Filamentous fungi serve a very important role in Nature where they break down organic matter, releasing nutrients that can be used by other organisms. Fungi and other microorganisms also produce a wide array of bioactive compounds, the secondary metabolites (SMs), used for such diverse roles as signaling, defense, or pigmentation. Compounds from microorganisms have a dual impact on human society: they have been used as drugs, or as inspiration for the development of drugs for centuries. However, fungal infection of crops and the subsequent contamination by mycotoxins, continue to pose a threat to human health. Because of this, methods for detection and analysis of these compounds are vital. Estimates suggest that there are around 1.5 million different fungal species on Earth, dwarfing the number of plants estimated to 300,000, meaning that there potentially are many more interesting compounds are still to be discovered. The main analytical technique used to investigate production of products from these diverse organisms is liquid-chromatography coupled to mass spectrometry (LC-MS). With the development of new and improved analytical instrumentation for chemical analysis, the time needed to perform a single analytical run has decreased, while the amount of information obtained from each of these analytical runs has increased drastically. Consequently, the limiting step in chemical analysis of a microorganism is no longer the analytical run itself, but rather analysis of the resulting data. Classical methods for manual interpretation of one single data file at a time are not sufficient to cope with this influx of data. Hence, there is a need for development of new methods for data analysis to extract valuable information in the data, and also speeding up the data analysis itself. A prime goal of my PhD study was to develop methods that allow for high-throughput analysis of metabolic extracts from filamentous fungi and other microorganisms, and to reduce the time spent on manual interpretation of LC-MS data. This lead to development of a method that utilizes compound libraries to screen the recorded LC-MS data and annotate known compounds, a process we have named aggressive dereplication. By overlaying automatically generated extracted-ion chromatograms from detected compounds on the base peak chromatogram, all major potentially novel peaks can be visualized, allowing for fast dereplication of samples. This was further developed to include the use of recorded MS/MS data, allowing for greater confidence in matched compounds.

Another goal of the present study has been to develop methods that allow for faster coupling of SMs to their biosynthetic genes, as coupling of genes to metabolites is of large commercial interest for production of the bioactive compounds of the future. One part of my study focused on identification and elucidation of the biosynthesis of a nonribosomal peptide (NRP) nidulanin A from Aspergillus nidulans. Although the study was successful several analogs were not structure elucidated due to very low production titers. Instead a novel approach was developed for probing the biosynthesis of NRPs using stable isotope labeled (SIL) amino acids and subsequent analysis by MS/MS. Recorded MS/MS data were analyzed using molecular networking, coupling together compounds that exhibit similar MS/MS spectra. The combination of stable isotope labeling and molecular networking proved very effective for detection of structurally related NRPs. Labeling alone aided in determining the cyclic-peptide sequence, and may be used to provide information on biosynthesis of bioactive compounds. In another study, the combined approach of targeted analysis methods and SIL precursors was used to elucidate the biosynthesis of yanuthone D in A. niger, and to determine compounds biosynthesized from the same precursor. Further studies on the biosynthesis of polyketides were conducted using feeding studies with SIL precursor in order to determine advantages and disadvantages of the approach. This led to determination of the biosynthetic origin of several compounds in Fusarium including antibiotic Y, and tentative identification of an intermediate in its biosynthetic pathway. Last, benzoic acid was identified as the precursor to asperrubrol in A. niger. Finally, I have developed an integrated approach to evaluate the biosynthetic richness in bacteria and mine the associated chemical diversity. Here, 13 strains related to the marine bacterial species Pseudoalteromonas luteoviolacea were investigated in an untargeted metabolomics experiment and the results were correlated to whole-genome sequences of the strains. We found that 30 % of all chemical features and 24 % of the biosynthetic genes were unique to a single strain, while only 2 % of the features and 7 % of the genes were shared between all. The list of chemical features, originally comprising 2,000 features, was reduced to 50 discriminating features using a genetic algorithm combined with support vector machine evaluation. These features were efficiently dereplicated by molecular networking, which lead to tentative identification of several known antibacterial compounds, some of which had not previously been described from this organism. By combining metabolomics and genomics data, it was possible to link metabolites to chemical pathways at a very early stage in the discovery process. Based on these results, the data analysis methods and methodologies developed during these studies have shown to be very effective and applicable to metabolite analysis of a wide range of microorganisms, and not restricted to fungi. The developed methods have revealed new insights into microbial SMs, and it is clear that even more discoveries can be made using these methods.