PAPER

Systems chemical analytics: introduction to the challenges of chemical complexity analysis

Philippe Schmitt-Kopplin, [iD] *[ab] Daniel Hemmler, [iD] [ab]
Franco Moritz, [iD] [a] Régis D. Gougeon, [iD] [c] Marianna Lucio, [iD] [a]
Markus Meringer, [iD] [d] Constanze Müller, [iD] [a] Mourad Harir [iD] [a]
and Norbert Hertkorn [iD] [a]

Understanding complex (bio/geo)systems is a pivotal challenge in modern sciences that fuels a constant development of modern analytical technology, finding innovative solutions to resolve and analyse. In this introductory paper to the Faraday Discussion "*Challenges in the analysis of complex natural systems*", we aim to present concepts of complexity, and complex chemistry in systems subjected to biotic and abiotic transformations, and introduce the analytical possibilities to disentangle chemical complexity into its elementary parts (*i.e.* compositional and structural resolution) as a global integrated approach termed *systems chemical analytics*.

Introduction/complexity

Over the past few weeks, we have asked friends and colleagues to give their examples of complex systems/phenomena/objects and to describe their definition of "complexity" in one sentence. The answers first looked extremely heterogeneous but after detailed analysis they seemed to converge to a few concepts that are illustrated as an introduction herein.

The "complex" systems, phenomena or objects named were extremely diverse and inspired by our direct environment, such as life, ecosystems, nature, the universe, humans, brains, food, flavours or even wine. Other notions were more

[a]HelmholtzZentrum Muenchen, German Research Center for Environmental Health, Department of Environmental Sciences, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany. E-mail: schmitt-kopplin@helmholtz-muenchen.de

[b]Technical University Munich, Chair of Analytical Food Chemistry, Maximus-von-Imhof-Forum 2, 85354 Freising Weihenstephan, Germany

[c]UMR PAM Université de Bourgogne/AgroSup Dijon, Institut Universitaire de la Vigne et du Vin, Jules Guyot, Dijon, France

[d]German Aerospace Center (DLR), Earth Observation Center (EOC), Münchner Straße 20, 82234 Oberpfaffenhofen-Wessling, Germany

abstract or linked to processes, such as climate, emotions, thinking or health. Even more interesting was that the word "complexity" also stimulated associations with society, languages, religions, mathematics, art, family (teenagers – wonder why?!) and psychology. These answers were mainly related to the background, the field of expertise and the current interests of the respondents. In many cases, the complexity was associated with the description of a "complicated" system in a given quest for knowledge. Complexity is not only related to the scientist's view of natural systems but is also expressed in the dimensions of social sciences, religions, emotions or arts. For example, while a scientific theory of music may describe in mathematical terms the harmony of rhythms and succession of sounds, music evolves more emotionally and thus involves personal levels of enjoyment[1] making this art rather "complex" beyond the combinatorial sound, rhythm, timbre, voices, or choice of instrument. Any art can be projected in a similar way and a song, a sculpture or a painting may be reflected as an emergence resulting from the connections of the elements – art is a complex act of creativity. Thus, it was not surprising to find one of the answers describing "complexity" as: "*Complexity is what we see when we open the eyes, complexity is beauty*" (thank you Régis Gougeon). Bringing a similar dimension to our scientific world, let us remember Feynman's beautiful monologue about art and science concerning the beauty of a flower:

"*I have a friend who's an artist and has sometimes taken a view which I don't agree with very well. He'll hold up a flower and say "look how beautiful it is," and I'll agree. Then he says "I as an artist can see how beautiful this is but you as a scientist take this all apart and it becomes a dull thing," and I think that he's kind of nutty. First of all, the beauty that he sees is available to other people and to me too, I believe... I can appreciate the beauty of a flower. At the same time, I see much more about the flower than he sees. I could imagine the cells in there, the complicated actions inside, which also have a beauty. I mean it's not just beauty at this dimension, at one centimeter; there's also beauty at smaller dimensions, the inner structure, also the processes. The fact that the colors in the flower evolved in order to attract insects to pollinate it is interesting; it means that insects can see the color. It adds a question: does this aesthetic sense also exist in the lower forms? Why is it aesthetic? All kinds of interesting questions which the science knowledge only adds to the excitement, the mystery and the awe of a flower. It only adds. I don't understand how it subtracts.*"

Richard Feynman, Nobel-laureate in physics (from BBC Interview for Horizon 'The Pleasure of Finding Things Out' (https://www.bbc.co.uk/sn/tvradio/programmes/ horizon/broadband/archive/feynman/) animated by Fraser Davidson/https:// vimeo.com/55874553).

