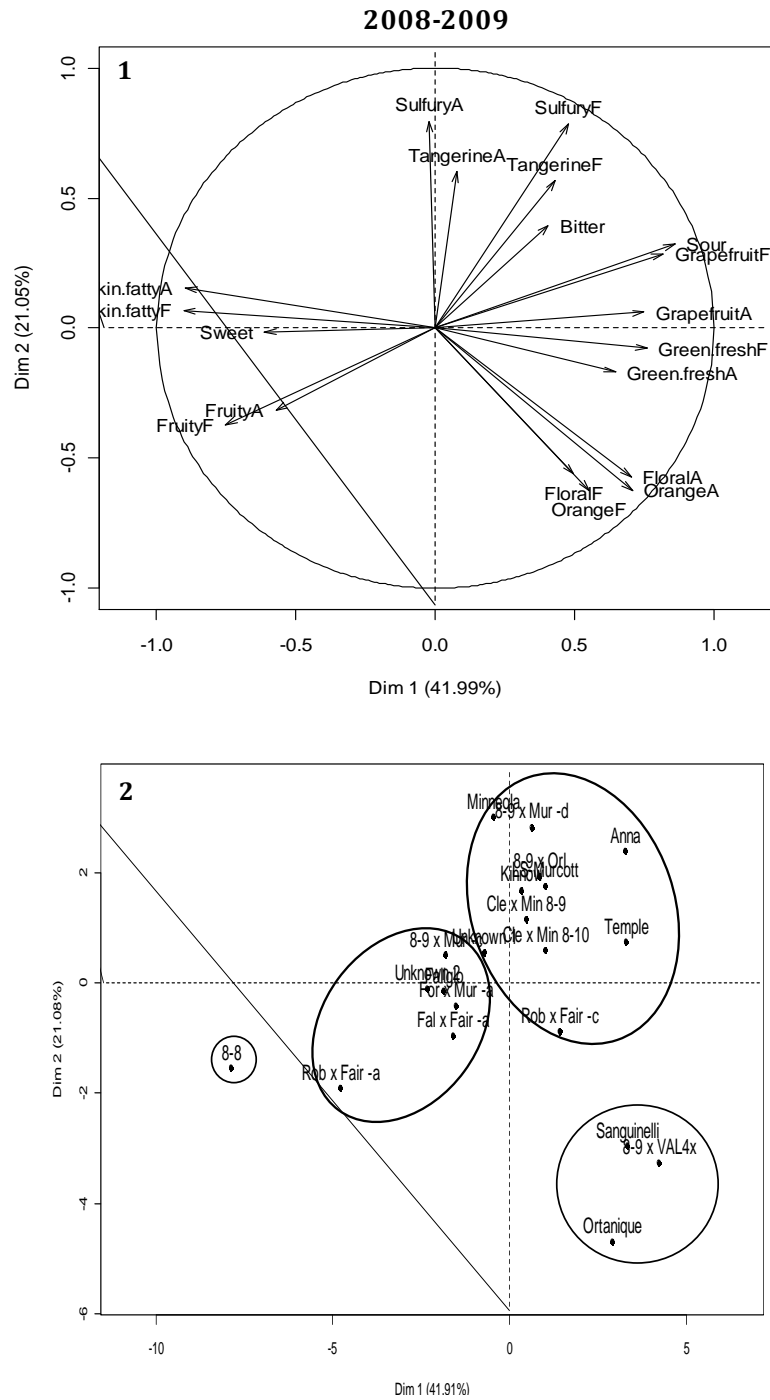
2008-2009

Figure 5.2: PCA score plot (1) and loadings plot (2) associated with the sensory analysis of the mandarin and hybrid juices evaluated in 2008-2009. The circles on the loadings plots represent the clusters determined by the HCPC.

The descriptors "Sweet", "Fruity non-citrus" and "Pumpkin/fatty" were negatively correlated with "Sour", "Green/fresh" and "Grapefruit" along the first dimension, which is not totally surprising as most of these attributes are related to the fruit maturity. Furthermore, panelists tended to associate "Grapefruit flavor" to sourness because of the reference standard (grapefruit juice) which was quite sour. The descriptors ["Sulfury", "Tangerine" and "Bitter"] were negatively correlated with ["Floral", "Orange"] along the second dimension and were not much correlated to the descriptors "Sweet", "Fruity non-citrus" and "Pumpkin/fatty". However, they had a positive correlation with "Sour", "Green/fresh" and "Grapefruit".

Although not represented on the loadings plot, the third dimension (12 % of the total variance) was significantly influenced by the descriptors "Tangerine", "Sweet", and "Bitter" which explains why these descriptors are not on the periphery of the correlation circle of the variables in the two first dimensions. As a side note, the aroma and flavor descriptors corresponding to the same term were very well associated between each other as seen by their proximity on the loadings plot.

The PCA score plot was examined to see how samples related to each other, recalling that samples being close to each other in the score plot have similar sensory profiles in those dimensions (PC 1 and PC 2). In addition, a hierarchical cluster analysis was performed on the results of the PCA to help us in categorizing the juices (Table 5.7). A total of four clusters were determined and are represented on the PCA score plot; samples with common genetic background (parents and hybrids) tended to be clustered together, indicating their similarities in sensory profiles. Figure 5.3 illustrates the main differences in aroma and flavor between the four clusters of mandarin and hybrid juices; for clarity, only one sample per cluster is represented.

Table 5.7: Sample clustering obtained by hierarchical clustering on principal component analyses (2008-2009, 2007-2008, 2006-2007 data) [a].

Cluster	2008-2009	2007-2008	2006-2007
1	8-8	Unknown1, Unknown2, Rob×Fair-a, 8-9×Orl	8-8
2	Fallglo, Fal×Fair-a, Rob×Fair-a, 8-9×Mur-c, For×Mur-a, Unknown2	Fal×Fair (b and c), Rob×Fair(c and d), 8-9×Mur-e, Unknown3	Rob×Fair-a
3	Kinnow, Anna, Temple, Minneola, Rob×Fair-c, Cle×Min (8-9 and 8-10), 8-9×Orl, 8-9×Mur-d, LS-Murcott, Unknown1	Fallglo, Rob×Fair-b, Cle×Min (8-9 and 8-10), 8-9×Mur (b and c), LS-Murcott, For×Mur-b	Unknown1, 8-9×Mur-b, For×Mur-a, Fal×Fair-a
4	Ortanique, Sanguinelli, 8-9×VAL4x	8-9×Mur-a, Fal×Fair-a, 9-4×Blood4x	Kinnow, Ortanique , Cle×Min8-9, 8-9×Mur-a
5		Temple, Ortanique, 8-9×VAL4x, Sanguinelli	LS-Murcott
6			Minneola, Temple
7			8-9×VAL4x
8			Sanguinelli

[a] Cle: Clementine, Fair: Fairchild, Fal: Fallglo, For:Fortune, Min: Minneola, Mur: Murcott, Orl: Orlando, Rob: Robinson.

Cluster 1 included only sample 8-8, which was strongly associated with a pumpkin/fatty flavor profile. Interestingly, this sample was selected by the breeder as one with extremely low citric acid content. This may suggest that there is a genetic relationship with low acid and pumpkin flavor, or that the lack of acids in the juice creates an imbalance in flavor perception. Cluster 2 included Fallglo, Fallglo × Fairchild tree a, Robinson × Fairchild tree a, 8-9 × Murcott tree c, Fortune × Murcott tree a, and hybrid of unknown lineage 2; this cluster had stronger fruity non-citrus and sweet notes. Cluster 3 included Kinnow, Anna, Temple, Minneola, Robinson × Fairchild tree c, two Clementine × Minneola hybrids (8-9 and 8-10), 8-9 × Orlando hybrid, 8-9 × Murcott tree d, LS-Murcott and hybrid of unknown lineage 1; these juices were described by more sulfury, tangerine, grapefruit, green/fresh and sour flavor profiles. It was found during training that some samples exhibited a stronger sulfury

aroma developed upon juicing and freezing. This may indicate the presence of sulfur compound precursors contained in the vacuoles that were liberated due to cell-breakage. Finally, cluster 4 included Ortanique, Sanguinelli and 8-9 × VAL4x hybrid; these juices had higher floral and orange characteristics (both aroma and flavor). These varieties and hybrids had orange (*Citrus sinensis* L. Osb.) in their background, and were found to have a dominance of ester volatiles (Kerbiriou et al. 2007a; Miyazaki et al. 2011), which are known to impart fruity and floral flavor to citrus juice (Shaw and Wilson 1980).

Figure 5.3: Sensory profile of the mandarin and hybrid juices for 2008-2009 (summarized by one juice sample per cluster).

5.3.2.3. Sensory data 2007-2008

The 2007-2008 sensory data set, which included all of the 19 descriptors, was described by means of PCA followed by HCPC; the PCA outputs are presented in Figure 5.4. The first two principal components (PC) accounted for about 61 % of the total variance of the data set (and 71 % with 3 PCs), indicating that the data is well summarized with a 2-D projection.

The sensory profile space was very similar to that of 2008-2009 and therefore, is not discussed here. The only small difference is for the descriptors ["Floral", "Orange"] which had a stronger influence on the second dimension and were negatively correlated to ["Sulfury", "Tangerine"]. Although not represented on the loadings plot, the third dimension (10 % of the total variance of the data set) was significantly influenced by the descriptors "Bitter", "FloralF", "OrangeF" and "TangerineF".

A total of five clusters (Table 5.7) were determined and are represented on the PCA score plot. Figure 5.5 illustrates the main differences in aroma and flavor between the different groups of mandarin and hybrid juices. Cluster 1 included the two hybrids of unknown lineage (1 and 2), Robinson × Fairchild tree a, and 8-9 × Orlando hybrid; these juices were perceived as very sweet with high fruity non-citrus and pumpkin/fatty flavor. Cluster 2 included two Fallglo × Fairchild hybrids (tree b and c), two Robinson × Fairchild hybrids (tree c and d), 8-9 × Murcott hybrid tree e, and one hybrid of unknown parent 3; this cluster had pumpkin/fatty flavor profile with tangerine and sulfury notes. Cluster 3 included Fallglo, Robinson × Fairchild tree b, two Clementine × Minneola hybrids (8-9 and 8-10), two 8-9 × Murcott hybrids (tree b and c), LS-Murcott and Fortune × Murcott tree b; the juices were described by more sulfury, tangerine and bitter flavor profiles. Cluster 4 included one 8-9 × Murcott hybrid (tree a), Fallglo × Fairchild tree a, and 9-4 × Blood4x hybrid; these juices had floral and orange characteristics, combined with sweet and fruity non-citrus attributes. Finally, cluster 5 included Temple, Ortanique, 8-9 × VAL4x hybrid and Sanguinelli; their profile was to some extent similar to cluster 4 but less sweet, more sour and, with additional green/fresh and grapefruit notes.

127

2007-2008

Figure 5.4: PCA score plot (1) and loadings plot (2) associated with the sensory analysis of the mandarin and hybrid juices evaluated in 2007-2008. The circles on the loadings plots represent the clusters determined by the HCPC.
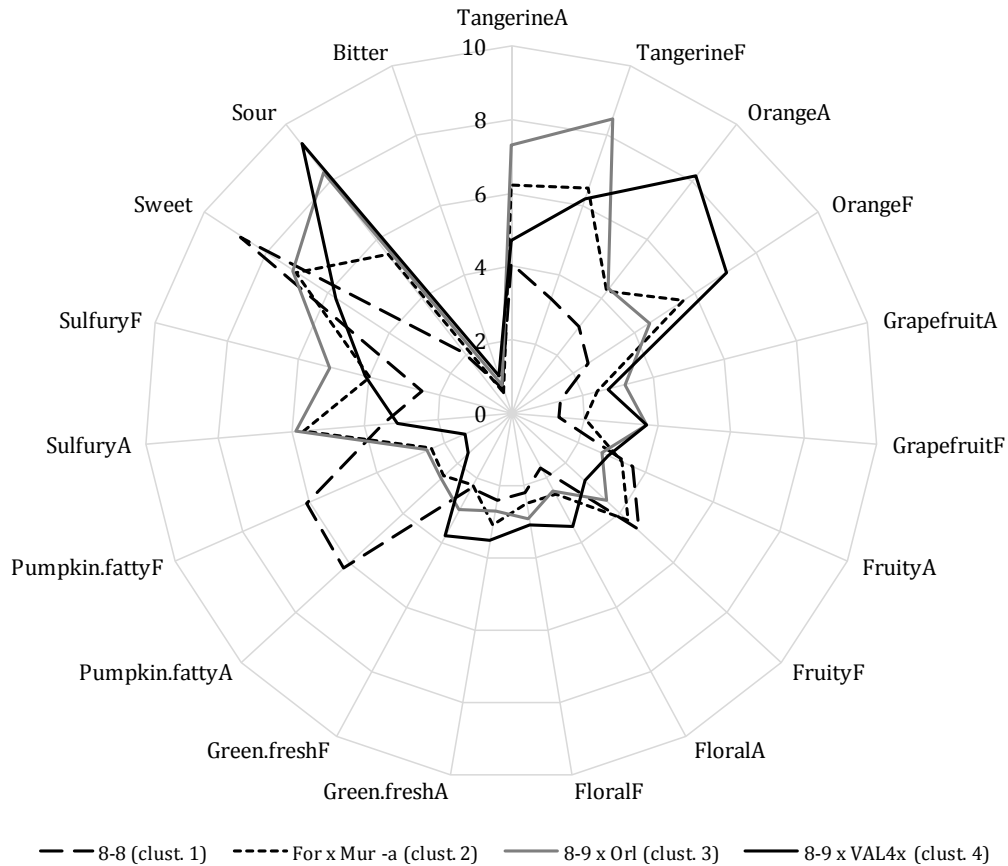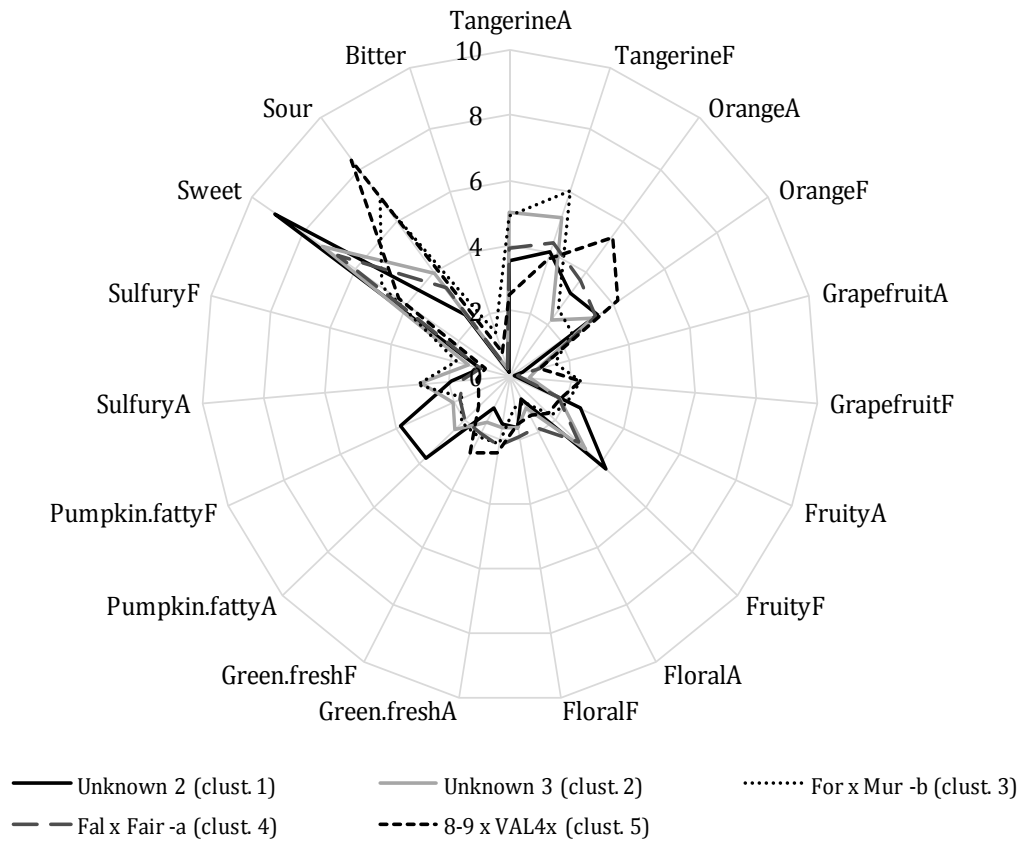
Figure 5.5: Sensory profile of the mandarin and hybrid juices for 2007-2008 (summarized by one juice sample per cluster).