These personalized definitions of "complexity" converged in their description of (super)systems and exceeded simple diversity showing various layers of organization and interactions/interconnections between their elements. In these terms, a representative answer and description of complexity, contributed by our friend Nancy Hinman, is "*Complexity is the nexus between parts and processes. If the parts are divided then the processes won't occur. And conversely, if the processes are occurring then the parts cannot be dismantled*".

Further analysis involved text mining by a word cloud generated from almost 120 000 abstracts listed in PubMed, which contained the word "complexity" (Fig. 1A). The concept of complexity is highly relevant in biology and health, which can be concluded from the frequent occurrences of words such as clinics,
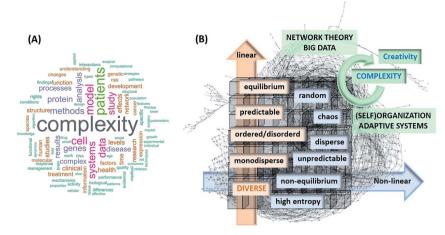
Fig. 1   (A) A word cloud created from abstracts of published articles listed in PubMed and containing the term "complexity". (B) An attempt of a schematic representation of concepts at various levels, as exerted from the definition of complexity within systems theory. While concepts such as diverse, predictable, ordered/disordered, and mono-disperse still show some linearity and, as a consequence, predictability, others such as high entropy, disperse, random, non-equilibrium and chaos may be closer to the concept of "complexity" and thus of the unpredictability of the phenomena. The latter concepts integrate more of the concepts of self-organization, adaptive systems, and network theory as needed and used in modern computational big data processing.

patients, human, risk, cancer, genes, and proteins. Additionally, its great relevance in the field of data science can be deduced from associations with data, systems, study, results, analysis, and network. Complexity is defined in various disciplines and became an important concept in modern social, natural and economic sciences, and "complexity theory" rose out of this need only a few decades ago. The word "complexity" has Latin roots: "*com*" means "*together*" and "*plectere*" means "*to plait*", showing the interconnectivity of the contributing objects/concepts. Some general definitions were given by Darley:[2] "*a typical complex system is one for which at least some of its global behaviors (that result from the interactions between a large number of relatively simple parts) cannot be predicted simply from the rules of underlying interactions*". Complexity thus not only implies interconnectivity but also possible dynamics.

"Complex systems" can be defined as being generally composed of many individual parts interacting with each other, following simple rules, synchronizing without any centralized control. The links between the activities of these parts relate to network approaches and network theory, and their dynamics in the interaction with the environments in cooperation or competition is typical of adaptive systems. This concept is illustrated in Fig. 1B. Out of their linear or non-linear dynamics, patterns can emerge, that can (self)organize and create islands of order within high entropic systems keeping a thermodynamic equilibrium of the whole. In short, as described by Ferreira,[3] "*a complex system is characterized by emergent behavior resulting from the interaction among parts*". The emerging complexity theory is interested in observations of systems and explanations of endogenous behavior, breaking down the complexity of the phenomena,[4] while

the related systems theory uses this information to trigger and optimize some functions.[5] The fundamentals of the theory behind complexity have their roots in mathematics and computer sciences, physics, chemistry, biology, ecology, social sciences, engineering, economics and arts, trying to understand or trigger the formation of complex systems such as swarms, social behaviors, business markets, ecosystems or the emergence of life and evolution in general. Even in the arts, creativity may emerge from complexity by reflecting local interactions of objects in global patterns projected by the artists.

Finally, a further example following Feynman's line with a remarkably contemporary – although not recent – representation of such complex systems, where interplay between factors as diverse as ecology, biology, chemistry, physics, sociology and economics is at the basis of an emergent behavior, is the French concept of the terroir of wines. Such emergent behavior, which is constantly challenged,[6] is indeed composed of individual parts, which all bear individual complex patterns, and yet which collectively find coherence, not only from a marketing point of view, but also from a compositional property point of view.[7,8] Richard Feynman wrote "…*But it is true that if we look at a glass of wine closely enough we see the entire universe. There are the things of physics: the twisting liquid which evaporates depending on the wind and weather, the reflections in the glass, and our imagination adds the atoms. The glass is a distillation of the Earth's rocks, and in its composition we see the secrets of the universe's age, and the evolution of stars. What strange arrays of chemicals are in the wine? How did they come to be? There are the ferments, the enzymes, the substrates, and the products. There in wine is found the great generalization: all life is fermentation*…". If a signature of stellar evolution may indeed be found in a glass of wine, individual contributions as subtle as the geographical origin of woods used for barrel aging can emerge in the final composition of a wine.[9]

Complexity in chemistry/complex natural mixtures

The concept of complexity is rather new in chemistry, a field of science that attempts to predict and control rather than simply observe and analyze.[10]

Chemistry defines us and our surroundings; *chemistry is everywhere* from nuclear and organic to polymer chemistry. According to the American Chemical Society (ACS):[11] "*Everything you hear, see, smell, taste, and touch involves chemistry and chemicals (matter). And hearing, seeing, tasting, and touching, all involve intricate series of chemical reactions and interactions in your body. With such an enormous range of topics, it is essential to know about chemistry at some level to understand the world around us*". Chemical complexity thus affects how we sense the world at all levels. Overcoming complexity, adaptation and usage of its intrinsic elements may have led to the emergence of life and further processes may be connected to evolution and progress in time and space. Thus, when defining "chemical complexity" in all environments, one needs to consider various scales in the time/space domain (short to long, tiny to infinitely large) and position these within the chemical evolution. In this context, Fig. 2 is an attempt at illustrating the linear or non-linear chemical evolution in time and space, assuming that each of the elements (presented as *time/space* voxelcubes) is highly entangled with others in
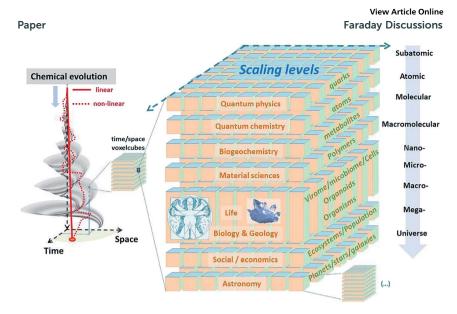
Fig. 2 Schematic representation of the connectivity in time and space of the linear/non-linear chemical evolution occurring at different scales. For example, complex co-evolution in the chemistry of living and non-living systems occurs in highly multifaceted reactivity pathways and the imperfections of its description is due to the extreme limitations in our holistic analytical observations and mathematical projections. "Complexity" is present in every voxelcube at any scale.

this process. Our main problem as humans certainly is how to represent multiple dimensions in our 3D world, and especially in this manuscript on 2D paper. The levels of complexity present in each voxelcube show scaling levels/sizes as well from atoms to molecules across organisms and ecosystems toward planetary dimensions and galaxies with mutual interactions between the respective research disciplines (from quantum physics to astronomy) with strong fractal interdependency. Chemistry itself represents only a limited part within these voxels as it is affected by biotic and abiotic processes in non-living and living systems within the described dimensions.

Chemistry is generally considered to be "complex" and "complicated" and a common strategy in chemistry is to always simplify and linearize.[10] It is generally possible on a reduced scale to explain the processes within small voxelcubes but linear interpolation across all levels and scales shown in Fig. 2 is not possible. In terms of complex chemistry, non-linear processes have not really found the focus yet, because of the limited analytical resolution of our observations leading to a generalization and to averaging if information. The resolution requirements and currently available analytical approaches are briefly discussed in this manuscript and are the focus of this Faraday Discussion on complex natural mixtures.

As illustrated in Fig. 2, scales are extremely important and each voxelcube presents its own chemistry with a high dynamic range in terms of the diversity and abundances of species that are entangled in a network of interactions. We divided the descriptions of diversity into two steps. The first level is considering all mathematically conceivable chemical structures based on the most abundant essential atoms of life (C, H, N, O, P, S) and their comparison with known

compounds in databases. In the second part we describe the chemical diversity/complexity of natural systems as they derive from abiotic and/or biotic synthesis.