5.3.2.4. Sensory data 2006-2007

The 2006-2007 sensory data set, which did not included the descriptors "FloralA" and "FloralF", was described by means of PCA followed by HCPC; the PCA outputs are presented in Figure 5.6. The first two principal components (PC) accounted for about 57 % of the total variance of the data set and 70 % with 3 PCs, indicating that the data are better summarized with a 3-D projection.
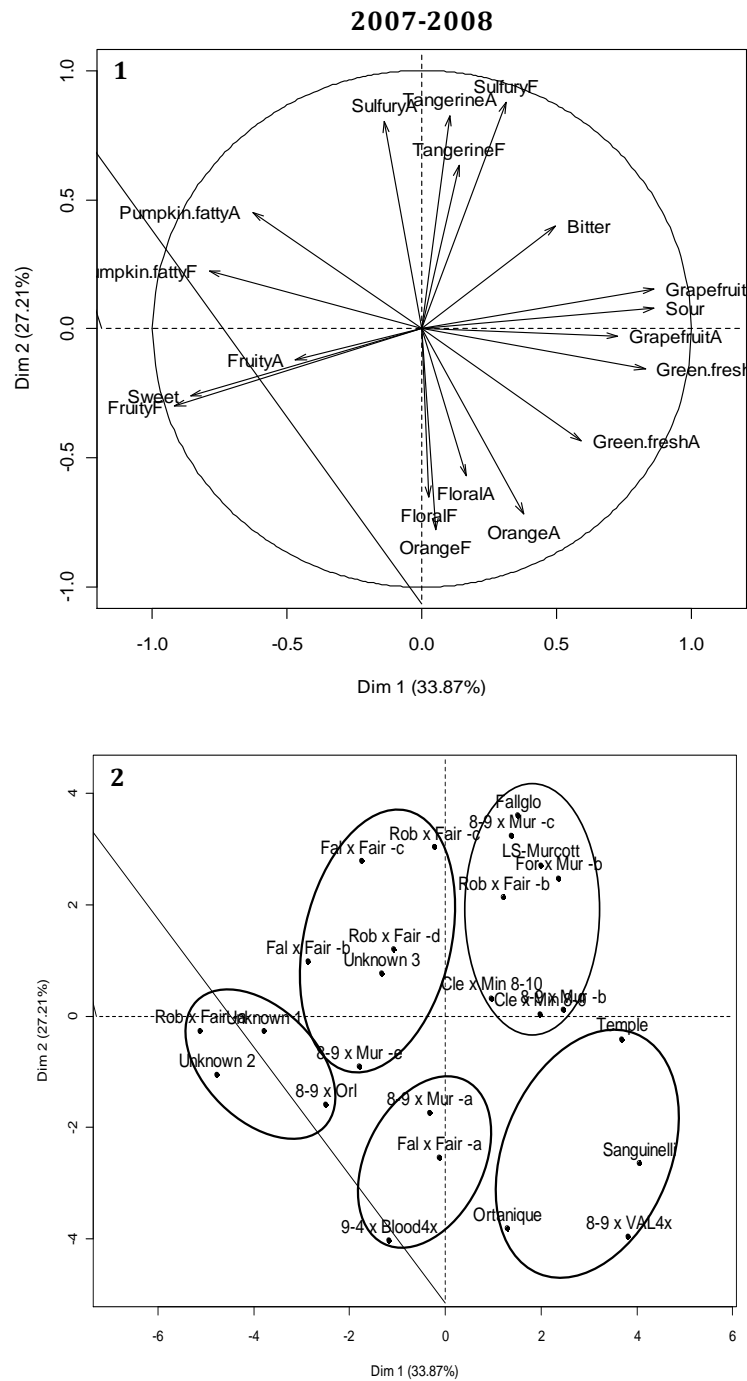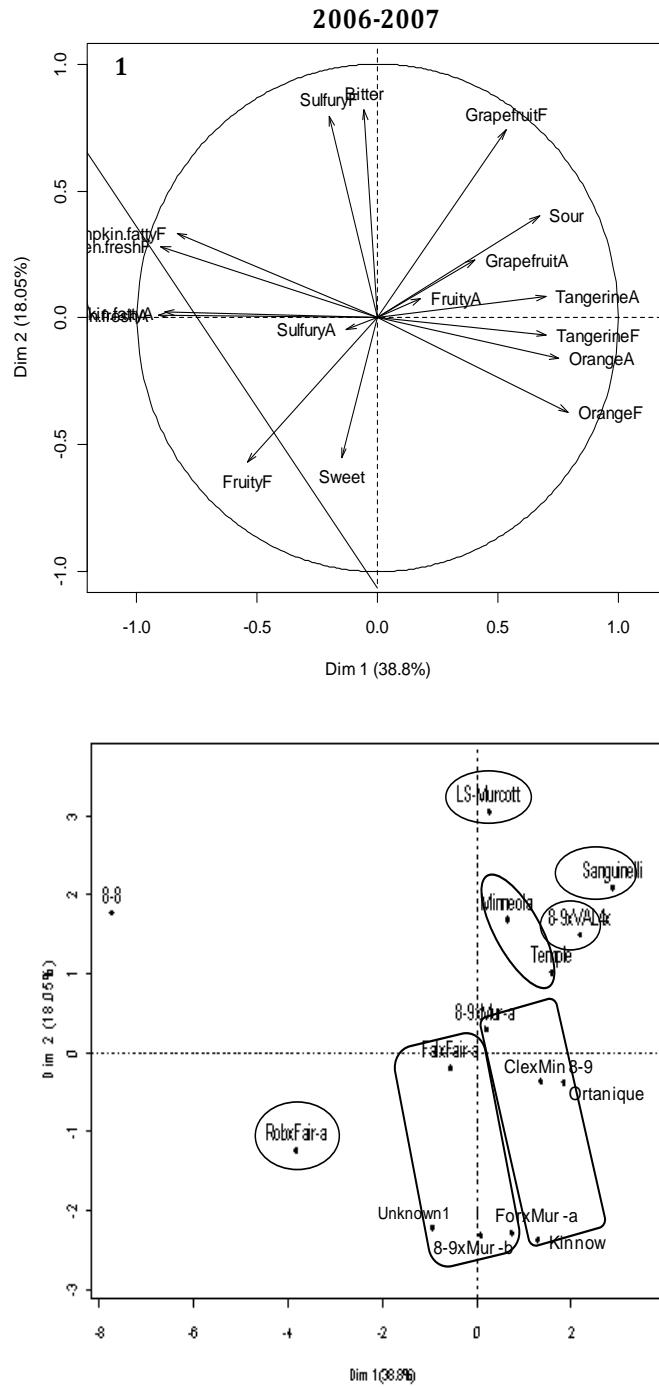
2006-2007

Figure 5.6: PCA score plot (1) and loadings plot (2) associated with the sensory analysis of the mandarin and hybrid juices evaluated in 2006-2007. The circles on the loadings plots represent the clusters determined by the HCPC.

130

The sensory profile space was somehow similar to those of 2007-2008 and 2008-2009; noting that for this set of juices there were no significant differences for the descriptor "Floral". One noticeable difference is for the descriptors ["Tangerine" and "Orange"] which were strongly and positively associated, whereas they were negatively correlated in the other years. A possible explanation is from the fact that only a subset of data is presented in this analysis; these two descriptors were better separated in the analysis with all the samples. Furthermore, there was a lack of familiarity of the panelists with the descriptors in 2006-2007; indeed, this year was an exploratory year (i.e., the panelists most likely confused both terms when evaluating the juices). Another difference is for the descriptor "Green/fresh" which was strongly correlated with "Pumpkin/fatty" and negatively correlated to "Sour" in 2006-2007, whereas it was strongly correlated with "Sour" and negatively correlated to "Pumpkin/fatty" in the following years. Once again, the panelists might have misused the descriptors. Although not represented on the loadings plot, the third dimension (13 % of the total variance of the data set) was significantly influenced by the descriptors "SulfuryA", "TangerineA" and "Fruity non-citrusA".

A total of eight clusters were determined by HCPC (Table 5.7) and are represented on the PCA score plot; the fact that there are many clusters is not surprising as the samples chosen for this exploratory year were expected to have very different sensory profiles. Figure 5.7 illustrates the main differences in aroma and flavor between the different groups of mandarin and hybrid juices. Cluster 1 included only sample 8-8, which was strongly associated with "Pumpkin/fatty" and "Green/fresh" notes. Cluster 2 included only one hybrid (Robinson × Fairchild tree a) that was characterized by a sweet and very fruity non-citrus flavor profile. Cluster 3 included one hybrid of unknown lineage (Unknown 1), 8-9 × Murcott tree b, Fortune × Murcott tree a and Fallglo × Fairchild tree a (which is not very well represented by the 2 PCs); these juices were described with sweet and high fruity non-citrus flavor. Cluster 4 included juices from Kinnow, Ortanique, one Clementine × Minneola hybrid (8-9) and one 8-9 × Murcott hybrid tree a (which is not very well represented by the 2 PCs), this group was characterized by sweet, tangerine and orange notes. Cluster 5 included only sample LS-Murcott and was described as very bitter with a sulfury flavor. Cluster 6 included Minneola and Temple juices, these juices were best represented in the

third dimension of the PCA with tangerine and sulfury aromas. Cluster 7 included only 8-9 ×
VAL4x hybrid that had strong sour notes (fruits harvested too early?) associated with
tangerine flavor and fruity non-citrus aroma (best visualized with PC 4). Finally, Sanguinelli
represented the last cluster and had the strongest grapefruit aroma and flavor, which
differentiated it from it from cluster 5, 6 and 7cluster 5, 6 and 7.
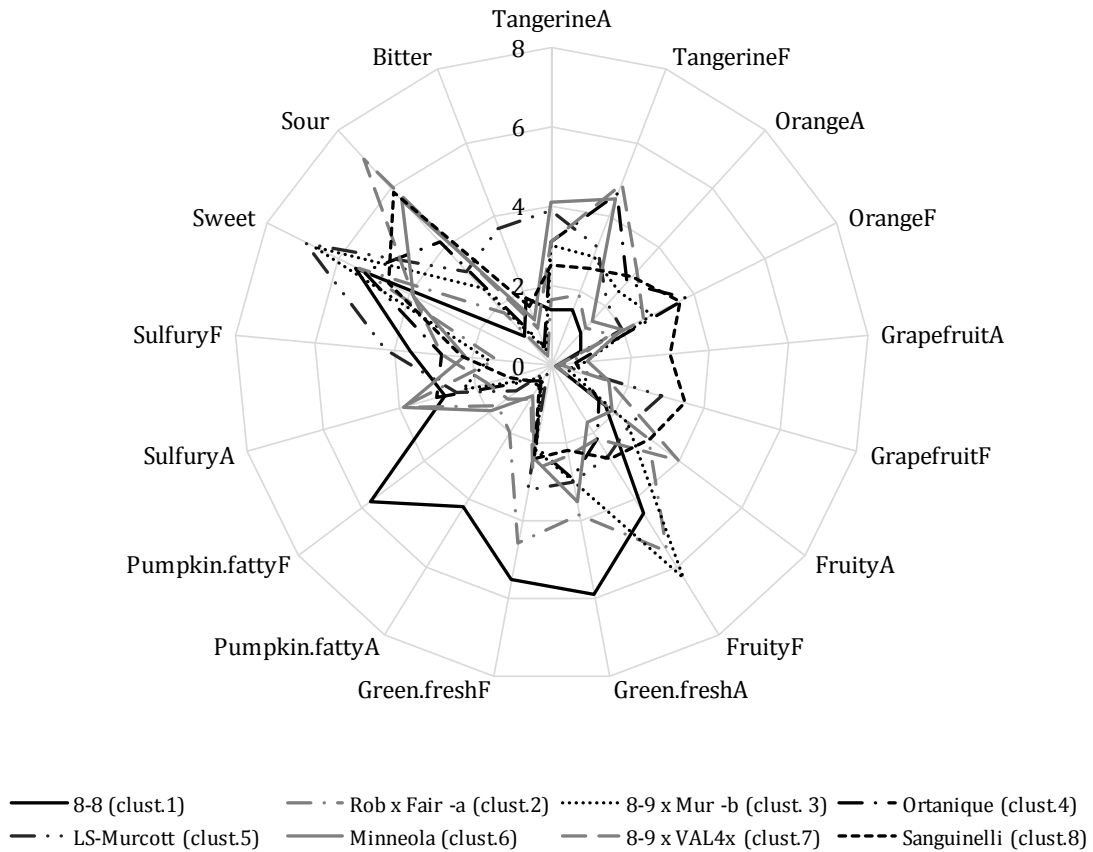
Figure 5.7: Sensory profile of the mandarin and hybrid juices for 2006-2007 (summarized
by one juice sample per cluster).

5.4. CONCLUSIONS

In summary, this chapter reports on the sensory characterization of juices made from commercial mandarin cultivars and new hybrids developed by the University of Florida Citrus Research and Education Center breeding program. A total of 61 mandarin juices were evaluated by descriptive sensory analysis between 2006 and 2009; some of the cultivars and hybrids were evaluated over the three harvesting seasons. Multivariate statistical tools such as PCA and HCPC were used to reveal the differences/similarities in flavor profiles and examine the influence of the parentage on the sensory quality.

Overall, there was a relatively good consensus between the sensory data collected on different years; some differences were noticed for the first year probably due to the lack of familiarity of the panelists with the descriptors. The sensory space was divided (in the first two dimensions of the PCA) in four groups of descriptors: ["Sweet", "Fruity non-citrus" and "Pumpkin/fatty"], ["Sulfury", "Tangerine" and "Bitter"], ["Sour", "Green/fresh" and "Grapefruit"] and ["Orange", "Floral"]. The descriptor "Tangerine" (both aroma and flavor) strongly influenced the third dimension, accounting for 39 %, 28 % and 22 % of its variation in 2008-2009, 2007-2008, 2006-2007, respectively.

The flavor profiles were very different between and within hybrids and cultivars, and juices made from fruits with common genetic background (parents and hybrids) tended to share some sensory characteristics. Juices made from fruits with orange in their lineage like Sanguinelli, 8-9 × VAL4x, Ortanique and Temple were described with floral and orange notes, or with grapefruit, green/fresh and sour profile. Similarly, some cultivars or hybrids with grapefruit in their background (Minneola, Clementine × Minneola, Robinson × Fairchild and 8-9 × Orlando) were described by bitter and grapefruit flavor and aroma. Other hybrids that were very particular in terms of sensory profile included sample 8-8 and one Robinson × Fairchild hybrid, they were respectively described by a pumpkin/fatty flavor profile and sweeter and stronger fruity non-citrus character. Some juices (LS-Murcott or some of its crossings) were characterized by sulfury and bitter flavor profiles which are usually considered as defects. It is possible that these juices developed bitterness after

juicing and in storage, which is a well-known phenomenon ("delayed bitterness") in citrus juices (Dea et al. 2010).

Differences in flavor profiles between harvesting seasons were noticed for some fruits (e.g., 8-9 × Orlando, 8-9 × Murcott tree b and c, Fallglo × Fairchild tree a, Fortune × Murcott tree a, Kinnow, Minneola, Unknown parent 2); there were most likely due to differences in maturity and/or seasonal effects. For instance, harvesting dates were adjusted based on data collected the previous year as the optimum harvest maturity was not known yet for many of the newly developed hybrids. Also, differences in weather between the seasons may have imparted stress or provided optimum growing conditions on the fruits, resulting in different flavor profiles across years.

In the next chapter, the sensory data was correlated to the chemical profile of the juices, obtained by instrumental measurements, to develop predictive models of the flavor profile. The correlations established can be used later to understand the relationships between some compounds and sensory characteristics, and also to reveal potential "flavor markers" specific to one sensory trait or specific to one fruit variety. This kind of information is valuable to understand the underlying genetic basis in the flavor generation pathways.

Chapter 6 : Prediction of mandarin juice flavor attributes using a multi-instrumental approach and chemometrics

Abstract. The objective of the study was to assess the feasibility of a non-targeted multi-instrumental approach to predict specific flavor attributes. Forty-six mandarin juice samples from different cultivars and hybrids were evaluated by instrumental (in 2010-2011) and descriptive sensory analyses (in 2007-2009). The instrumental analyses included the characterization of volatiles and non-volatiles by gas chromatography and ultra high performance liquid chromatography, respectively. The relationships between presence and relative intensities of chemical compounds (given as mass spectral signals RT-m/z) and flavor attributes (provided by means of descriptive sensory analysis) were investigated by means of partial least squares regression (PLSR), and then used to predict the scores of various flavor attributes. Several predictive models were developed, compared, and tested by external validation. The explanatory and predictive performances of the models were improved when combining all instrumental data into one single data set as opposed to individual ones. The best PLS model was obtained with mid-level data fusion, for which a preliminary variable selection was done before merging the different instrumental data sets. The predictive power of the selected model was tested using calibration and prediction sample sets (38 and 8 juices, respectively). A fairly robust model was obtained and a strong relationship between instrumental and sensory measurements was observed. The best predictions were obtained for the attributes grapefruit, sour, fruity non-citrus, orange and pumpkin/fatty ($0.5 < Q^2Y < 0.7$), whereas tangerine, bitter and floral yielded the poorest ones ($Q^2Y < 0.35$).

6.1. INTRODUCTION

The ability to predict flavor or sensory quality of food based on chemical composition has always been of interest for food scientists and food manufacturers, hoping it would serve as

an alternative to sensory evaluations (Noes and Kowalski 1989). Even though sensory analyses reflect "reality" better than instrumental methods, panels are costly, time consuming in preparation and cannot be used extensively over a long time period (fatigue). Instrumental-sensory correlations have been used since the late 1970s (Aishima and Nakai 1991) to establish associations between chemical stimuli (measured instrumentally) and flavor perceived (measured by sensory analyses) which can be used later for predictions. However, the task remains a challenge.

One major barrier to accurate predictions of flavor is due to the multi-dimensionality of the chemical stimuli involved in flavor perception. Even though the importance of both volatile and non-volatile compounds to flavor perception has been reported in the literature (Laing and Jinks 1996; Noble 1996; Keast et al. 2004), there are few studies that have used both to predict flavor (Tandon et al. 2003; Cabezas et al. 2006). In addition, most works considered in developing predictive models of flavor used only aroma compounds or known sensory-active compounds, and generally overlooked the contribution of other compounds.

It is reasonable to hypothesize that using a non-targeted approach, like flavoromics, to predict flavor can improve the predictive performance of models by including inputs from more chemical compounds. The concept and reasons for using flavoromics were introduced elsewhere (de Vos et al. 2008; Reineccius 2008; Charve et al. 2011). There are some reports in the literature that used a non-targeted approach to study flavor, however, the range of chemical stimuli investigated was often restricted to one type of instrumentation (either taste- or aroma-related compounds) (Lindinger et al. 2008; Pongsuwan et al. 2008a; Lindinger et al. 2009; Sugimoto et al. 2010). The feasibility of using a data-driven approach (rather than targeted) to predict flavor was demonstrated by Lindinger et al. (2008). These researchers successfully predicted the flavor (aroma) of coffee by monitoring the most discriminating ion traces in the headspace of coffee using proton transfer reaction- mass spectrometry (PTR-MS), irrespective of whether the chemical identity of the instrumental signals was known or not.

The use of chemometrics, and more precisely multivariate data analyses, is well-suited to establishing sensory predictions from instrumental measurements as it accounts

simultaneously for changes in several variables (vs. univariate techniques), thereby allowing a better picture of the flavor perception phenomenon. Partial least squares regression (PLSR), also known as projection to latent structures, is a method that finds a linear multivariate model to link two data matrices, X (predictor variables) and Y (dependent variables), to each other (Wold et al. 2001). For instance in this study, PLSR was used to search for a quantitative relationship between the intensities of mass spectral signals recorded across samples (X) and the associated sensory scores (Y). Once the mathematical relationship is established, the model can be used for predictions of new samples. PLSR has been successfully used in food science to correlate odor profile and sensory attributes of cheese (Biasioli et al. 2006; Cabezas et al. 2006), to predict the sensory quality of watermelon (Tarachiwin et al. 2008) and aroma properties of wine (Aznar et al. 2003; Cozzolino et al. 2008), among others. It is worthwhile noting that PLS is compatible with multi-block data where several blocks of data are available (Wangen and Kowalski 1989; Wold et al. 1996); for instance in comprehensive studies, data collected from different instruments on the same set of samples. One example is the study of Kreutzmann et al. (2008) who applied multi-block PLS to predict the sensory quality of raw carrots based on measurements of the dry matter and non-volatiles (block 1) and volatile compounds (block 2).

One important point to consider when building predictive models is the use of appropriate data pre-processing steps, which has been mentioned to significantly influence the results (e.g., the selection of markers) (van den Berg et al. 2006). As pointed out by Lindinger et al. (2008), sensory scores and instrumental measurements are very different in their nature and should be processed accordingly. Appropriate scaling, normalization and eventually transformation are some of the steps often needed; as well weighting and/or scaling factors may be used when combining blocks of data to adjust for differences in numerical range and/or size between blocks (Smilde et al. 2005; Eriksson et al. 2006b). In the case of the MS-based measurements, mean-centering and Pareto or unit-variance scaling are commonly used. There are disagreements in the way to process sensory scores. Most commonly, the sensory mean scores (i.e., average of the judges' scores for each sensory attribute) are used. However, it is well accepted that judges (even trained) use

differently scales, either in the level used or in the range given between products. Generalized procrustes analysis (GPA) (Chung et al. 2003) and isotropic scaling (Kunert and Qannari 1999) have been proposed to adjust for differences in ratings among judges but were not used here.