In April 2019, the PubChem database and human metabolome database (HMDB, http://www.hmdb.ca/) contained 97.3 million and more than 113 000 compounds, respectively. Adopting the approach of Kauffman,[12] in Fig. 3 we plotted the number of molecules per nominal mass, showing a close to Gaussian distribution of up to 400 000 compounds per nominal mass at around 350 Da. A data reduction was possible in projecting this information into the compositional space by computing the number of chemical formulas per nominal mass and calculating the corresponding multiplicity (average number of isomers per formula) for each nominal mass. For the chemical space (as represented by the PubChem database), a maximum in multiplicity is reached around 350 Da with a maximum number of up to 7000 formulas per nominal mass around 500 Da. The HMDB is a subset of the PubChem data set and showed more than 113 000 compounds reduced to only 9118 elementary formulas having a maximum distribution in the lower mass range around 400 Da. Here, multiplicity is heavily biased by the lipid compounds in the database having an very high number of structural isomers and thus multiplicity. Lipids are among the most diverse metabolites in living systems, and although lipids are composed from a structural combination of only a few defined building blocks, they generate more than several 100 000s of individual lipid structures that vary in chain length, saturation and double bond positions.[13] Lipidomics is a rising complex discipline within the field of metabolite profiling (and metabolomics) and is essentially based on high resolution chromatography hyphenated to tandem mass spectrometry.[14]
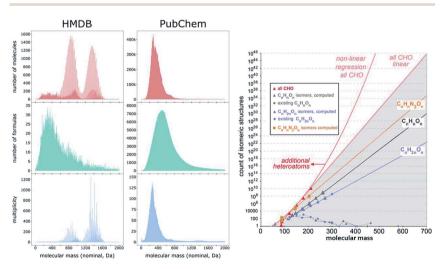

Fig. 3   Left: Abundances of molecules and multiplicity (number of molecules per formula — isomers) of related molecular compositions (formulas) over a wide mass range as found in the metabolite database (HMDB) and general chemical database (PubChem). Right: Diversity and complexity illustrated with the computed number of isomers for given restricted elementary compositions as compared to the corresponding PubChem database. The dotted lines at the bottom depict known molecular structures for the series $C_nH_{2n}O_n$ and $C_nH_nO_n$.

Compositional space, diversity/complexity

When playing with numbers, it is instructive to compare the counts of existing molecular compositions and structures as provided above with mathematically conceivable molecular compositions and isomeric molecular structures that can be computed from given molecular compositions by graph theory.[15,16] Fig. 3 provides the counts of chemically reasonable isomers for compositions $C_nH_{2n}O_n$ (one DBE) and $C_nH_nO_n$ (several DBEs), which can be computed up to $n = 10$, with extrapolation to 700 Da. Molecules of the $C_nH_{2n}O_n$ type feature a single DBE and therefore can carry only a single double bond or a single alicyclic ring; even so, the projected count of isomers at 500 Da and 700 Da reaches $\sim 10^{16}$ and $\sim 10^{22}$, respectively. Molecules of the $C_nH_nO_n$ type feature several DBEs, which give rise to many more conceivable chemical structures, including those with aromatic substructures, which are inaccessible for $C_nH_{2n}O_n$ compositions.

When clusters of nominal masses are recognized, counts of CHO molecular compositions increase moderately with increasing mass,[15] whereas the count of isomeric structures for any of these CHO molecules increases rapidly with the mass. In short, the number of CHO structures per nominal mass (called a cluster) dramatically increases with mass. For example, 18 112 feasible elemental compositions define the compositional space of CHO compounds at m/z 180–700, with H/C $\leq$ 2 and O/C $\leq$ 1.[15] If we assume 260 nominal mass clusters for CHO compounds and 20 mass peaks per cluster, we will observe a total of 5200 mass peaks susceptible to a continually growing number of isomers, ranging from $\sim 10^4$ (at m/z 180) to $\sim 10^{23}$ (at m/z 700) that will be projected on any single mass peak with a cumulative count of $>10^{35}$ conceivable isomers for CHO compounds alone (Fig. 3B). Cumulative counts of CHO compounds define the structural space without additional heteroatoms; the effects of adding heteroatoms, such as N, are exemplified using the $C_nH_nO_nN_2$ formula. The provided isomer counts are vastly underestimated because stereoisomers and several N-bearing functional groups in elevated oxidation states are *not* recognized by the used method of computation. An upward curvature with a considerable progressive size increment is observed after non-linear regression (for the sake of clarity, linear regression is applied here with one exception), which likely reflects continuous evolution of overarching structural motifs, which become accessible only when a certain mass limit is exceeded.[16]