The sensory profile of mandarin fruits is influenced by their chemical composition, which depends on intrinsic (genetic background, maturity) and extrinsic (growing conditions) factors. While there is little control on external factors, it is possible to select and breed fruits with a given flavor profile. Understanding the relationships between the chemical composition and the sensory profile can assist breeders in developing fruits with enhanced/ improved sensory quality. The sensory characterization of newly developed hybrids is, therefore, critical in the selection process. Though sensory profiling with a trained panel is the method of choice, an instrumental methodology capable of predicting flavor would definitely accelerate the breeding program. Further, the identification of compounds ("flavor markers") associated with a specific quality trait would assist breeders in finding DNA-based molecular markers and enable them to select fruits with a particular flavor profile in their early stage of development (Hall 2006).

The objective of this study was to examine the feasibility of using "Flavoromics", a non-targeted, multi-instrumental approach in combination with chemometrics, to predict the sensory quality (flavor) of mandarin juices. Forty-six mandarin juices from different cultivars and hybrids were evaluated. The instrumental analyses included measurements of volatiles and non-volatiles by gas chromatography (GC) and ultra high performance liquid chromatography (UHPLC) coupled with time-of-flight mass spectrometry (TOF-MS), respectively; and the sensory profile of the juices was described by a trained panel. The relationships between chemical compounds (given as mass spectral signals RT-m/z) and the sensory descriptors were investigated by means of PLSR analysis, and used to predict the intensities of various flavor attributes. Several models were developed and compared, and the best model was subsequently tested by external validation.

6.2. MATERIALS AND METHODS

Mandarin juices. A total of 46 mandarin juices from different cultivars and hybrids was characterized by instrumental and sensory measurements; the juices were sourced from a larger study (Plotto et al. 2010). The mandarin cultivars and hybrids used to prepare the juices were from the University of Florida Citrus Research and Education Center breeding program and were collected over two harvesting seasons (2007-2009). They were selected by the breeder for their diversity in flavor profiles. The growing conditions and the fruits have been described elsewhere (Kerbiriou et al. 2007a; Miyazaki et al. 2011), and juices were presented previously (Table 5.1).

Chemicals. Purchase details and information related to the materials used are described in the Materials and Methods section in Chapter 3. In addition, ammonium hydroxide (5N, Fluka, Milwaukee, WI), D-mannitol (98 %, Fisher Scientific, Pittsburg, PA), (+)-catechin (> 95 %, TCI America, Portland, OR) and 3-hexanone (98 %, Acros Organics/ Fisher Scientific, Pittsburg, PA) were used.

Sample preparation. All juices were prepared on the day of harvest at the USDA/ARS Citrus and Subtropical Products Laboratory (for more details see Miyazaki et al. 2011). Upon juicing, samples were kept in glass containers at -20 °C until analysis. Juices were evaluated by sensory analyses between December 2007 and January 2009 at the USDA/ARS Citrus and Subtropical Products Laboratory, and by instrumental analyses at the University of Minnesota between December 2010 and February 2011. The juices were shipped and received frozen at the University of Minnesota, samples were stored at -20 °C until sample analysis. All samples were prepared in duplicate and analyzed in random order.

For the sensory evaluation, juices were thawed at room temperature on the day of the analysis and prepared in 3-digit coded 120 mL cups with lids (SOLO® Cup Company, Urbana, IL). Samples were presented in 45 mL portions and served at 14 °C ±2. Sample presentation was monadic and randomized (Williams design); the randomization pattern was only for the samples presented at one sitting.

For the GC analysis, juices were thawed overnight at 4 °C before being used the next day. Samples were prepared by adding 10 mL of juice to 20 mL headspace vials containing 3 g of salt (NaCl). Twenty μL of 3-hexanone (250 ppm in water) were added to each sample as an external reference to check for variation in MS ionization efficiency over time. The next day, samples were loaded into a cooling tray (10 °C) mounted on a Gerstel MPS2 autosampler equipped with a SPME option. The Maestro software (Gerstel, Baltimore, MD) was used as cycle composer with the co-current sample preparation feature enabled. Blanks (empty vials) were analyzed randomly throughout the experiment.

For the UHPLC analyses, solid phase extraction (SPE) was used for sample preparation. Samples were prepared on the same day than GC samples and were maintained at -20 °C until analysis. After thawing and centrifugation at 6800 g/ 5 min, the juices were used immediately or maintained on ice. C18-SPE cartridges (SampliQ, 500 mg sorbent, 3 mL tubes, Agilent, Wilmington, DE) were conditioned with 3 mL acetonitrile and equilibrated with 3 mL water/acetonitrile (95/5) before loading 2 mL of clarified juice onto the cartridge. Then, the SPE cartridge was washed with 3 mL water/acetonitrile (95/5) and eluted using 3 mL acetonitrile. The fractions collected are referred to as SPE-wash and SPE-eluate fractions, respectively. SPE blanks (cartridge loaded with nanopure water) were prepared similarly. A SPE Visiprep 12-port vacuum manifold (Sigma-Aldrich) fitted with disposable liners was used to increase sample preparation throughput and reproducibility. An aliquot of the fractions was used for the UHPLC analyses and transferred to 2 mL autosampler vials. Samples were diluted (1/4) using nanopure water with (+)-catechin (20 ppm), and acetonitrile/water (70/30) with mannitol (160 ppm) as diluents for SPE-eluate and SPE-wash fractions, respectively. Mannitol and (+)-catechin were used as an external references to check for variation in MS ionization efficiency over time. For organizational purposes, the UHPLC analyses of SPE wash and SPE eluate fractions were conducted one week apart.

Descriptive sensory analysis. A detailed description of the sensory procedure is given in the Materials and Methods section in Chapter 5. Briefly, a descriptive sensory analysis of the samples was performed at the USDA/ARS Citrus and Subtropical Products Laboratory by 13-15 trained panelists over a three-year period (Note: data from the first year 2006-2007

were not included here; this year was exploratory for sample selection and panel training). A set of 19 descriptors was used to characterize the sensory profile of the juices including aroma (orthonasal perception), flavor (retronasal perception after swallowing sample) and basic taste descriptors. For the present study, only taste and flavor descriptors were kept. It was shown previously that aroma and flavor descriptors had similar scorings (see Chapter 5). The descriptors selected were: Tangerine, Orange, Grapefruit, Fruity non-citrus, Floral, Green/fresh, Sulfury, Pumpkin/fatty, Sweet, Sour and Bitter. Panelists were asked to rate the intensity of the descriptors using a 15-point category scale (1=low, 15=high); reference standards, fresh water and crackers were provided at each session. A maximum of four samples were presented at one sitting, and the sensory evaluation lasted over one month each year of the study.

Instrumental analyses. To obtain a comprehensive view of the chemical composition of the mandarin juices, samples were analyzed by headspace solid phase micro-extraction-gas chromatography (SPME-GC) and by solid phase extraction-ultra high performance liquid chromatography (SPE-UHPLC) for volatiles and non-volatiles, respectively. The applicability of these methods for a non-targeted approach of chemical stimuli of flavor has been reported (Charve et al. 2011).

For the GC analysis (volatiles), samples were incubated at 50 °C for 5 min under constant stirring (500 rpm) before the SPME extraction. A 75 μm carboxen/ polydimethylsiloxane fiber (Supelco, Bellefonte, PA) was exposed for 20 min to the headspace of the mandarin juice maintained at 50 °C. The fiber was then desorbed into the injection port of a gas chromatograph (260 °C/ 10 min). Prior to extraction, the SPME fiber was conditioned according to the manufacturer's recommendations. Analyses were performed on an Agilent GC 7890A equipped with a HP-5MS capillary column (30 m length, 0.25 mm i.d., 0.25 μm film thickness) (J&W Scientific, Folsom, CA) and coupled with an electron impact, time-of-flight mass-spectrometer (EI-TOF-MS) (GCT, Waters, Milford, MA). The operating conditions were as follows: GC-MS: inlet split 5:1; initial oven temperature: 40 °C increased at 4 °C/min to 120 °C, then 40 °C/min to 250 °C and hold for 1 min; TOF-MS: 40-400 amu, centroid mode, scan time: 0.2 sec, inter-scan time: 0.02 sec, 2750 volts, 70 eV. An inlet liner (78.5 mm

x O.D. 6.5 mm x I.D. 0.75 mm) designed to achieve sharper peaks with SPME/GC analyses was used.

For the UHPLC analyses (non-volatiles), analyses were performed on a Acquity UPLC quadrupole-TOF-MS system (Synapt MS, Waters, Milford, MA) equipped with an electrospray ionization (ESI) source operating first in positive and then, in a separate analysis, in negative mode. Two complementary separation modes (hydrophilic interaction liquid chromatography HILIC and reversed-phase RP) were used. The SPE-wash fraction (highly hydrophilic) was injected onto a Bridged Ethyl Hybrid (BEH) amide column (2.1 mm x 100 mm, 1.7 μm, Waters, Milford, MA) and the SPE-eluate fraction (relative hydrophobic) onto a BEH C18 column (2.1 mm x 50 mm, 1.7 μm, Waters, Milford, MA), the columns were maintained at 40 °C. Three μL (partial loop overfill mode) of sample was injected, the autosampler chamber was maintained at 12 °C. The mobile phases were: A: ammonium acetate (10 mM + 0.1 % ammonium hydroxide), B: 95/5 acetonitrile/ ammonium acetate (10 mM) + 0.1% ammonium hydroxide for HILIC separation; and A: formic acid (0.1 %); B: acetonitrile + 0.1 % formic acid for RP separation. For the RP separation, the gradient (flow rate: 0.5 mL/min) ramped linearly from 0.5-20 % B over 4 min, then to 100 % over 3.5 min and held for 1 min before returning to initial conditions and re-equilibrated for 1.5 min. For the HILIC separation, the gradient (flow rate: 0.3 mL/min) started at 90 % B, held for 2 min then decreased linearly to 50 % over 5 min and held for 0.5 min before returning to initial conditions and re-equilibrated for 2.5 min (flow rate: 0.5 mL/min). The MS operating conditions were as follows: source temperature: 120 °C; desolvation temperature: 350 °C; desolvation gas: 700 L/h; cone gas: 50 L/h; capillary voltage: 3 kV (ESI-), 3.2 kV (ESI+); cone voltage: 35 V (ESI-), 30 V (ESI+); extraction cone: 4 V (ESI-), 4 V (ESI+); TOF-MS: 50-1500 amu, centroid mode, scan time: 0.3 sec. Leucine enkephalin (m/z 556.2771 and 554.2615 in positive and negative mode, respectively) at a concentration of 0.5 ng/μL was used as a lock-mass for mass accuracy and infused at a flow of 40 μL/min during the experiment.

Data pre-processing. Prior to the modeling, data were organized and pre-processed. Instrumental data sets were designated as matrix X_i (predictors) and sensory data as matrix Y (responses).

For the instrumental data, mass spectral peaks (given as retention time RT – mass-to-charge ratio m/z) from GC and UHPLC chromatograms were extracted using Metalign freeware (available at http://www.metalign.nl). The program has been described elsewhere (de Vos et al. 2007; Lommen 2009) and successfully used in non-targeted metabolomic research. The output is a matrix containing a list of the mass spectral signals (also referred to as variables or features) detected across samples: [I samples × J instrumental variables RT-m/z] with their corresponding signal intensity. Five matrices were obtained, one per type of analysis: GC (X_1), Reversed-Phase UHPLC analysis in negative (RP neg. X_2) and positive (RP pos. X_3) ionization mode, HILIC UHPLC analysis in negative (HILIC neg. X_4) and positive (HILIC pos. X_5) ionization mode.

The UHPLC data sets were manually filtered to remove isotopes and dimer signals. The GC data set was reduced using the multivariate mass spectra reconstruction (MMSR) approach to filter out redundant signals from EI-GC-MS and to narrow down to one major m/z per detected peak (Tikunov et al. 2005); and output was inspected for misalignment or missed peaks. The MMSR procedure was graciously performed by Dr Yury Tikunov (University of Wageningen, The Netherlands). Variables corresponding to the references (3-hexanone, (+)-catechin and D-mannitol) were removed from the data sets. All variables were normalized to total sum intensity for each sample (intensity of the variable/ sum of intensities of all variables in the sample $\times$ 10,000). Then, all X matrices were reduced using a in-house script developed in R (http://www.r-project.org), which discarded features that did not differ significantly between samples (one-way analysis of variance ANOVA, P < 0.05); meaning, retaining only features that contribute the most to the variance of the data set. Finally, for all instrumental data, the mean intensity of each variable in each sample was calculated by averaging replicate samples.

For the sensory data, the panel and individual panelist performances were evaluated as detailed in Chapter 5. After selection of the panelists, the mean scores of each descriptor were calculated for each sample by averaging the judges' scores. The sensory results were then organized, by year, in a matrix Y [I samples × J descriptors] with their corresponding mean scores.

Overview of the data sets. All matrices were imported into Simca-P+ statistical software (v. 12.0, Umetrics, Umea, Sweden) for the multivariate data analyses. Instrumental data sets (X_1-X_5) were initially processed separately and then compiled in a unique matrix in such a way that samples were in the shared mode. Similarly, sensory data (2007-2008, 2008-2009) were first examined separately before being merged. Instrumental data were column-centered and scaled to unit variance ($1/SD_k$, where SD is the standard deviation of the variable k); sensory data were column-centered and scaled to unit variance, by year.

Before applying PLSR analysis, each data set was first inspected by principal component analysis (PCA). All PCA models were inspected for their quality of fit and predictive ability (statistics R^2 and Q^2, respectively). For the instrumental data, the presence of outliers was checked by looking at samples in the score plot being outside of the 95 % tolerance ellipse given by Hotelling's T^2. For the sensory data, the loadings plot was examined to see how Y-variables were correlated to each other.

Partial least squares regression modeling (PLSR). PLSR was used to build predictive models of flavor using instrumental data as independent variables (predictors X) and the sensory attributes as response variables (Y). Initially, the entire sample set (46 juices) was used for the development of the models and their comparison. The following PLS models were developed: a) five individual PLS models for each instrumental data set (models 1-5), b) one "low-level fusion" PLS model with the instrumental data sets merged (model 6), c) one "mid-level fusion" PLS model with preliminary variable selection (model 7). A weight factor equal to $1/\sqrt{k_{block}}$ (with k, number of variables in a block) was applied to each block of instrumental data in model 6 and 7. For model 7, a variable selection was done for each instrumental data set before merging: variables with a VIP (Variable Influence on Projection) value inferior to 1 were discarded.

The explanatory and predictive qualities of the models were respectively evaluated through $R^2Y(cum)$, the cumulative fraction of the variation of Y explained by the components of the model, and $Q^2Y(cum)$, the cumulative fraction of the variation of Y that can be predicted by the components of the model according to cross-validation (Eriksson et al. 2006e). Values close to 1 indicate an excellent model. The models were compared and

the one having the best overall performance was chosen for future predictions of flavor scores.

Validation of the PLS model. The predictive validity of the selected PLS model was then examined. The total sample set was divided in two: one calibration set (38 juices) used for model development (the selected PLS model was initially built with all samples which may over-fit the data), and one prediction set (8 juices) used for model testing (Eriksson et al. 2006c). The samples were chosen in a manner that they were similar to the calibration samples and to assess the model over a large range of responses. With the developed model, the sensory descriptors' scores of samples in the prediction set were predicted (predicted Y-data set) from the instrumental data collected. For each modeled response (sensory scores), the predicted values from the model and observed values given by descriptive sensory analysis were plotted against each other. The coefficient of determination r^2 for the regression line represented the predictive power of the model for a given Y, it is also referred to as $Q^2Y_{external}$. The prediction error of the model was also evaluated with the parameters RMSEE (Root Mean Squared Error Estimate) and RMSEP (Root Mean Squared Error of Prediction) calculated with the calibration and prediction sample sets, respectively. RMSEE and RMSEP describe the error between observed and predicted values, and therefore are considered as a measure of accuracy of the model.

Flavor markers - tentative identification. A few flavor markers contributing the most to the prediction of some sensory descriptors were tentatively identified. No efforts were made to identify all marker compounds or to confirm their identity as this task was not relevant to the present objectives and would have required considerable work beyond the expectations of this doctoral research. Variables with a high positive regression coefficient in the prediction model of some sensory descriptors were selected. For variables belonging to the GC data set (volatiles), the mass spectrum associated with the selected variable (RT-m/z) was retrieved from the chromatogram and compared to Wiley Registry with NIST 08 mass spectral library. Compounds were tentatively identified based on the similarity of their mass spectrum with the library hit, their retention indices and sample knowledge. For variables belonging to the UHPLC data set (non-volatiles), the accurate mass of the feature was used to calculate the elemental composition and to search the online Scripps Center

Metlin database (http://metlin.scripps.edu/). A tandem mass spectrometry (MS/MS) strategy (collision energy ramped from 10 to 45 V) was also used to gain further information on compound's structure (and identity). When available to us, a reference compound was analyzed to confirm identity.

6.3. RESULTS AND DISCUSSION

The main objective of this study was to investigate the feasibility of using Flavoromics, a non-targeted instrumental approach, to predict the flavor quality of mandarin juices. Juices were analyzed by GC-TOF-MS, UHPLC-TOF-MS and by descriptive sensory analysis; and partial least-squares regression was applied to link information on presence and relative intensity of chemical signals to variations in sensory profiles.

6.3.1. Overview of the data sets

Upon analyses, each juice sample was characterized by a set of chemical signals (X_i: from instrumental analyses) and flavor descriptor scores (Y: from sensory analyses). Before applying PLSR, each data set was first inspected by PCA for outliers or anomalies and also to examine how Y-variables (sensory attributes) were correlated to each other.

6.3.1.1. Sensory data

Results describing the selection of judges and descriptors, and the sensory space were provided in detail in Chapter 5. As mentioned earlier, aroma and flavor descriptors corresponding to the same term were very well associated and therefore, only flavor descriptors were included here.

PCA was applied to the combined sensory data set and a three-component model was obtained with an explained variation (R^2X) of 79.4 % (PC1: 46.5% and PC2: 21.9 %) and a predicted variation (Q^2X) of 44.1 %. The PCA score plot was then examined to see how samples related to each other, recalling that samples being close to each other have similar sensory profiles. As seen in Figure 6.1 (left), the flavor profiles were very different between

and within mandarin hybrids and cultivars confirming their applicability of use for PLS modeling. An important fact to consider when developing predictive models is the proper selection of samples used: the selection of representative and diverse samples is crucial to ensure that the developed model will be applicable to unknown samples. Sample selection can be achieved either through statistical experimental design (factorial, D-optimal, mixture designs, etc.) or proper sampling strategy (Eriksson et al. 2006c; Trygg et al. 2007). In this study, samples were selected by the breeder on the basis of their genetic background and sensory results obtained during the exploratory phase of the project (2006-2007). Most hybrids of the same tree had similar sensory profile across years; some differences in flavor profiles between harvesting seasons were noticed for some fruits (e.g., 8-9 × Orlando, 8-9 × Murcott tree b and c) and were most likely due to differences in maturity and/or seasonal effects.
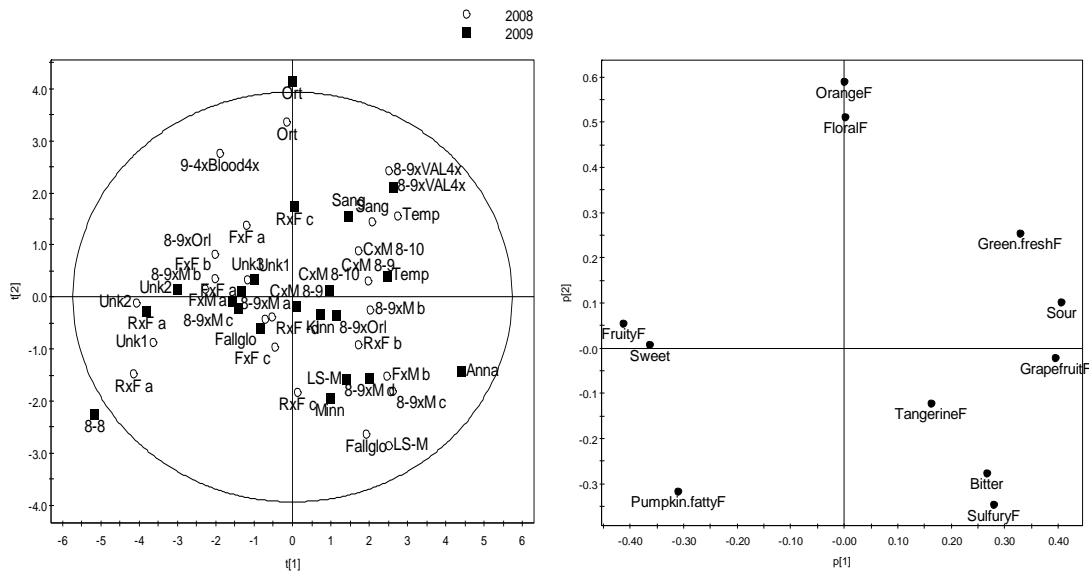

Figure 6.1: PCA score plot (left) and loadings plot (right) associated with the sensory analysis of mandarin and hybrid juices (combined data sets for 2007-2008 and 2008-2009).

One advantage of PLSR over other multivariate techniques is its ability to analyze and model several responses in a single model, which gives an easier understanding of the overall system and also a stronger model (Wold et al. 2001). Hence, the loadings plot (Figure 6.1, right) obtained for the combined sensory data set (2007-2008 and 2008-2009) was examined to see how Y-variables were correlated to each other and if they should be modeled together. Most variables were correlated between them and well summarized by the first three principal components (PCs), explaining about 79.4 % of their variation. In the first two dimensions, the sensory space was roughly divided in four groups of descriptors (Figure 6.1): ["Sweet", "Fruity non-citrus" and "Pumpkin/fatty"], ["Sulfury", "Tangerine" and "Bitter"], ["Sour", "Green/fresh" and "Grapefruit"] and ["Orange", "Floral"]. Although not represented on the loadings plot, the third dimension (explaining 11 % of the total variance) was mostly influenced by the descriptor "Tangerine". Based on these results, all sensory descriptors were modeled together (instead of having a separate model for each response).

6.3.1.2. Instrumental data

Instrumental signals were extracted from chromatograms using the specialized freeware Metalign. Briefly, the program extracted mass spectral signals from the chromatograms after baseline and noise correction, and aligned chromatograms across samples. Upon processing and reduction of the data sets using one-way ANOVA, the GC, RP neg./UHPLC, RP pos./UHPLC, HILIC neg./UHPLC and HILIC pos./UHPLC data sets resulted in 134, 439, 309, 300 and 255 features (or instrumental variables) detected across samples, respectively. Mass spectral signals were mean-centered and scaled to unit variance so that instrumental response factors were removed from the data. A preliminary examination of the data (before averaging replicates) showed good clustering among duplicates, indicating the methods' repeatability.

PCA was applied to each instrumental data set for inspection purposes (Table 6.1). Overall, the goodness of fit (R^2) for the PCA models obtained for the instrumental data were good unlike the goodness of prediction (Q^2) which was more inconsistent across instrumental analyses. The poor Q^2 values were not totally unexpected given the large

number of variables and indicated noise in the data sets that was not accounted for by the PCA models, meaning a there was a lot of systematic variation in the data. This may affect the quality of the PLS models as they are negatively affected by systematic variation in the X matrix that is not related to the Y matrix (Trygg et al. 2007). The use of algorithm filters (e.g., Orthogonal Signal Correction, derivatives) (e.g., Biasioli et al. 2006; Pongsuwan et al. 2008b) or a modified PLS technique (Orthogonal Projections to Latent Structures OPLS and O2PLS) has been proposed to diminish artifact and noise associated with the instruments and to separate the unrelated systematic variation from the data set (Gabrielsson et al. 2006); but they were not considered here.

Table 6.1: Summary of the PCA results obtained for the instrumental analyses of juices (all samples included).

Instrumental analysis	A [a]	$R^2X_{(cum)}$ [b]	$Q^2X_{(cum)}$ [c]	Outlier
GC	8	0.69	0.15	Sanguinelli (2008-2009), Temple (2008-2009)
UHPLC/ RP negative	5	0.71	0.56	Sanguinelli (2008-2009), Robinson × Fairchild tree a & c (2009)
UHPLC/ RP positive	6	0.71	0.50	Robinson × Fairchild tree a (2009)
UHPLC/ HILIC negative	9	0.71	0.22	Anna (2009)
UPHLC/ HILIC positive	9	0.69	0.22	Sanguinelli (2009), 8-8 (2009)

[a] Number of significant principal components determined by cross-validation (auto-fit option in SIMCA-P+); [b] Cumulative fraction of the variation of the data set X explained by the model; [c] Cumulative fraction of the variation of X that can be predicted by the model according to the cross-validation.

Even though some samples were located outside of the 95 % tolerance ellipse, no strong outliers (except for Sanguinelli 2009 in GC) were detected in the score plots (not presented); these samples were then included in the remainder of the analyses. Sample grouping obtained with the PCA models was different between instrumental methods (RP-

UHPLC vs. HILIC-UHPLC vs. GC), though juices made from same hybrids were close to each other. This indicates that each instrumental analysis conveyed different information regarding sample chemical composition, reinforcing the usefulness of using complementary instrumental methods for this study.

6.3.2. Predictive model – selection

As noted earlier, the primary objective of this study was to develop a predictive model relating instrumental measurements to the sensory descriptors of mandarin juices using PLSR. First, several predictive models were developed and compared in terms of explanatory $R^2Y(cum)$ and predictive qualities $Q^2Y(cum)$ for each sensory descriptor (Table 6.2) using the entire set of samples (2007-2008, 2008-2009). For comparison purposes across models, additional PLS components were forced into the models even though cross-validation suggested a different number of significant components. The model having the best overall performance was then chosen for flavor predictions, and tested using external validation.

For the different PLS models, the number of significant PLS components needed to explain the majority of the variation in Y varied from 0 (no structure in the multivariate space) to 4. As expected, the number of components needed was higher for models 6 and 7 for which the complexity of the information modeled was increased since the different data sets were fused.

Table 6.2: Summary of the PLS models [a] obtained with the entire sample set (46 juices).

		Model ID						
		GC (M$_1$)	RP neg. (M$_2$)	RP pos. (M$_3$)	HILIC neg. (M$_4$)	HILIC pos. (M$_5$)	Low-level fusion (M$_6$)	Mid-level fusion (M$_7$)
	A [b]	4 (2)	4 (0)	4 (3)	4 (3)	4 (1)	4	4
Tangerine	R^{2}Y [c]	0.52	0.22	0.17	0.54	0.55	0.64	0.62
	Q^{2}Y [d]	0.31	-0.09	-0.14	0.04	0.25	0.36	0.37
Orange	R^{2}Y [c]	0.65	0.57	0.76	0.57	0.69	0.86	0.84
	Q^{2}Y [d]	0.22	0.07	0.14	-0.002	0.4	0.46	0.54
Floral	R^{2}Y [c]	0.38	0.31	0.33	0.33	0.37	0.43	0.4
	Q^{2}Y [d]	-0.03	-0.14	-0.1	-0.06	0.08	0.01	0.1
Fruity non-citrus	R^{2}Y [c]	0.67	0.67	0.56	0.64	0.63	0.7	0.73
	Q^{2}Y [d]	0.35	0.27	0.22	0.3	0.22	0.35	0.51
Sulfury	R^{2}Y [c]	0.49	0.51	0.43	0.56	0.59	0.6	0.62
	Q^{2}Y [d]	0.11	0.01	0.007	0.12	0.23	0.27	0.4
Pumpkin/ fatty	R^{2}Y [c]	0.63	0.52	0.6	0.63	0.64	0.7	0.73
	Q^{2}Y [d]	0.43	0.23	0.22	0.32	0.35	0.39	0.5
Grapefruit	R^{2}Y [c]	0.76	0.63	0.56	0.67	0.79	0.83	0.82
	Q^{2}Y [d]	0.42	0.27	0.18	0.43	0.55	0.59	0.63
Green/ fresh	R^{2}Y [c]	0.5	0.53	0.55	0.52	0.49	0.6	0.59
	Q^{2}Y [d]	0.32	0.36	0.38	0.36	0.34	0.45	0.48
Sour	R^{2}Y [c]	0.68	0.58	0.63	0.84	0.8	0.85	0.84
	Q^{2}Y [d]	0.46	0.29	0.32	0.63	0.56	0.64	0.7
Sweet	R^{2}Y [c]	0.6	0.56	0.51	0.55	0.57	0.6	0.59
	Q^{2}Y [d]	0.3	0.18	0.29	0.39	0.31	0.38	0.45
Bitter	R^{2}Y [c]	0.41	0.55	0.48	0.36	0.38	0.48	0.52
	Q^{2}Y [d]	0.05	0.16	-0.05	-0.08	0.02	0.06	0.21

[a] For each PLS model, all responses were analyzed together. These models were not validated with external samples; [b] Number of significant PLS components. For comparison purposes, additional components were fitted if the number of significant LVs (given between parentheses) determined by cross-validation (auto-fit option in SIMCA-P+) was not equal; [c] Cumulative fraction of the variation of Y explained by the model; [d] Cumulative fraction of the variation of Y that can be predicted by the model according to the cross-validation.

When looking at the individual models of the instrumental analyses (model 1-5), each one contributed to generate some information about the relationships between chemical composition and sensory data (Table 6.2). When the response (sensory descriptor) is well modeled (i.e., for both fit R^2 and prediction Q^2) by one type of instrument, it indicates that this instrument contains most of the relevant chemical information that is related to the variation in the response. For instance, the flavor descriptors pumpkin/fatty, grapefruit, orange, tangerine and fruity non-citrus were best explained and predicted by UHPLC/HILIC and GC measurements; whereas sulfury was best modeled by UHPLC/HILIC neg. and green/fresh was equally explained and predicted by each instrument. The taste descriptors sweet and sour were better modeled for measurements from UHPLC/HILIC, whereas bitter was better explained and predicted by UHPLC/RP neg. On the other hand, the descriptor floral was poorly explained and predicted by all models.

We also examined the potential gain in explanatory and predictive power when combining chemical information (i.e., relative to volatiles and non-volatiles) collected from different instruments on the same set of samples. Our hypothesis was if the information carried by the different instruments is complementary, the predictive performance of the model will improve. Indeed, data fusion has been proposed in "omics" studies as a means to obtain a comprehensive view on the metabolome of an organism or biological system (Smilde et al. 2005). There are several examples in the literature where different analytical techniques were combined to describe complex samples (e.g., Forshed et al. 2007; Cozzolino et al. 2008; Tikunov et al. 2010). In this study, two levels of data fusion were considered: low and mid-level with 1437 and 576 variables in total, respectively. Low-level fusion (model 6) consisted of simply merging the data sets together, whereas mid-level fusion (model 7) involved a preliminary variable selection prior merging. For the mid-level fusion model, the VIP parameter (Variable Influence on Projection) was used to eliminate the instrumental variables that contributed the least to the variation in the responses: for each instrumental data set, the variables were ranked according to their VIP value and variables with a VIP<1 were discarded. A weighting factor was also applied to each block of data to adjust for the differences in size between blocks in terms of number of variables, as larger

blocks are not synonyms with higher importance for the model (Smilde et al. 2005; Eriksson et al. 2006b).

Overall, the performances of the models were increased when combining all instrumental data into one single data set as opposed to individual ones: models 6 and 7 gave the highest explanatory (R^2Y) and predictive power (Q^2Y) for most of the descriptors. This confirmed that each individual subset conveyed complementary information and the fact of combining them improved the overall description of the sensory profile. These results agreed with previous studies that reported improvements in the predictive abilities of the models when different instrumental measurements were combined. For instance, the prediction of soybean flour quality properties was improved when combining near-infrared (NIR) and mid-infrared spectra (Brás et al. 2005). Similarly, Cozzolino et al. (2008) reported better PLS models for the sensory attributes "developed" and "floral" of Australian Riesling wines when MS-eNose, visible (VIS) and NIR measurements were combined.

Even though models 6 and 7 had relatively similar statistics, model 7 was chosen for the remainder of the study given its ability to give slightly better predictions. The overall variation in the sensory data Y was well explained by the model ($R^2Y_{cum}>0.66$) and the overall predictive performance was judged fair ($Q^2Y_{cum}=0.44$). The individual R^2Y values for each descriptor ranged from fair (40 % of the variation in the Y-data is modeled) to excellent (85 %), and the Q^2Y from very poor (10 % of the variation of Y is predicted) to excellent (70 %). Of note, this PLS model was re-fitted (not shown) after removing "floral" but resulted in very little improvements in overall R^2 and Q^2 values.

While we could have developed individual models for each sensory descriptor, we have decided to use a single model to describe the entire sensory quality. We wanted to examine the feasibility of developing one unique predictive model of flavor using a non-targeted instrumental approach. One possible disadvantage of such comprehensive model is a reduced performance in the prediction accuracy of the individual responses, as has been previously reported (Biasioli et al. 2006). Hence, we compared the explanatory and predictive statistics for the sensory attributes of the PLS model when the responses were modeled individually (also referred to as PLS1 modeling algorithm) or all together (also

referred to as PLS2 modeling algorithm, and here corresponding to model 7). Of note, these models were not tested by external validation. As seen in Figure 6.2, there were no drastic differences between the performances obtained from the two PLS models. However, one possible benefit of using separate models is for picking variables (i.e., compounds) linked to a particular sensory attribute. Even so, we decided to keep one unique model for the purpose of this study. Model 7 was then tested for its ability to provide predictions for the sensory attributes of mandarin juices (prediction set) from instrumental measurements.
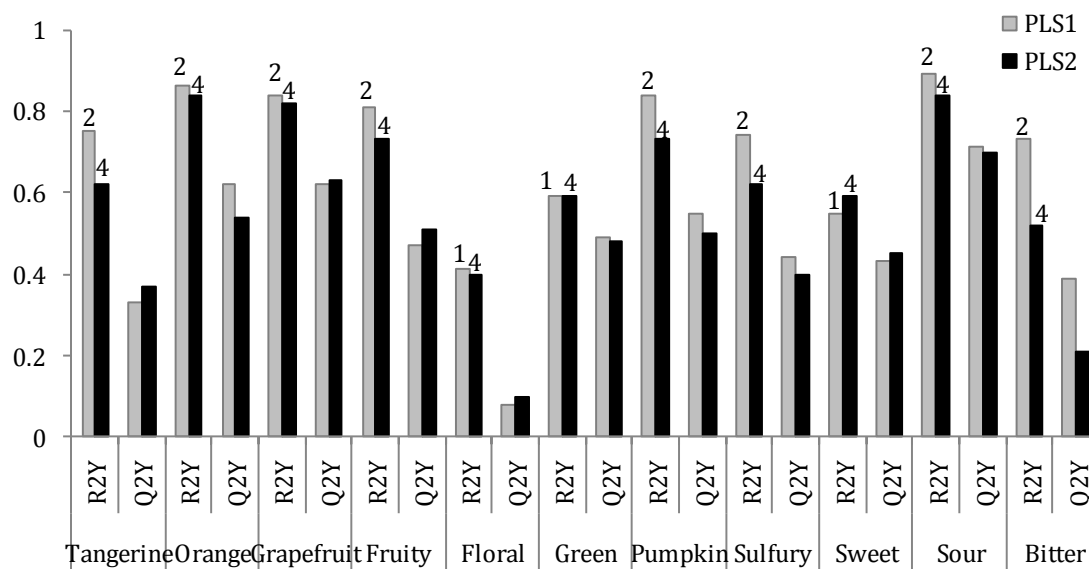
Figure 6.2: Comparison of the performances of the PLS model (mid-level data fusion) obtained for all samples when responses were modeled individually (PLS1) or together (PLS2). Numbers indicate the number of significant components included in the model.

6.3.3. Predictive model – validation

Next, the predictive validity of the selected PLS model (model 7) was examined (recalling that all samples were used for the initial model selection). The entire sample set was split in two as follows: 38 juices were chosen to form the calibration set (model development) and the remaining 8 juices (four samples from each year) were used as the prediction set

(model testing). Such a validation procedure is referred to as external validation since the developed model is evaluated with a set of external juice samples that have not been used in model building. It is the most rigorous way to evaluate the predictive power of a model (Eriksson et al. 2006c). The following samples constituted the prediction set: 8-9 × VAL4x (2008), Unknown 1 (2008), Fallglo × Fairchild tree c (2008), 9-4 × Blood4x (2008), 8-9 × Murcott tree d (2009), Fallglo (2009), Temple (2009) and Minneola (2009). These samples represented diversity in flavor profile across the juices.

The PLS model obtained (4 significant components) for the calibration set explained and predicted 68 % and 43 % of the total variation in Y, respectively. The individual performances of the sensory attributes are provided in Table 6.3. In summary, sour, grapefruit and fruity non-citrus were the best modeled, whereas floral, and to a lesser extent bitter, were the worst. The attributes orange and pumpkin/fatty were also well modeled; and green/fresh, sweet, sulfury and tangerine were moderately well modeled.

Table 6.3: Performances of the PLS model for the calibration and prediction sample sets.

	Calibration set			**Prediction set**	
	R^2Y [a]	Q^2Y [b]	RMSEE [c]	r^2 (Q^2Y_{ext}) [d]	RMSEP [e]
Sour	0.84	0.68	0.42	0.82	0.40
Grapefruit	0.81	0.60	0.47	0.90	0.25
Fruity non-citrus	0.72	0.54	0.56	0.76	0.57
Pumpkin/fatty	0.79	0.50	0.49	0.66	0.65
Orange	0.83	0.50	0.44	0.88	0.30
Green/fresh	0.57	0.45	0.64	0.76	0.78
Sweet	0.59	0.43	0.66	0.47	0.79
Sulfury	0.69	0.42	0.55	0.35	1.05
Tangerine	0.66	0.32	0.65	0.24	0.64
Bitter	0.53	0.26	0.75	0.27	0.63
Floral	0.39	0.06	0.76	0.28	1.09

[a] Cumulative fraction of the variation of Y explained by the model; [b] Cumulative fraction of the variation of Y that can be predicted by the model according to the cross-validation; [c] Root Mean Square Error Estimate; [d] Coefficient of determination for the regression line between observed and predicted scores; [e] Root Mean Square Error of Prediction.

We also examined the correlation structure between X- and Y-data described by the PLS model by means of the score plot t_1/u_1. Ideally, the relation between X and Y is linear (slope of 1) with little scattering of the samples around the diagonal line. As noticed in Figure 6.3, there was a good correlation between the instrumental measurements and the sensory responses (r^2=0.77). Most samples were located around the diagonal line, meaning that when the chemical composition changed (X) there was a good relation with the changes in sensory profile (Y). Samples located outside of the diagonal line (e.g., Temple 2008, 8-9 × Murcott tree e, Fallglo, etc.) in the first PLS component were accounted in the other dimensions. Also, the large spread of samples along the diagonal line indicated that the samples used for PLS modeling covered a broad range in chemical/sensory properties, which is ideally preferred for PLS model calibration.
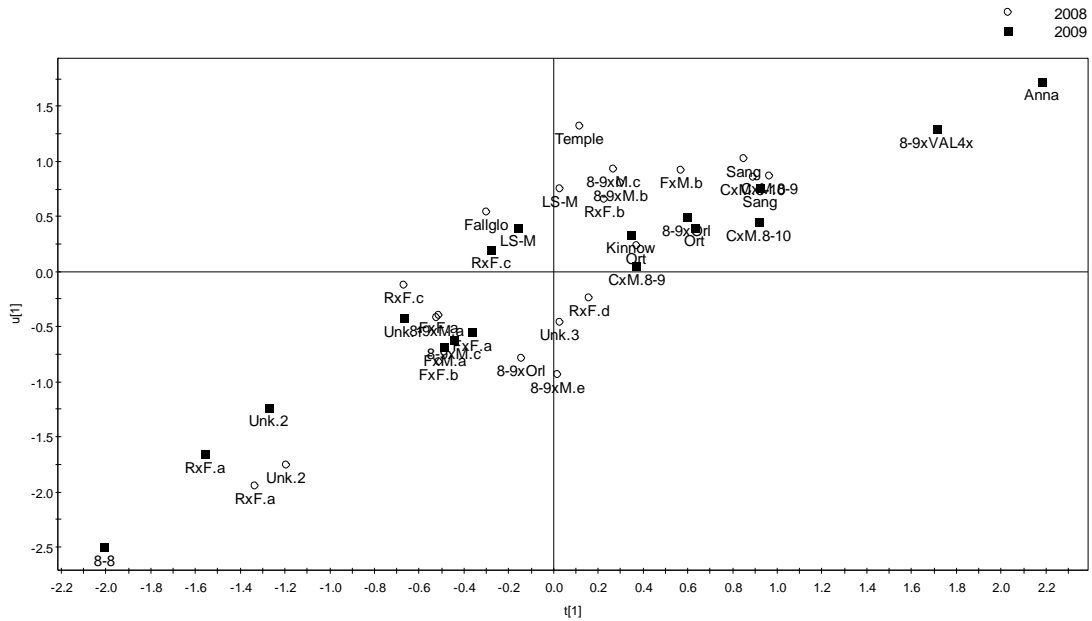

Figure 6.3: Scatter plot of the X-scores (t) and the Y-scores (u), with the first PLS component, summarizing the correlation structure between instrumental and sensory data for the calibration set (38 juices).

Using the PLS model developed with the calibration sample set, the sensory descriptor scores of the samples in the prediction set were calculated (predicted Y-data set) based on the instrumental data collected. For each modeled response, the predicted scores (given by the model) and the observed values (given by descriptive sensory analysis) were plotted against each other. The coefficient of determination r^2 for the regression line was used as a measure of the predictive power of the model using the prediction set (also referred to as Q^2Y_{ext}). Figure 6.4 is given as an example for the attributes that were well predicted (e.g., sour) and poorly predicted (e.g., floral); we added on the same plot both calibration and prediction samples.

We observed that samples from the prediction set were homogeneously spread across the ideal regression line for most of the sensory attributes (except floral and bitter), indicating that they spanned the range of responses fully and therefore were appropriate as a prediction set. The results (Table 6.3) showed good agreement between the predicted and observed scores for the attributes grapefruit, orange, sour, fruity non-citrus, and green/fresh ($r^2 > 0.75$), and moderate for pumpkin/fatty and sweet ($0.45 > r^2 > 0.75$). The best correlations were for grapefruit and orange, and the poorest were found for tangerine, bitter and floral. Also, the prediction errors of the prediction set (RMSEP) were relatively comparable to the errors of the calibration set (RMSEE) except for sulfury and floral. In conclusion, there was a good agreement overall between results from the calibration and prediction sets (Table 6.3), confirming that the developed PLS model did not over-fit the data. The PLS model was able to predict well the attributes grapefruit, sour, fruity non-citrus, orange and pumpkin/fatty, and to some extent green/fresh and sweet. The predictions for the other descriptors (tangerine, sulfury, bitter and floral) were poorer. Of note, the sensory data for these latter descriptors were more variable: even after training, panelists had difficulties to come to an agreement for these attributes.
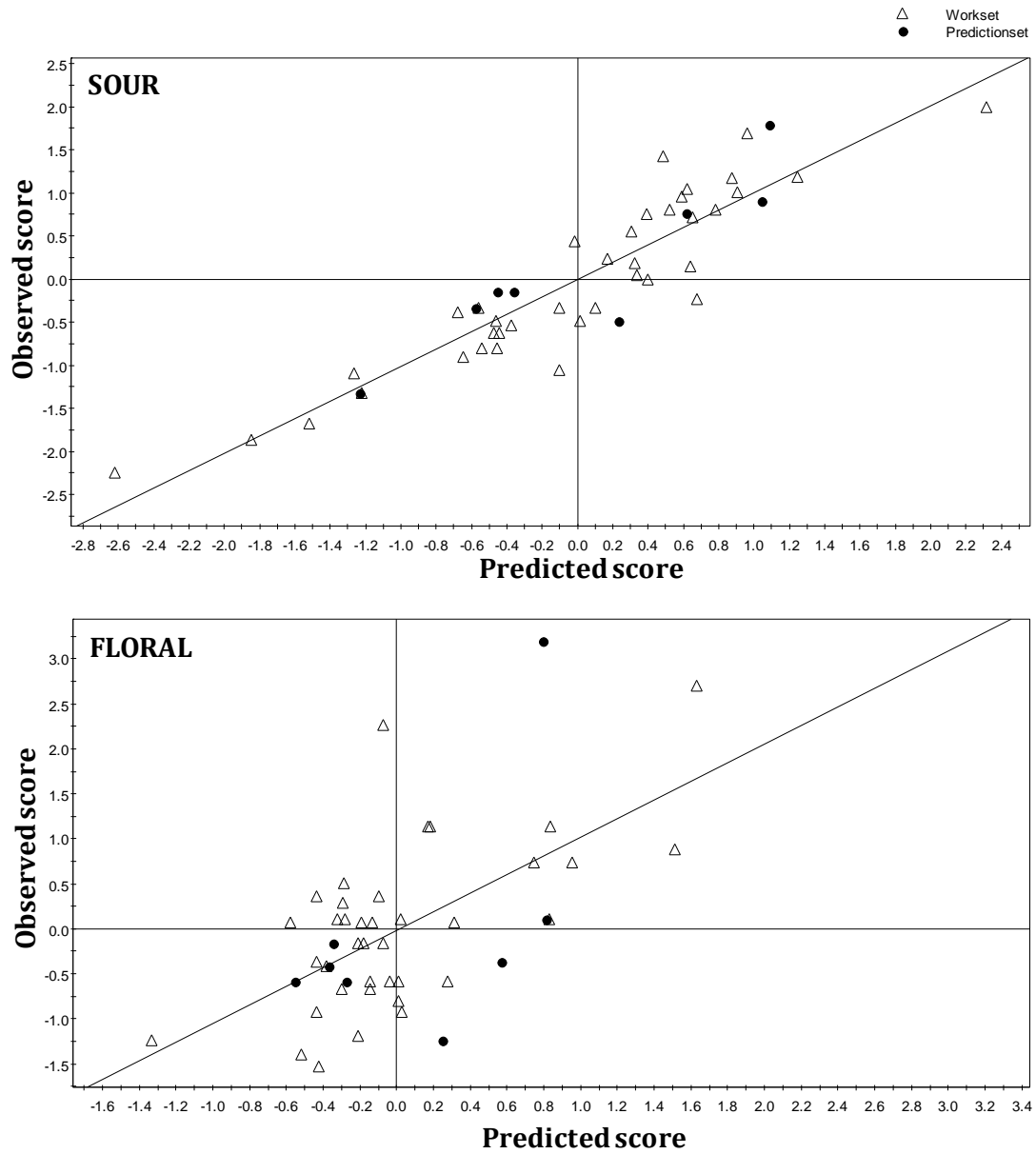
Figure 6.4: Relationship between observed and predicted scores for the descriptors sour (upper) and floral (lower) in mandarin juices by PLS modeling.

6.3.4. Tentative identification of flavor markers

Although this was not a primary objective of this study, we examined some of the variables that most influenced the prediction model for a few of the sensory descriptors in an attempt to identify compounds that are related to a specific trait ("flavor markers"). This introduces another application of using a non-targeted approach: the identification of flavor markers. Of course, the sensory relevance of such compounds will need to be validated by subsequent sensory studies (see Baldwin et al. 2008 for an example of how one may proceed).

One advantage of such strategy is that only compounds associated with specific flavor characteristics (as selected by the multivariate analyses) are identified, thereby saving time and energy. However, as mentioned earlier, the fact that all responses were modeled together in this study (instead of separate models) might complicate this task. In future studies, it might be more appropriate to develop separate models if identification of flavor markers is the goal, and/or use different experimental design and multivariate tools (e.g., cluster analysis, network analysis).

Table 6.4: Examples of "flavor markers" tentatively identified.

Sensory attribute	Instrumental variables	Identification (Δppm)	Method for identification
Orange	GC_3.5172_88	Ethyl butanoate	
	GC_22.0816_105	γ-Selinene	Mass spectral library and retention indices
	GC_22.1402_105	Valencene	
	GC_22.1915_189	α-Selinene	
	GC_22.3456_122	α-Panasinsene	
Sour	An_6.2338_191.0167	Citric acid (Δppm =12)	Database search , MS/MS, and confirmation with standard
Pumpkin /fatty	Ap_4.2167_138.0541	Trigonelline (Δppm =6)	Database search , MS/MS, and confirmation with standard

To select the variables of interest, we inspected the regression coefficient plots which provide information on the amplitude and direction of the relationship between variables and a specific response (therefore, there is one coefficient plot per response). We focused only on variables having a high positive correlation with a given attribute. Only few examples are reported here (Table 6.4). As noted earlier, this approach is not without limitations, one being that the correlations established are not causative. The correlations (chemical-sensory) established by the predictive models will require subsequent studies (focused on those specific compounds) to test the causal nature of the relationship. However, it provides ground for future targeted studies and the compounds to focus on.

As anticipated, most of the compounds selected are already identified as "sensory active" in the literature. For instance, citric acid which was highly ranked in the regression coefficient plots correlated to the prediction of sour. Citric acid is one of the main organic acid in mandarin juices and its quantity reported ranges between 0.86 and 1.22% (Nagy and Shaw 1990). Sugar/acid ratio is correlated to fruit maturity, and indirectly to flavor development since volatiles (secondary plant metabolites) are generated during fruit ripening (Ranganna et al. 1983a). Therefore, it is an important quality indicator for flavor and is used to grade the legal maturity of citrus fruits in Florida. Ahmed et al. (1978c) also reported that the sugar-acid ratio affected the sensory ratings of orange juice but a clear explanation of the reasons (e.g., interactions with the matrix or aroma compounds) was lacking. The contribution of sugars and acids to flavor have been examined in other food systems also. For instance, Baldwin et al. (2008) found that the intensity and profile of perceived flavor in tomato puree depended on sugar and acid levels.

As well, markers associated with the descriptor "orange" have been previously reported in citrus juices. Ethyl butanoate has a fruity/pineapple aroma and is sensitive to thermal processing, therefore it is associated with the "fresh orange juice" character (Perez-Cacho and Rouseff 2008b; Perez-Cacho and Rouseff 2008a). Even though ethyl butanoate is present in very low concentrations (ppm to ppb), its low sensory threshold (1.71 and 1 μg/L for orthonasal and retronasal detection thresholds in reconstituted pump-out matrix, respectively) make it one of the most intense odorants in orange juice (Plotto et al. 2008). Of note, the type of cultivar and fruit maturity have been reported to significantly impact

the levels of this compound in orange fruits (Buettner and Schieberle 2001; Arena et al. 2006). The other volatile compounds are sesquiterpenes and constitute one of the volatile groups present in the greatest quantity in orange fruits. They have been previously reported to be in higher concentrations in mandarin fruits with sweet orange (*Citrus sinensis* L. Osb.) in their lineage (Kerbiriou et al. 2007a; Miyazaki et al. 2011). However, they have low aroma activity due to high sensory thresholds (Plotto et al. 2004). It is possible that these compounds are correlated with other markers that were not identified here in this analysis (for example, some nor-isoprenoids that are very potent but are hardly detected by GC-MS headspace techniques). Valencene has been associated with aroma quality but not for its own sensory properties but rather as an indicator of fruit maturity, which is linked to increased orange flavor quality (Shimada et al. 2004; Elston et al. 2005).

For the prediction of the descriptor "Pumpkin/fatty", trigonelline was identified as one influent compound (Figure 6.5). Trigonelline is a plant alkaloid derived from nicotinic acid (Holman and De Lange 1950). Its presence has been reported in Satsuma mandarin juice (Yamada 1977; Zhang et al. 2011). Trigonelline is in high contents in green coffee and is an important flavor precursor upon thermal processing (Flament 2002). Besides its slightly bitter properties (a quarter the strength that of caffeine) (Flament 2002), we were not able to find any reports on other sensory properties.
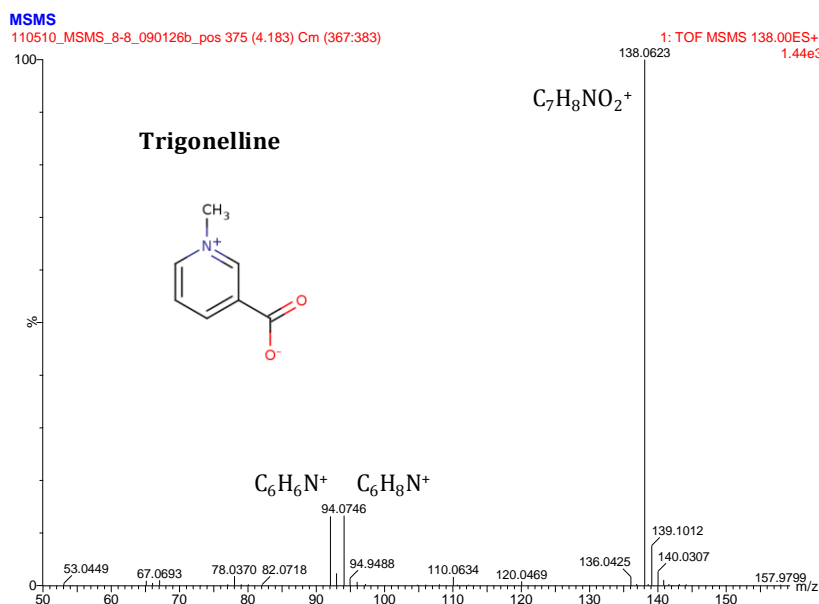
Figure 6.5: Identification of the marker associated with pumpkin/fatty flavor by ESI-MS/MS approach.

Even though we did not perform quantification for the selected markers, we examined their relative abundance across samples (Figure 6.6). Two types of flavor markers were distinguished. The first type included "univariate markers", which are compounds that can potentially be used on their own to predict a particular sensory attribute. For instance, citric acid (considered as a univariate marker) for which a strong linear relationship between instrumental signal intensity and sensory score was observed. Yet, one needs to keep in mind that the relationship between concentration and sensory perception is not linear across a large range. As well, the amplitude of a chemical compound does not relate systematically to its importance in terms of sensory meaning. The second type of flavor markers can be refered to as "multivariate", implying that they can not be used alone for predictions. Such compounds may play a key role in differentiating samples with high- and low- scores for a given attribute (e.g., valencene), but do not exhibit a continuous relationship between sensory and instrumental intensity. This is not unexepected given the fact that the perception of flavor is multivariate in nature and involves inputs from multiple chemical stimuli.
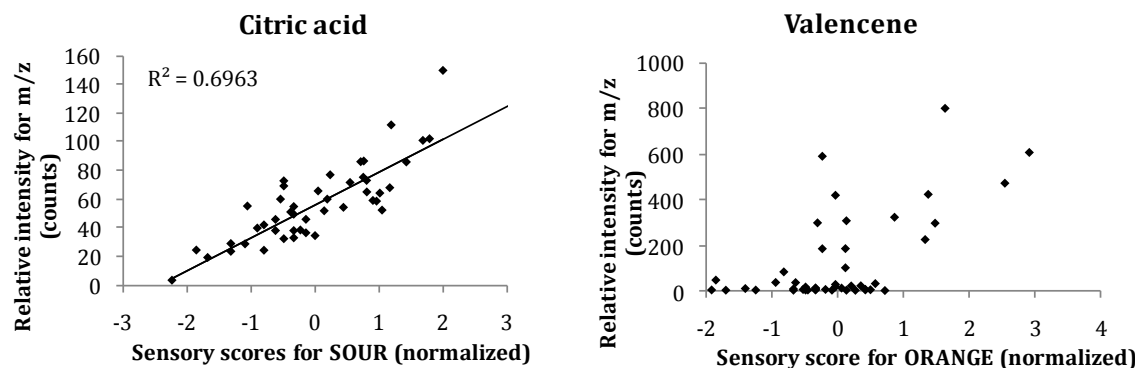
162

Figure 6.6: Relationships between marker's intensity and sensory score (normalized units).

6.4. CONCLUSIONS

This study investigated the feasibility of using flavoromics to predict the flavor quality of mandarin juice. This approach used a non-targeted instrumental methodology so that the chemical information included in the predictive models was comprehensive. The chemical composition of 46 mandarin juices from different cultivars and hybrids was characterized by GC- and UHPLC-TOF-MS, and their sensory profile established by a trained panel. The fruits used in this study were selected by the breeder to represent a range of sensory profiles, which was confirmed with a preliminary examination of the sensory data by PCA. Compositional variations across mandarin juice samples and their sensory profile were correlated using partial least squares regression, from which predictive models of sensory quality were developed.

Initially, the entire sample set was used to compare the performances of different PLS models. The explanatory and predictive performances were improved when combining all instrumental data into one single data set as opposed to individual data sets. The best PLS model was obtained with mid-level data fusion (model 7), for which a preliminary variable selection was done before merging the different instrumental data sets. The predictive power of the selected model was subsequently tested using calibration and prediction sample sets (38 and 8 juices, respectively).

A fairly robust model was obtained as both cross-validation (for calibration set) and validation by external predictions gave comparable estimates of the predictive ability of the model for most of the attributes; and a strong relationship between instrumental and sensory measurements was observed. The resulting model showed that prediction of sensory scores was possible to a certain extent for the majority of the sensory descriptors thereby demonstrating the applicability of flavoromics. The best predictions were obtained for the attributes grapefruit, sour, fruity non-citrus, orange and pumpkin/fatty ($0.5 < Q^2Y < 0.7$), whereas tangerine, bitter and floral yielded the poorest ones ($Q^2Y < 0.35$).

The approach developed in this research might be a valuable tool in plant breeding programs to predict the sensory properties of newly developed hybrids, allowing faster screening. As well, other insights may be drawn from such a research approach, for instance by comparing the different hybrid populations against each other and identifying markers that contribute to their separation. This kind of information may assist in understanding the underlying genetic basis of the differences in sensory quality observed.

Chapter 7 : Conclusions and remarks

The ability to predict flavor or sensory quality of food based on instrumental-sensory correlations has been a long term goal for food flavor scientists. However, the task remains a challenge mainly because of the multi-dimensionality of flavor perception. This dissertation work addressed some of the issues neglected in the past by introducing flavoromics, a non-targeted chemometric research methodology that links chemical compositional variations to flavor perception.

7.1. VALUE OF FLAVOROMICS AND APPLICATIONS

Flavoromics is proposed here as a complementary tool for flavor research that offers several advantages over conventional methods. It is a comprehensive strategy that considers "all" (ideally) chemical signals as inputs of flavor perception, instead of focusing only on compounds already known to influence flavor quality. This dissertation work demonstrated the feasibility of flavoromics to predict flavor with a set of forty-six mandarin juices (differences in hybrids and harvest years).

Methods using complementary instrumental platforms (GC and UHPLC) were developed and used to gain the most comprehensive overview of sample chemical composition. The relationships between the presence and relative intensities of chemical compounds (given as mass spectral signals) and flavor attribute scores (provided by means of descriptive sensory analysis) were modeled by means of partial least squares regression, and then used for predictions of sensory scores. This study showed improved explanatory and predictive performances of the models when all instrumental information was combined into one single data set as opposed to individual data sets, illustrating the usefulness of a holistic approach to get more accurate flavor predictions. A fairly robust predictive model was developed for the majority of the sensory descriptors and did not require knowing the chemical identity of the instrumental variables measured.

While this methodology is not fully developed, it is evident from this work that there is a large potential for the use of flavoromics in flavor research. The two major areas where flavoromics will be a valuable addition are sensory prediction and sensory understanding. Given its non-targeted nature, flavoromics is appropriate for the study of complex food systems as it accounts for the presence of all compounds and their possible interactions, without the need for chemical identification. Moreover, better predictions of flavor are achieved by including more inputs from chemical compounds. As a data-driven approach, flavoromics also provides the opportunity to more quickly screen for compounds that are related to flavor ("markers") and only compounds linked to perception are identified in subsequent experiments. Other applications of flavoromics are expected in the future: for instance, to follow the kinetics of flavor development (flavor genesis pathways), or to pinpoint markers of sensory quality by comparative analysis of food chemical compositions.

7.2. LIMITATIONS/ CHALLENGES OF THE RESEARCH

This work opens new opportunities for flavor investigation, however, there are still many analytical and data processing challenges as well as room for improvements in the methodology.

7.2.1. Instrumental

One of the first limitations is related to the comprehensive characterization of the sample. No matter how careful one is in choosing sample preparation techniques or how sophisticated an instrumental platform is, there will be always some (minimum) bias in the compounds being extracted and analyzed. Currently, it is not possible to have a single analytical method that has virtually no bias (de Vos et al. 2006). Therefore, it is still a requisite to use multiple and complementary instrumental platforms and detection techniques (which implies substantial additional work) to gain a comprehensive overview. Limitations of MS-based approaches are discussed in more detail in Scalbert et al. (2009).

166

Another limitation is the quantification of compounds. Due to matrix interferences, or differences in instrument response between compounds, or limitations in the linear range measured, it is not possible to obtain an absolute quantitative analysis of all compounds in the sample. Also, it is not possible to do calibration curves because of the large number of compounds recorded and because their chemical identity is not always known. Therefore, semi-quantitative analyses are usually the only option. (Absolute quantification of a given compound may be conducted in later stages of the research). Normalization of the data is generally used to express the relative amounts of each compound in the sample; however, there is a lot of disagreement among published works regarding the method to apply for normalization. For instance, whether to normalize the data relative to total count or to a single or multiple internal/external standards (Katajamaa and Oresic 2007). This brings up an issue we faced in this dissertation work. We initially decided to normalize the data relative to an external standard (i.e., added after sample extraction/preparation) but we realized that its signal intensity varied depending on several factors related to suppression of the signal (matrix effect) and differences in MS ionization efficiency over time for UHPLC analyses, and due to differences in sample-to-fiber partition coefficient, differences between successive SPME fibers (replacement) and fiber aging for GC analyses. These problems have been mentioned in the literature (de Vos et al. 2007; Gika et al. 2007; Parab et al. 2009).

The inherent biological variability requires that a large number of replicates should be analyzed, which clearly affects throughput. Fiehn et al. (2000) estimated the biological variability for a series of primary metabolites from 18 leaf extracts of *Arabidopsis thaliana* measured by GC/MS ranged from 17 to 56 % s.d. Recommendations on how to minimize biological and analytical variations can be found in Scalbert et al. (2009).

Finally, the identification of compounds remains a challenge especially for non-volatile compounds. Even though the number of searchable databases is growing (see Werner et al. 2008 for a list), it is still difficult to find potential leads using only the elemental composition. Tandem mass-spectrometry MS/MS or NMR strategies are required for further structural elucidation.

7.2.2. Data pre-processing and chemometrics

One hurdle of non-targeted approaches is that they require more sophisticated data processing techniques, including data extraction from the instrumental measurements and subsequent mathematical modeling.

7.2.2.1. Data pre-processing

It is evident from this study and other published works that comprehensive approaches depend strongly on the computational and statistic tools available. Indeed, the massive amounts of data generated by "omics" strategies prevent anyone from extracting conclusions of biological relevance without software assistance. Advances in dedicated software (free or commercial) to extract and automatically process the raw data are still required even though they are becoming more available and powerful.

At this time, common problems of automated data extraction tools include: chromatogram misalignment, missed peaks (e.g., intensity below threshold) and integration errors. One way to minimize these errors is to use a reference sample which is a mixture of several samples from the study (Lommen 2011). In this manner, the analysis of the reference standard is used as a starting point for chromatogram alignment and peak picking, and also allows for checking stability of the instrumentation. Unfortunately, we did not consider early enough in our study the use of such samples.

The lack of tools to minimize data redundancy (i.e., multiple variables from the same compound) is also an issue; there is a lack of software that deals with this and their application is limited (Werner et al. 2008). For example, it would be very valuable to filter out redundant signals related to in-source fragmentation, isotopes, dimers or adducts. Similarly, to retain only one representative m/z variable for EI-GC-MS data where multiple fragments occur for a given compound, or in studies combining data from multiple instruments or from different detection methods (i.e., same compound reported multiple times). Currently, research groups use their own "in-house" developed tools for performing such task, and therefore, they are not publicly available. The presence of redundant signals

slows down the identification step by increasing the number of fragments to be searched and also "artificially" inflates the contribution of a given compound.

One last point to mention is the use of appropriate data pre-processing steps, which have been mentioned to significantly influence the results (e.g., the selection of markers) (van den Berg et al. 2006). Appropriate scaling, normalization and eventually transformations are often needed; as well weighting and/or scaling factors may be used when combining blocks of data to adjust for differences in numerical range and/or size between blocks (Smilde et al. 2005; Eriksson et al. 2006b). There are currently no standards for data processing.

7.2.2.2. Chemometrics

One important point we would like to emphasize here is that chemometric tools (i.e., PLSR) aid in the interpretation/use of the data but do not provide final conclusions. In this study, PLSR was used to obtain a mathematical model relating variations in chemical data with sensory perception. However, the correlations established did not imply causative relationships. The results (correlations) always need to be checked through careful targeted studies to test their sensory relevance. Identifications of key stimuli will ultimately need to be done, judgments made about each stimulus identified as to the likely hood of it being causative and much sensory work done to determine if a selected chemical marker may truly contribute to sensory perception.

Of importance, the models developed in this study are valid only for the laboratory where data was collected and limited to the type of samples analyzed. Indeed, variation specific to the instruments is embedded in the dataset used to develop the models. Therefore, if the same sample was analyzed on a different instrument the predictions may be erroneous. Further work would need to be done before applying the predictive model to other types of juices (e.g., juice) or in another location. However, the goal of this work was to examine the potential of using a non-targeted multi-instrumental method to predict flavor.

One last limit of this study related to chemometric processing is that it considered only linear models. It is possible that other multivariate data analyses, such as artificial neural networks and heatmaps, may have provided a better description of the instrument-sensory

relationships because of their ability to solve non-linear relationships (Sugimoto et al. 2010). However, the interpretation of the findings might be more difficult.

7.2.3. Sensory

Finally, there are few points to consider regarding the sensory aspects of our approach. First, our proposed methodology did not account for other aspects potentially affecting the flavor scores given by the panelists such as the physiological, environmental and emotional factors (e.g., previous exposure to the food). Even though from a strict mechanistic point of view flavor perception results from chemical stimuli, it is difficult to separate "liking" from perception in judges' ratings. As well, the overall sensory quality of the sample (i.e., flavor profile) might have affected the perception of the intensity of the individual sensory attributes as reported by Lindinger et al. (2008). Similarly, if a sensory descriptor was missing to describe a particular flavor characteristic the panelist may have been tempted to transfer its perception onto other descriptors ("dumping effect"), thereby biasing the "true" scores. These facts may account for some of the differences between predicted and observed sensory scores.

Second, the processing of the sensory scores could have been different. In this dissertation work, we used the sensory mean scores (i.e., average of the judges' scores for each sensory attribute) to build the predictive models. However, it is well accepted that judges (even trained) use differently scales, either in the level used or in the range given between products. As well, individual sensitivities to some stimuli may result in large disagreements between judges (e.g., bitterness). Some "scaling methods" such as generalized procrustes analysis (GPA) (Chung et al. 2003) and isotropic scaling (Kunert and Qannari 1999) have been proposed to adjust for differences in ratings among judges but were not considered in this work.

Third and last, the sensory ratings used to build the predictive models were obtained using a trained sensory panel and therefore cannot be extrapolated to consumers. It has been acknowledged that there are differences in scores generated between a trained and untrained panel (Roberts and Vickers 1994). However, the direction of the correlation

(positive or negative) between sensory scores and chemical compounds should remain similar, independently of trained and untrained panel (Moskowitz 1996). Nonetheless, the models were designed to predict the flavor in terms of product characteristics (objective) and NOT in terms of consumer preference (very subjective to a lot of environmental factors). To this end, trained assessors are –in theory– more accurate and sensitive than untrained ones, which guarantied us better precision for the model development.

References

Abegaz EG, Tandon KS, Scott JW, Baldwin EA & Shewfelt RL. 2004. Partitioning taste from aromatic flavor notes of fresh tomato (*lycopersicon esculentum, mill*) to develop predictive models as a function of volatile and nonvolatile components. Postharvest biology and technology 34(3):227-235.

Ahmed EM, Dennison RA, Dougherty RH & Shaw PE. 1978a. Effect of nonvolatile orange juice components, acid, sugar, and pectin on the flavor threshold of d-limonene in water. Journal of Agricultural and Food Chemistry 26(1):192-194.

Ahmed EM, Dennison RA, Dougherty RH & Shaw PE. 1978b. Flavor and odor thresholds in water of selected orange juice components. Journal of Agricultural and Food Chemistry 26(1):187-191.

Ahmed EM, Dennison RA & Shaw PE. 1978c. Effect of selected oil and essence volatile components on flavor quality of pumpout orange juice. Journal of Agricultural and Food Chemistry 26(2):368-372.

Aishima T & Nakai S. 1991. Chemometrics in flavor research. Food reviews international 7(1):33 - 101.

Andersson CA. 2000. Exploratory multivariate data analysis with applications in food technology. Departement of Dairy and Food Science*.* Frederiksberg, Denmark: The Royal Veterinary and Agricultural University. p. 202.

Araujo AS, da Rocha LL, Tomazela DM, Sawaya ACHF, Almeida RR, Catharino RR & Eberlin MN. 2005. Electrospray ionization mass spectrometry fingerprinting of beer. The Analyst 130(6):884-889.

Arena E, Guarrera N, Campisi S & Nicolosi Asmundo C. 2006. Comparison of odour active compounds detected by gas-chromatography-olfactometry between hand-squeezed juices from different orange varieties. Food chemistry 98(1):59-63.

Attaway JA, Barron RW, Blair JG, Buslig BS, Carter RD, Dougherty MH, Fellers PJ, Fisher JF, Hill EC, Huggart RL, Maraulja MD, Petrus DR & Ting SV. 1972. Some new analytical indicators of processed orange juice quality, 1971-72. Florida State Horticultural Society 85:192-203.

Attaway JA & Carter RD. 1971. Some new analytical indicators of processed orange juice quality. Florida State Horticultural Society 84:200-205.

Aznar M, López R, Cacho J & Ferreira V. 2003. Prediction of aged red wine aroma properties from aroma chemical composition. Partial least squares regression models. Journal of Agricultural and Food Chemistry 51(9):2700-2707.

Bakker J, Brown W, Hills B, Boudaud N, Wilson C & Harrison M. 1996. Effect of the food matrix on flavor release and perception. In: Taylor, A. J. & Mottram, D. S., editors. *Flavor science: Recent developments.* Special publication ed. Cambridge, UK: Royal Society of Chemistry. p. 369.

Baldwin EA, Goodner K & Plotto A. 2008. Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors. Journal of food science 73(6):S294-S307.

Bell GA. 1996. Molecular mechanisms of olfactory perception: Their potential for future technologies. Trends in Food Science & Technology 7(12):425-431.

Biasioli F, Gasperi F, Aprea E, Endrizzi I, Valeria F, Marini F, Mott D & Mark TD. 2006. Correlation of ptr-ms spectral fingerprints with sensory characterisation of flavour and odour profile of "Trentingrana" Cheese. Food Quality and Preference 17(1-2):63-75.

Boland AB, Delahunty CM & van Ruth SM. 2006. Influence of the texture of gelatin gels and pectin gels on strawberry flavour release and perception. Food chemistry 96(3):452.

Braddock RJ, Goodrich RM & Bryan CR. 2004. Variables affecting solid-phase microextraction headspace analysis of orange juice volatiles. In: Deibler, K. D. & Delwiche, J., editors. *Handbook of flavor characterization. Sensory analysis, chemistry, and physiology.* New York, USA: Marcel Dekker, Inc. p. 423-436.

Brás LP, Bernardino SA, Lopes JA & Menezes JC. 2005. Multiblock pls as an approach to compare and combine nir and mir spectra in calibrations of soybean flour. Chemometrics and Intelligent Laboratory Systems 75(1):91-99.

Brat P, Rega B, Alter P, Reynes M & Brillouet J-M. 2003. Distribution of volatile compounds in the pulp, cloud, and serum of freshly squeezed orange juice. Journal of Agricultural and Food Chemistry 51(11):3442-3447.

Breslin PAS. 1996. Interactions among salty, sour and bitter compounds. Trends in Food Science & Technology 7(12):390-399.

Brown M, Dunn WB, Dobson P, Patel Y, Winder CL, Francis-McIntyre S, Begley P, Carroll K, Broadhurst D, Tseng A, Swainston N, Spasic I, Goodacre R & Kell DB. 2009. Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. Analyst 134(7):1322-1332.

Brown WE & Wilson CE. 1996. Individuality of flavour perceptions - the influence of mastication patterns. In: Taylor, A. J. & Mottram, D. S., editors. *Flavour science : Recent developments.* Special publication ed. Cambridge, UK: Royal Society of Chemistry. p. 451-455.

Bruce SJ, Tavazzi I, Parisod V, Rezzi S, Kochhar S & Guy PA. 2009. Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. Analytical Chemistry 81(9):3285-3296.

Buettner A & Schieberle P. 2001. Evaluation of aroma differences between hand-squeezed juices from valencia late and navel oranges by quantitation of key odorants and flavor reconstitution experiments. Journal of Agricultural and Food Chemistry 49(5):2387-2394.

Burgard DR. 1995. Chemical and sensory correlations for orange juice. In: Rouseff, R. L. & Leahy, M. M., editors. *Fruit flavors. Biogenesis, characterization and authentification.*: American Chemical Society. p. 21-32.

Buttery RG. 1999. Flavor chemistry and odor thresholds. In: Teranishi, R., Wick, E. L. & Hornstein, I., editors. *Flavor chemistry. Thirty years of progress.* New York: Kluwer Academic/Plenum Publishers. p. 353-365.

Cabezas L, González-Viñas M, Ballesteros C & Martín-Alvarez P. 2006. Application of partial least squares regression to predict sensory attributes of artisanal and industrial manchego cheeses. European food research and technology 222(3):223-228.

Carbonell L, Bayarri S, Navarro JL, Carbonell I & Izquierdo L. 2009. Sensory profile and acceptability of juices from mandarin varieties and hybrids. Food science and technology international 15(4):375-385.

Carbonell L, Izquierdo L & Carbonell I. 2007. Sensory analysis of spanish mandarin juices. Selection of attributes and panel performance. Food Quality and Preference 18(2):329-341.

Carbonell L, Navarro JL, Izquierdo L & Carbonell I. 2008. Sensory profile of mandarin chilled juices and consumers' acceptability. Food science and technology international 14(2):157-166.

Carter RD & Buslig BS. 1975. Statistical inferences on processed orange juice quality from the 1972-73 and 1973-1974 juice definition program. Florida State Horticultural Society 88:358-370.

Charve J, Chen C, Hegeman AD & Reineccius GA. 2011. Evaluation of instrumental methods for the untargeted analysis of chemical stimuli of orange juice flavor. Flavour and fragrance journal *In press*.

Chung S-J, Heymann H & Grün IU. 2003. Application of gpa and plsr in correlating sensory and chemical data sets. Food Quality and Preference 14(5-6):485-495.

Cozzolino D, Smyth HE, Cynkar W, Janik L, Dambergs RG & Gishen M. 2008. Use of direct headspace-mass spectrometry coupled with chemometrics to predict aroma properties in australian riesling wine. Anal. Chim. Acta 621(1):2-7.

Cubbon S, Bradbury T, Wilson J & Thomas-Oates J. 2007. Hydrophilic interaction chromatography for mass spectrometric metabonomic studies of urine. Analytical Chemistry 79(23):8911-8918.

Dalton P, Doolittle N, Nagata H & Breslin PAS. 2000. The merging of the senses: Integration of subthreshold taste and smell. Nature Neuroscience 3(5):431-432.

Davidson JM, Linforth RST, Hollowood TA & Taylor AJ. 1999. Effect of sucrose on the perceived flavor intensity of chewing gum. Journal of Agricultural and Food Chemistry 47(10):4336-4340.

de Vos CHR, Borem FM, Bouwmeester HJ & Bino RJ. 2006. Untargeted metabolomics as a novel tool in coffee research. Colloq. Sci. Int. Cafe 21st:125-132.

de Vos CHR, Moco S, Lommen A, Keurentjes JJB, Bino RJ & Hall RD. 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nature Protocols 2(4):778-791.

de Vos CHR, Tikunov Y, Bovy AG & Hall RD. 2008. Flavour metabolomics: Holistic versus targeted approaches in flavour research. In: Blank, I., Wust, M. & Yeretzian, C., editors. Expression of Multidisciplinary Flavour Science. Proceedings of the 12th Weurman Symposium. Interlaken, Switzerland: Zürcher Hochschule für Angewandte and Institut Für Chemie und Biologische Chemie. p. 573-580.

de Wijk RA, Rasing F & Wilkinson C. 2003. Texture of semi-solids: Sensory flavor-texture interactions for custard desserts. Journal of Texture Studies 34:131-146.

Dea S, Plotto A, Manthey J, Baldwin E & McCollum G. 2010. Insights regarding sensory evaluation of bitterness development in citrus juice. HortScience 45(8):S307.

Deibler KD, Lavin EH, Linforth RST, Taylor AJ & Acree TE. 2001. Verification of a mouth simulator by in vivo measurements. Journal of Agricultural and Food Chemistry 49(3):1388-1393.

Delwiche J. 1996. Are there basic tastes? Trends in Food Science & Technology 7(12):411-415.

Dettmer K, Aronov PA & Hammock BD. 2007. Mass spectrometry-based metabolomics. Mass Spectrom. Rev. 26(1):51-78.

Downard K. 2004. Mass spectrometry: A foundation course. Cambrigde, UK: The Royal Society of Chemistry.

Dravnieks A. 1976. Approaches to subjective/objective correlations in flavor. ASTM symposium: Correlating sensory objective measurements - new methods for answering old problems. Philadelphia, PA: American Society for Testing and Materials. p. 5-25.

Druaux C, Le Thanh M, Seuvre A-M & Voilley A. 1998. Application of headspace analysis to the study of aroma compounds-lipids interactions. Journal of the American Oil Chemists' Society 75(2):127-130.

Dunn WB & Ellis DI. 2005. Metabolomics: Current analytical platforms and methodologies. Trends in Analytical Chemistry 24(4):285-294.

Elston A, Lin J & Rouseff RL. 2005. Determination of the role of valencene in orange oil as a direct contributor to aroma quality. Flavour and fragrance journal 20(4):381-386.

Elston AM. 2005. Developing an objective flavor measurement tool for orange juice. Phd thesis. Institute of Food and Agricultural Science, Citrus Research and Education Center. Lake Alfred: University of Florida. p. 135.

Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C & Wold S. 2006a. Basic concepts and principles of projections. *Multi- and megavariate data analysis. Part i. Basic principles and applications.* 2nd revised and enlarged ed. Umea, Sweden: Umetrics AB. p. 21-38.

Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C & Wold S. 2006b. Centering and scaling. *Multi- and megavariate data analysis. Part i. Basic principles and applications.* 2nd revised and enlarged ed. Umea, Sweden: Umetrics AB. p. 207-219.

Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C & Wold S. 2006c. Multivariate calibration. *Multi- and megavariate data analysis. Part i. Basic principles and applications.* 2nd revised and enlarged ed. Umea, Sweden: Umetrics AB. p. 121-145.

Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C & Wold S. 2006d. Pca. *Multi- and megavariate data analysis. Part i. Basic principles and applications.* 2nd revised and enlarged ed. Umea, Sweden: Umetrics AB. p. 39-62.

Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C & Wold S. 2006e. Pls. *Multi- and megavariate data analysis. Part i. Basic principles and applications.* 2nd revised and enlarged ed. Umea, Sweden: Umetrics AB. p. 63-102, 181-189.

Faber NM & Rajkó R. 2007. How to avoid over-fitting in multivariate calibration - the conventional validation approach and an alternative. Anal. Chim. Acta 595(1-2):98-106.

Fake C. 2004. A mandarin by any other name. Publication # 31-111. http://ceplacer.ucdavis.edu/files/58523.pdf ed.: University of California, Cooperative extension.

Fanciullino A-L, Dhuique-Mayer C, Luro F, Casanova J, Morillon R & Ollitrault P. 2006. Carotenoid diversity in cultivated citrus is highly influenced by genetic factors. Journal of Agricultural and Food Chemistry 54(12):4397-4406.

Fiehn O. 2002. Metabolomics - the link between genotypes and phenotypes. Plant Molecular Biology 48:155-171.

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN & Willmitzer L. 2000. Metabolite profiling for plant functional genomics. Nature Biotechnology 18(11):1157-1161.

Fishman GM & Bednyak RB. 1974. Optimum content of volatile oil in tangerine juice and its control. Konservnaya i Ovoshchesushil'naya Promyshlennost' in Food Science and Technology Abstracts (9):39-40.

Flament I. 2002. Coffee flavour chemistry. Chichester, UK: John Wiley & Sons, Ltd.

Forshed J, Idborg H & Jacobsson SP. 2007. Evaluation of different techniques for data fusion of lc/ms and 1h-nmr. Chemometrics and Intelligent Laboratory Systems 85(1):102-109.

Frank IE & Kowalski BR. 1982. Chemometrics. Analytical Chemistry 54(5):232R-243R.

Gabrielsson J, Jonsson H, Airiau C, Schmidt B, Escott R & Trygg J. 2006. Opls methodology for analysis of pre-processing effects on spectroscopic data. Chemometrics and Intelligent Laboratory Systems 84(1-2):153-158.

Gattuso G, Barreca D, Gargiulli C, Leuzzi U & Caristi C. 2007. Flavonoid composition of citrus juices. Molecules 12(8):1641-1673.

Gika HG, Theodoridis GA & Wilson ID. 2008. Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography tof-ms for metabonomic analysis of zucker rat urine. Journal of separation science 31(9):1598-1608.

Gika HG, Theodoridis GA, Wingate JE & Wilson ID. 2007. Within-day reproducibility of an hplc-ms-based method for metabonomic analysis: Application to human urine. J. Proteome Res. 6(8):3291-3303.

Gil-Izquierdo A, Gil MI, Ferreres F & Tomas-Barberan FA. 2001. In vitro availability of flavonoids and other phenolics in orange juice. Journal of Agricultural and Food Chemistry 49(2):1035-1041.

Gmitter FG, Jr. , Grosser JW, Castle WS & Moore GA. 2007. A comprehensive citrus genetics improvement program. In: Khan, I. H., editor). *Citrus genetics, breeding and biotechnology.* Wallingford, UK: CAB International. p. 9-18.

Grata E, Boccard J, Guillarme D, Glauser G, Carrupt P-A, Farmer EE, Wolfender J-L & Rudaz S. 2008. Uplc-tof-ms for plant metabolomics: A sequential approach for wound marker analysis in *arabidopsis thaliana*. Journal of Chromatography B 871(2):261-270.

Green BG. 1996. Chemesthesis: Pungency as a component of flavor. Trends in Food Science & Technology 7(12):415-420.

Grove WM & Andreasen NC. 1982. Simultaneous tests of many hypotheses in exploratory research. Jounal of Nervous and Mental Disease 170(1):3-8.

Guinard JX & Marty C. 1995. Time-intensity measurement of flavor release from a model gel system: Effect of gelling agent type and concentration. Journal of food science 60(4):727-730.

Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD & Bramley PM. 2005. Chemical derivatization and mass spectral libraries in metabolic profiling by gc/ms and lc/ms/ms. J. Exp. Bot. 56(410):219-243.

Hall RD. 2006. Plant metabolomics: From holistic hope, to hype, to hot topic. New phytologist 169(3):453-468.

Hall RD. 2007. Food metabolomics: Meta-phor, a new european research initiative. AgroFOOD industry hi-tech 18(1):14-16.

Harrison M. 1998. Effect of breathing and saliva flow on flavor release from liquid foods. Journal of Agricultural and Food Chemistry 46(7):2727-2735.

Harrison M, Hills BP, Bakker J & Clothier T. 1997. Mathematical models of flavor release from liquid emulsions. Journal of food science 62(4):653-658.

Hasegawa S, Miyake M, Ozaki Y & Berhow M. 1996. Limonin: A non-volatile bitter principle in citrus juice. In: Pickenhagen, W., Ho, C.-T. & Spanier, A. M., editors. *The contribution of low- and nonvolatile materials to the flavor of foods.* Allured Publishing Corp. p. 137-145.

Havekotte M, Hofman T, Glabasnia A, Nagle C, Morello MJ, Rakofsky TA & Jordan RL, inventors; Tropicana Products, Inc., assignee. 2009. Control of flavor characteristics of and indicator of freshness in fruit juice. US patent WO 2009/049046 A1.

Holman WIM & De Lange DJ. 1950. Metabolism of nicotinic acid and related compounds by humans. Nature 165(4198):604-605.

Hort J & Hollowood TA. 2004. Controlled continuous flow delivery system for investigating taste-aroma interactions. Journal of Agricultural and Food Chemistry 52(15):4834-4843.

Huang MT, Ho C-t, Rosen RT, Ghai G, Lipkin M, Chen KY, Telang N, Boyd C & Csiszar K, inventors. 2007. Extracts of orange peel for prevention and treatment of cancer. Us patent 7201928.

Idborg H, Zamani L, Edlund P-O, Schuppe-Koistinen I & Jacobsson SP. 2005a. Metabolic fingerprinting of rat urine by lc/ms: Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry. Journal of Chromatography B 828(1-2):9-13.

Idborg H, Zamani L, Edlund P-O, Schuppe-Koistinen I & Jacobsson SP. 2005b. Metabolic fingerprinting of rat urine by lc/ms: Part 2. Data pretreatment methods for handling of complex data. Journal of Chromatography B 828(1-2):14-20.

Jung D-M, de Ropp JS & Ebeler SE. 2000. Study of interactions between food phenolics and aromatic flavors using one- and two-dimensional 1h nmr spectroscopy. Journal of Agricultural and Food Chemistry 48(2):407-412.

Kalivas JH. 2009. Calibration methodologies. In: Brown, S., Tauler, R. & Walczak, B., editors. *Comprehensive chemometrics : Chemical and bochemical data analysis.* Elsevier.

Katajamaa M & Oresic M. 2007. Data processing for mass spectrometry-based metabolomics. Journal of Chromatography A 1158(1-2):318-328.

Keast RSJ, Dalton P & Breslin PAS. 2004. Flavor interactions at the sensory level. In: Taylor, A. J. & Roberts, D. D., editors. *Flavor perception.* Iowa, USA: Blackwell Publishing Ltd. p. 228-254.

Kell DB & Oliver SG. 2004. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays 26(1):99-105.

Kerbiriou P, Plotto A, Goodner K, Baldwin E & Gmitter FG, Jr. 2007a. Distribution of aroma volatiles in a population of tangerine hybrids. Proceedings of the Florida State Horticultural Society 120:267-275.

Kerbiriou P, Plotto A, Goodner K, Baldwin E & Gmitter FG, Jr. 2007b. Quality evaluation of tangerines from the crec breeding program. Subtropical Technology Conference Proceedings 58:14-15.

King BM & Solms J. 1982. Interactions of volatile flavor compounds with propyl gallate and other phenols as compared with caffeine. Journal of Agricultural and Food Chemistry 30(5):838-840.

Kjeldahl K & Bro R. 2010. Some common misunderstandings in chemometrics. Journal of Chemometrics 24(7-8):558-564.

Koliandris A, Lee A, Ferry A-L, Hill S & Mitchell J. 2008. Relationship between structure of hydrocolloid gels and solutions and flavor release. Food Hydrocolloids 22(4):623-630.

Kreutzmann S, Svensson VT, Thybo AK, Bro R & Petersen MA. 2008. Prediction of sensory quality in raw carrots (*daucus carota* l.) using multi-block ls-parpls. Food Quality and Preference 19(7):609-617.

Kunert J & Qannari EM. 1999. A simple alternative to generalized procrustes analysis: Application to sensory profiling data. Journal of Sensory Studies 14(2):197-208.

Labbe D, Rytz A, Morgenegg C, Ali S & Martin N. 2007. Subthreshold olfactory stimulation can enhance sweetness. Chemical Senses 32(3):205-214.

Laing DG & Jinks A. 1996. Flavor perception mechanisms. Trends in Food Science & Technology 7(12):387-389.

Laing DG, Panhuber H, Willcox ME & Pittman EA. 1984. Quality and intensity of binary odor mixtures. Physiology and Behavior 33(2):309-319.

Lawlor JB, Delahunty CM, Sheehan J & Wilkinson MG. 2003. Relationships between sensory attributes and the volatile compounds, non-volatile and gross compositional constituents of six blue-type cheeses. International Dairy Journal 13(6):481-494.

Lê S & Husson F. 2008. Sensominer: A package for sensory data analysis. Journal of Sensory Studies 23(1):14-25.

Lê S, Josse J & Husson F. 2008. Factominer: An r package for multivariate analysis. Journal of Statistical Software 25(1):1-18.

Lee WEI. 1986. A suggested instrumental technique for studying dynamic flavor release from food products. Journal of food science 51(1):249-250.

Li S & Weber SG. 1999. Selectivity in spme. In: Pawliszyn, J., editor). *Applications of solid phase micro extraction.* Cambridge, UK: The Royal Society of Chemistry. p. 49-56.

Lindemann B. 1996. Taste reception. Physiological Reviews 76(3):719-766.

Lindemann B. 2001. Receptors and transduction in taste. Nature (London, United Kingdom) 413(6852):219-225.

Lindinger C, Labbe D, Pollien P, Rytz A, Juillerat MA, Yeretzian C & Blank I. 2008. When machine tastes coffee: Instrumental approach to predict the sensory profile of espresso coffee. Analytical Chemistry 80(5):1574-1581.

Lindinger C, Pollien P, de Vos RCH, Tikunov Y, Hageman JA, Lambot C, Fumeaux R, Voirol-Baliguet E & Blank I. 2009. Identification of ethyl formate as a quality marker of the fermented off-note in coffee by a nontargeted chemometric approach. Journal of Agricultural and Food Chemistry 57(21):9972-9978.

Lindinger W, Hansel A & Jordan A. 1998. Online monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (ptr-ms). Medical applications, food control and environmental research. International Journal of Mass Spectrometry and Ion Processes 173(3):191-241.

Linforth R & Taylor A. 2006. The process of flavour release. In: Voilley, A. & Etiévant, P., editors. *Flavour in food.* Cambridge, UK: Woodhead Publishing Ltd. p. 287-307.

Lommen A. 2009. Metalign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Analytical Chemistry 81(8):3079-3086.

Lommen A. 2011. Metalign manual: Experimental design and checks.Ppt. p. available for download at http://www.metalign.wur.nl/UK/Download+and+publications.

Lotong V, Chambers EI & Chambers DH. 2003. Categorization of commercial orange juices based on flavor characteristics. Journal of food science 68:722-725.

Luthria DL, Lin L-Z, Robbins RJ, Finley JW, Banuelos GS & Harnly JM. 2008. Discriminating between cultivars and treatments of broccoli using mass spectral fingerprinting and analysis of variance-principal component analysis. Journal of Agricultural and Food Chemistry 56(21):9819-9827.

Lynch PJ. 2006. Head lateral mouth detail. p. http //commons.wikimedia.org/wiki/Image Head_lateral_mouth_detail.jpg.

Mahattanatawee K, Rouseff R, Valim MF & Naim M. 2005. Identification and aroma impact of norisoprenoids in orange juice. Journal of Agricultural and Food Chemistry 53(2):393-397.

Marsili R. 2007. Sensory-directed flavor analysis. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Matsunami H, Montmayeur J-P & Buck LB. 2000. A family of candidate taste receptors in human and mouse. Nature 404(6778):601-604.

Max M, Shanker YG, Huang L, Rong M, Liu Z, Campagne F, Weinstein H, Damak S & Margolskee RF. 2001. Tas1r3, encoding a new candidate taste receptor, is allelic to the sweet responsiveness locus sac. Nature Genetics 28(1):58-63.

Meilgaard M, Civille GV & Carr BT. 2007a. Measuring responses. *Sensory evaluation techniques.* 4th ed. Boca Raton, FL: CRC Press, Taylor & Francis Group. p. 45-63.

Meilgaard M, Civille GV & Carr BT. 2007b. Sensory evaluation techniques, 4th ed. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Mestres M, Moran N, Jordan A & Buettner A. 2005. Aroma release and retronasal perception during and after consumption of flavored whey protein gels with different textures. 1. In vivo release analysis. Journal of Agricultural and Food Chemistry 53(2):403.

Mills GA & Walker V. 2000. Headspace solid-phase microextraction procedures for gas chromatographic analysis of biological fluids and materials. Journal of Chromatography A 902(1):267-287.

Miyazaki T. 2009. Evaluation of aroma volatiles and their odor activity in a population of tangerine hybrids. Ms thesis. Institute of Food and Agricultural Sciences - Citrus Research and Education Center. Lake Alfred, FL: University of Florida. p. 99.

Miyazaki T, Plotto A, Baldwin E & Gmitter FG, Jr. 2008. Flavor evaluation of fresh tangerine hybrids. Subtropical Technology Conference Proceedings 59:20-21.

Miyazaki T, Plotto A, Goodner K & Gmitter FG, Jr. 2011. Distribution of aroma volatile compounds in tangerine hybrids and proposed inheritance. Journal of the Science of Food and Agriculture 91(3):449-460.

Mizrahi S & Berk Z. 1970. Physicochemical characteristics of orange juice cloud. Journal of the Science of Food and Agriculture 21(5):250-253.

Mori H & Begley T. 2008. Metabolomic analysis of microorganisms. Mol. BioSyst. 4(2):108-109.

Moshonas MG & Shaw PE. 1974. Quantitative and qualitative analysis of tangerine peel oil. Journal of Agricultural and Food Chemistry 22(2):282-284.

Moshonas MG & Shaw PE. 1994. Quantitative determination of 46 volatile constituents in fresh, unpasteurized orange juices using dynamic headspace gas chromatography. Journal of Agricultural and Food Chemistry 42(7):1525-1528.

Moshonas MG & Shaw PE. 1997. Quantitation of volatile constituents in mandarin juices and its use for comparison with orange juices by multivariate analysis. Journal of Agricultural and Food Chemistry 45(10):3968-3972.

Moskowitz HR. 1996. Experts versus consumers: A comparison. Journal of Sensory Studies 11(1):19-37.

Nagy S & Shaw PE. 1990. Factors affecting the flavor of citrus fruit. Food flavours, part. C. Developments in Food Science 3C:93-124.

Noble AC. 1996. Taste-aroma interactions. Trends in Food Science & Technology 7(12):439-444.

Noes T & Kowalski B. 1989. Predicting sensory profiles from external instrumental measurements. Food Quality and Preference 1(4-5):135-147.

Nongonierma A, Cayot P, Le Quéré J-L, Springett M & Voilley A. 2006. Mechanisms of extraction of aroma compounds from foods using adsorbents. Effect of various parameters. Food reviews international 22:51-94.

Nordstrom A, Want E, Northen T, Lehtio J & Siuzdak G. 2007. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. Analytical Chemistry 80(2):421-429.

Oliver SG, Winson MK, Kell DB & Baganz F. 1998. Systematic functional analysis of the yeast genome. Trends in Biotechnology 16(9):373-378.

Omics.org. 2008. Alphabetically ordered list of omes and omics. Http://omics.Org/index.Php/alphabetically_ordered_list_of_omes_and_omics.

Overbosch P, Afterof WGM & Haring PGM. 1991. Flavor release in the mouth. Food reviews international 7(2):137-184.

Pangborn RM. 1981. A critical review of threshold, intensity and descriptive analyses in flavor research. In: Schreier, P., editor). *Flavour'81.* 3rd Weurman symposium proceedings of the international conference, Munich, April 28-30, 1981 ed. New York, NY: Walter de Gruyter. p. 3-32.

Pangborn RM & Szczesniak AS. 1974. Effect of hydrocolloids and viscosity on flavor and odor intensities of aromatic flavor compounds. Journal of Texture Studies 4(4):467-482.

Parab GS, Rao R, Lakshminarayanan S, Bing YV, Moochhala SM & Swarup S. 2009. Data-driven optimization of metabolomics methods using rat liver samples. Analytical Chemistry 81(4):1315-1323.

Park GL, Byers JL, Fritz CM, Nelson DB, Navarro JL, Smolensky DC & Vandercook CE. 1983. Characteristics of california navel orange juice and pulpwash. Journal of food science 48(2):627-632.

Parker GH. 1912. The relation of smell, taste, and the common chemical sense in vertebrates. Journal of the Academy of Natural Sciences of Philadelphia 15:221-234.

Perez-Cacho PR, Galan-Soldevilla H, Mahattanatawee K, Elston A & Rouseff RL. 2008. Sensory lexicon for fresh squeezed and processed orange juices. Food science and technology international 14(5 (suppl.)):131-141.

Perez-Cacho PR, Mahattanatawee K, Smoot JM & Rouseff R. 2007. Identification of sulfur volatiles in canned orange juices lacking orange flavor. Journal of Agricultural and Food Chemistry 55(14):5761-5767.

Perez-Cacho PR & Rouseff RL. 2008a. Fresh squeezed orange juice odor: A review. Critical reviews in food science and nutrition 48(7):681-695.

Perez-Cacho PR & Rouseff RL. 2008b. Processing and storage effects on orange juice aroma: A review. Journal of Agricultural and Food Chemistry 56(21):9785-9796.

Pérez-López A, Saura D, Lorente J & Carbonell-Barrachina Á. 2006. Limonene, linalool, α-terpineol, and terpinen-4-ol as quality control parameters in mandarin juice processing. European food research and technology 222(3):281-285.

Pérez-López AJ & Carbonell-Barrachina ÁA. 2006. Volatile odour components and sensory quality of fresh and processed mandarin juices. Journal of the Science of Food and Agriculture 86(14):2404-2411.

Petersen MA, Tønder D & Poll L. 1998. Comparison of normal and accelerated storage of commercial orange juice-changes in flavour and content of volatile compounds. Food Quality and Preference 9(1-2):43-51.

Pierce KM, Hoggard JC, Mohler RE & Synovec RE. 2008. Recent advancements in comprehensive two-dimensional separations with chemometrics. Journal of Chromatography A 1184(1-2):341-352.

Piggott JR, Simpson SJ & Williams SAR. 1998. Sensory analysis. International Journal of Food Science and Technology 33(1):7-18.

Pino J. 1982. Correlation between sensory and gas-chromatographic measurements on orange volatiles. Acta Alimentaria 11(1):1-9.

Plotto A, Gmitter FG, Jr., Kerbiriou P, Miyazaki T, Valim MF & Baldwin E. 2010. A three-year sensory fruit quality study in a group of new tangerine hybrids. HortScience 45(8):S180-181.

Plotto A, Margaria CA, Goodner KL & A. BE. 2008. Odour and flavour thresholds for key aroma components in an orange juice matrix: Esters and miscellaneous compounds. Flavour and fragrance journal 23(6):398-406.

Plotto A, Margaria CA, Goodner KL, Goodrich R & Baldwin EA. 2004. Odour and flavour thresholds for key aroma components in an orange juice matrix: Terpenes and aldehydes. Flavour and fragrance journal 19:491-498.

Pongsuwan W, Bamba T, Harada K, Yonetani T, Kobayashi A & Fukusaki E. 2008a. High-throughput technique for comprehensive analysis of japanese green tea quality assessment using ultra-performance liquid chromatography with time-of-flight mass spectrometry (uplc/tof ms). Journal of Agricultural and Food Chemistry 56(22):10705-10708.

Pongsuwan W, Bamba T, Yonetani T, Kobayashi A & Fukusaki E. 2008b. Quality prediction of japanese green tea using pyrolyzer coupled gc/ms based metabolic fingerprinting. Journal of Agricultural and Food Chemistry 56(3):744-750.

Prescott J & Stevenson RJ. 1995. Effects of oral chemical irritation on tastes and flavors in frequent and infrequent users of chili. Physiology and Behavior 58(6):1117-1127.

Prithiviraj B, Vikram A, Kushalappa AC & Yaylayan V. 2004. Volatile metabolite profiling for the discrimination of onion bulbs infected by *erwinia carotovora ssp. Carotovora*, *fusariumoxysporum* and *botrytis allii*. Eur. J. Plant Pathol. 110(4):371-377.

Pupin AM, J. DM & F. TMC. 1999. Hplc analysis of carotenoids in orange juice. Food chemistry 64(2):269-275.

Quilez J, Ruiz JA & Romero MP. 2006. Relationships between sensory flavor evaluation and volatile and nonvolatile compounds in commercial wheat bread type baguette. Journal of food science 71(6):S423-S427.

Radford T, Kawashima K, Friedel PK, Pope LE & Gianturco MA. 1974. Distribution of volatile compounds between the pulp and serum of some fruit juices. Journal of Agricultural and Food Chemistry 22(6):1066-1070.

Ramana KV, Govindarajan VS & Ranganna S. 1981. Citrus fruits - varieties, chemistry, technology, and quality evaluation. Part i: Varieties, production, handling, and storage. Critical reviews in food science and nutrition 15(4):353-431.

Ranganna S, Govindarajan VS & Ramana KVR. 1983a. Citrus fruits - varieties, chemistry, technology, and quality evaluation. Part ii. Chemistry, technology, and quality evaluation. A. Chemistry. Critical reviews in food science and nutrition 18(4):313-386.

Ranganna S, Govindarajan VS & Ramana KVR. 1983b. Citrus fruits. Part ii. Chemistry, technology, and quality evaluation. B. Technology. Critical reviews in food science and nutrition 19(1):1-98.

Rega B, Fournier N & Guichard E. 2004a. Suspended solids influence flavour profile in processed orange juice. In: Hofmann, T., Rothe, M. & Schieberle, P., editors. *State-of-the-art in flavour chemistry and biology,.* Garching, Germany: Deutsche Forschungsanstalt fuer Lebensmittelchemie. p. 412-415.

Rega B, Fournier N, Nicklaus S & Guichard E. 2004b. Role of pulp in flavor release and sensory perception in orange juice. Journal of Agricultural and Food Chemistry 52(13):4204-4212.

Reineccius GA. 2008. Flavoromics - the next frontier? Abstracts of Papers, 235th ACS National Meeting, New Orleans, LA, United States*.* American Chemical Society. p. AGFD-061.

Rentmeister-Bryant H & Green BG. 1997. Perceived irritation during ingestion of capsaicin or piperine: Comparison of trigeminal and non-trigeminal areas. Chemical Senses 22(3):257-266.

Risticevic S, Lord H, Gorecki T, Arthur CL & Pawliszyn J. 2010. Protocol for solid-phase microextraction method development. Nature Protocols 5(1):122-139.

Roberts AK & Vickers ZM. 1994. A comparison of trained and untrained judges' evaluation of sensory attribute intensities and liking of cheddar cheeses. Journal of Sensory Studies 9(1):1-20.

Roberts DD & Acree TE. 1995. Simulation of retronasal aroma using a modified headspace technique: Investigating the effects of saliva, temperature, shearing, and oil on flavor release. Journal of Agricultural and Food Chemistry 43(8):2179-2186.

Roberts DD, Elmore JS, Langley KR & Bakker J. 1996. Effects of sucrose, guar gum, and carboxymethylcellulose on the release of volatile flavor compounds under dynamic conditions. Journal of Agricultural and Food Chemistry 44(5):1321-1326.

Rospars J-P, Lansky P, Chaput M & Duchamp-Viret P. 2008. Competitive and noncompetitive odorant interactions in the early neural coding of odorant mixtures. J Neurosci 28(10):2659-2666.

Rouseff RL, Ruiz Perez-Cacho P & Jabalpurwala F. 2009. Historical review of citrus flavor research during the past 100 years. Journal of Agricultural and Food Chemistry 57(18):8115-8124.

Salles C, Hollowood TA, Linforth RST & Taylor AJ. 2003. Relating real time flavour release to sensory perception of soft cheeses. In: Le Quéré, J. L. & Etiévant, P. X., editors. *Flavour research at the dawn of the twenty-first century, proceedings of the weurman flavor research symposium, 10th, beaune, france, june 25-28, 2002.* New York: Intercept Ltd. p. 170.

Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal B, van Ommen B, Pujos-Guillot E, Verheij E, Wishart D & Wopereis S. 2009. Mass-spectrometry-based metabolomics: Limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics 5(4):435-458.

Schauer N & Fernie AR. 2006. Plant metabolomics: Towards biological function and mechanism. Trends in plant science 11(10):508-516.

Schirra HJ, Anderson CG, Wilson WJ, Kerr L, Craik DJ, Waters MJ & Lichanska AM. 2008. Altered metabolism of growth hormone receptor mutant mice: A combined nmr metabonomics and microarray study. PLoS ONE 3(7):e2764.

Schultz HC & Wahl OL. 1981. Consumer perception of the relative importance of appearance, flavor and texture to food acceptance. In: Solms, J. & Hall, R. L., editors. *Criteria of food acceptance : How man chooses what he eats.* Zürich, Switzerland: Forster. p. 97-116.

Shaw PE. 1979. Review of quantitative analyses of citrus essential oils. Journal of Agricultural and Food Chemistry 27(2):246-257.

Shaw PE, Buslig BS & Moshonas MG. 1993. Classification of commercial orange juice types by pattern recognition involving volatile constituents quantified by gas chromatography. Journal of Agricultural and Food Chemistry 41(5):809-813.

Shaw PE & Moshonas MG. 1997. Quantification of volatile constituents in orange juice drinks and its use for comparison with pure juices by multivariate analysis. Food science and technology 30(5):497-501.

Shaw PE, Moshonas MG, Buslig BS, Barros SM & Widmer WW. 1999. Discriminant and principal component analyses to classify commercial orange juices based on relative amounts of volatile juice constituents. Journal of the Science of Food and Agriculture 79(13):1949-1953.

Shaw PE & Wilson CW. 1980. Importance of selected volatile components to natural orange, grapefruit, tangerine, and mandarin flavors. *Citrus nutrition and quality.* American Chemical Society. p. 167-190.

Shepherd GM. 2006. Smell images and the flavour system in the human brain. Nature (London, United Kingdom) 444(7117):316-321.

Shimada T, Endo T, Fujii H, Hara M, Ueda T, Kita M & Omura M. 2004. Molecular cloning and functional characterization of four monoterpene synthase genes from *citrus unshiu* marc. Plant Science 166(1):49-58.

Shirey RE & Mindrup RF. 1999. Spme-adsorption versus absorption: Which fiber is best for your application? The Reporter. <http://www.sigmaaldrich.com/etc/medialib/docs/Supelco/The_Reporter/7943.Par.0001.File.tmp/7943.pdf>: Supelco Application Notes.

Siegmund B, Derler K & Pfannhauser W. 2003. The suitability of solid-phase microextraction (spme) for the analysis of volatile compounds from fruit juices or nectars. In: Le Quéré, J. L. & Etiévant, P. X., editors. *Flavour research at the dawn of the twenty-first century, proceedings of the weurman flavor research symposium, 10th, beaune, france, june 25-28, 2002.* New York: Intercept Ltd. p. 719-722.

Smilde AK, van der Werf MJ, Bijlsma S, van der Werff-van-der Vat BJC & Jellema RH. 2005. Fusion of mass spectrometry-based metabolomics data. Analytical Chemistry 77(20):6729-6736.

Steffen A & Pawliszyn J. 1996. Analysis of flavor volatiles using headspace solid-phase microextraction. Journal of Agricultural and Food Chemistry 44(8):2187-2193.

Stevens SS. 1957. On the psychophysical law. Psychological Review 64(3):153-181.

Sugimoto M, Koseki T, Hirayama A, Abe S, Sano T, Tomita M & Soga T. 2010. Correlation between sensory evaluation scores of japanese sake and metabolome profiles. Journal of Agricultural and Food Chemistry 58(1):374-383.

Sumner LW, Urbanczyk-Wochniak E & Broeckling CD. 2007. Metabolomics data analysis, visualization, and integration. In: Edwards, D., editor). *Plant bioinformatics.* Totowa, NJ, US: Humana Press Inc. p. 409-436.

Swartz ME. 2005. Uplc: An introduction and review. J. Liq. Chromatogr. Relat. Technol. 28(7-8):1253-1263.

Tandon KS, Baldwin EA, Scott JW & Shewfelt RL. 2003. Linking sensory descriptors to volatile and nonvolatile components of fresh tomato flavor. Journal of food science 68(7):2366-2371.

Tarachiwin L, Masako O & Fukusaki E. 2008. Quality evaluation and prediction of *citrullus lanatus* by 1h nmr-based metabolomics and multivariate analysis. Journal of Agricultural and Food Chemistry 56(14):5827-5835.

Taylor A, Hollowood T, Davidson J, Cook D & Linforth R. 2003. Relating taste, aroma and mouthfeel stimuli to overall flavour perception. In: Le Quéré, J. L. & Etiévant, P. X., editors. *Flavour research at the dawn of the twenty-first century, proceedings of the weurman flavor research symposium, 10th, beaune, france, june 25-28, 2002.* New York: Intercept Ltd. p. 194-199.

Taylor AJ, Linforth RST, Harvey BA & Blake A. 2000. Atmospheric pressure chemical ionisation mass spectrometry for in vivo analysis of volatile flavour release. Food chemistry 71(3):327.

Tenenhaus M, Pagès J, Ambroisine L & Guinot C. 2005. Pls methodology to study relationships between hedonic judgements and product characteristics. Food Quality and Preference 16(4):315-325.

Tietel Z, Plotto A, Fallik E, Lewinsohn E & Porat R. 2011. Taste and aroma of fresh and stored mandarins. Journal of the Science of Food and Agriculture 91(1):14-23.

Tikunov Y, Lommen A, Ric de Vos CH, Verhoeven HA, Bino RJ, Hall RD & Bovy AG. 2005. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. Plant physiology 139(3):1125-1137.

Tikunov YM, de Vos RCH, Gonzalez Paramas AM, Hall RD & Bovy AG. 2010. A role for differential glycoconjugation in the emission of phenylpropanoid volatiles from tomato fruit discovered using a metabolic data fusion approach. Plant physiology 152(1):55-70.

Tolstikov VV & Fiehn O. 2002. Analysis of highly polar compounds of plant origin: Combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. Anal. Biochem. 301(2):298-307.

Trygg J, Holmes E & Lundstedt T. 2007. Chemometrics in metabonomics. J. Proteome Res. 6(2):469-479.

University of California Riverside. 2011. Citrus variety collection. http://www.citrusvariety.ucr.edu/citrus/mandarins.html.

University of Florida IFAS Extension. 2011. Citrus cultivars. http://polkhort.ifas.ufl.edu/a_z_publications.shtml.

USDA Foreign Agricultural Service. 2008. The u.S. And world situation: Citrus. Http://www.Fas.Usda.Gov/htp/horticulture/citrus/07-30-08_citrus_presentation.Pdf.

USDA Foreign Agricultural Service. 2011. Citrus: World markets and trade. Http://www.Fas.Usda.Gov/htp/2011_jan_citrus.Pdf.

van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK & van der Werf MJ. 2006. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. BMC Genomics 7:142.

van Ruth SM. 2001. Methods for gas chromatography-olfactometry: A review. Biomol Eng 17(4-5):121-128.

van Ruth SM, O'Connor CH & Delahunty CM. 2000. Relationships between temporal release of aroma compounds in a model mouth system and their physico-chemical characteristics. Food chemistry 71(3):393-399.

van Ruth SM & Roozen JP. 2000. Influence of mastication and saliva on aroma release in a model mouth system. Food chemistry 71(3):339-345.

Velez C, Costell E, Orlando L, Nadal MI, Sendra JM & Izquierdo L. 1993. Multidimensional scaling as a method to correlate sensory and instrumental data of orange juice aromas. Journal of the Science of Food and Agriculture 61(1):41-46.

Wangen LE & Kowalski BR. 1989. A multiblock partial least squares algorithm for investigating complex chemical systems. Journal of Chemometrics 3(1):3-20.

Waybright TJ, Van QN, Muschik GM, Conrads TP, Veenstra TD & Issaq HJ. 2006. Lc-ms in metabonomics: Optimization of experimental conditions for the analysis of metabolites in human urine. J. Liq. Chromatogr. Relat. Technol. 29(17):2475-2497.

Weel KGC, Boelrijk AEM, Burger JJ, Verschueren M, Gruppen H, Voragen AGJ & Smit G. 2004. New device to simulate swallowing and in vivo aroma release in the throat from liquid and semiliquid food systems. Journal of Agricultural and Food Chemistry 52(21):6564-6571.

Weihs C. 1993. Multivariate exploratory data analysis and graphics: A tutorial. Journal of Chemometrics 7(5):305-340.

Werner E, Heilier J-F, Ducruix C, Ezan E, Junot C & Tabet J-C. 2008. Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends. Journal of Chromatography B 871(2):143-163.

Westerhuis JA, Kourti T & Macgregor JF. 1998. Analysis of multiblock and hierarchical pca and pls models. Journal of Chemometrics 12(5):301-321.

Westerhuis JA & Smilde AK. 2001. Deflation in multiblock pls. Journal of Chemometrics 15(5):485-493.

Williams AA. 1994. Flavour quality - understanding the relationship between sensory responses and chemical stimuli. What are we trying to do? The data, approaches, problems. Food Quality and Preference 5:3-16.

Wishart DS. 2008a. Applications of metabolomics in drug discovery and development. Drugs in R&D 9(5):307-322.

Wishart DS. 2008b. Metabolomics: Applications to food science and nutrition research. Trends in Food Science & Technology 19(9):482-493.

Wold S. 1995. Chemometrics; what do we mean with it, and what do we want from it? Chemometrics and Intelligent Laboratory Systems 30(1):109-115.

Wold S, Kettaneh N & Tjessem K. 1996. Hierarchical multiblock pls and pc models for easier model interpretation and as an alternative to variable selection. Journal of Chemometrics 10(5-6):463-482.

Wold S, Sjöström M & Eriksson L. 2001. Pls-regression: A basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58(2):109-130.

Woo HM, Kim KM, Choi MH, Jung BH, Lee J, Kong G, Nam SJ, Kim S, Bai SW & Chung BC. 2009. Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. Clin. Chim. Acta 400(1-2):63-69.

Yajima I, Yanai T, Sakakibara H & Hayashi K. 1979. Compositions of the volatiles of peel oil and juice from *citrus unshiu*. Agricultural and Biological Chemistry 43(2):259-264.

Yamada K. 1977. On the betaine compounds of satsuma mandarin juice. Nippon Shokuhin Kogyo Gakkaishi 24:25-30.

Yang X & Peppard T. 1994. Solid-phase microextraction for flavor analysis. Journal of Agricultural and Food Chemistry 42(9):1925-1930.

Zhang X, Breksa AP, Mishchuk DO & Slupsky CM. 2011. Elevation, rootstock, and soil depth affect the nutritional quality of mandarin oranges. Journal of Agricultural and Food Chemistry 59(6):2672-2679.