The dotted lines in Fig. 3 depict known molecular structures for the homologous series of type $C_nH_{2n}O_n$ and $C_nH_nO_n$; the discrepancy between the vast theoretical structural space and the space very sparsely occupied by known molecules is immediately obvious. This implies that reliable extrapolations of structures within biogeochemical materials are difficult to confirm from current coverage of databases and contemporary comprehension of biochemistry. Databases overly rely on biological CHO molecules (HMDB) and islands of products from targeted chemosynthesis (PubChem), often produced by combinatorial chemistry.[17]

Natural chemical mixtures

Taking again the definitions of "complex system" from colleagues, one of the answers was "*complex is synergistic interactions of organic molecules and microbial*

transformations in natural systems" (thank you Michael Gonsior). This illustrates nicely the setup of complex chemistry involved in the interactions of elements, their adaptation, and their organization to reach homeostasis. The scheme in Fig. 4 integrates these interfaces between *biomes* and *abiomes* in biogeochemical systems and thus sets the challenges for analytical chemistry in describing the dynamic chemistry herein. Living systems (from ubiquitous microbiomes through to higher organisms through to entire ecosystems) are involved in their specific interactions and more general ecosystem adaptations, which are globally transformed on short term geological time scales to complex organic matter and geopolymers. Complexity can be found in interorganismic interactions at the macrolevel of the ecosystems[18] or at the organism level when taking account of supersystems such as holobionts.[19]

Natural chemical mixtures are key challenging samples at the center of focus for the analytical techniques presented during this Faraday Discussion. These cover a wide field of interests in this volume, such as heavy oils, natural/soil/dissolved organic matter, body fluids and plant extracts, to name just a few.

Natural chemical mixtures occur in soil,[20] freshwater,[21] estuarine,[22] marine,[23,24] arid[25] and hydrothermal environments,[26–28] and in the atmosphere,[29] and these
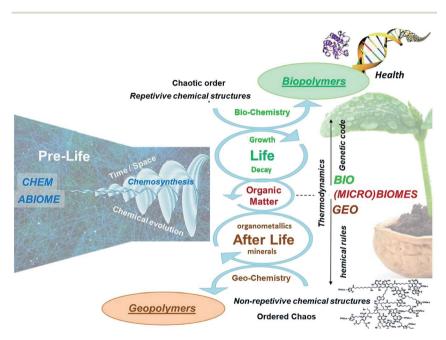

Fig. 4 *"Life in a nutshell"*: natural chemical mixtures are present in all fields of life, pre-life and after-life, with natural organic matter being at the interface of life bioprocesses and abiotic chemical complexity as an entire part of geochemistry. The chemical diversity in biology is mainly a result of and is limited to highly regulated bioprocesses. In comparison, geochemistry involves more diverse interactions including radical reactions and catalysis with metals and mineral phases in the environment and reflecting higher chemical reaction potential leads to a higher molecular diversity. NOM results from biotic and abiotic chemosynthesis and therefore structural assignments are difficult because of non-existing database knowledge. Analytical tools with higher resolution are mandatory for a trustworthy description of abiotic and biogeochemical molecular complexity.

represent exceedingly complex mixtures of organic compounds that collectively exhibit a nearly continuous range of properties (size–reactivity continuum). Their composition and structure in the bio- and geosphere are established and governed according to the rather fundamental constraints of thermodynamics and kinetics. In these intricate materials, the "classical" signatures of the (geogenic or ultimately biogenic) precursor molecules, like lipids, glycans, proteins and natural products, have been attenuated, often beyond recognition, during a succession of biotic and abiotic (*e.g.* photo and redox chemistry) reactions. Natural organic mixtures incorporate the hugely disparate characteristics of abiotic and biotic complexity.

Natural organic matter (NOM) is a conceptual subset within these complex mixtures and a key component of the global carbon cycle as well as one of the most intricate mixtures of organic compounds on earth.[15] NOM molecular composition and structure follow a dynamic equilibrium that is shaped by ecosystem characteristics with contributions from biochemical and abiotic reactions. Fundamental concepts, such as NOM lability, recalcitrance and persistence, were derived from relative differences in bulk parameters, which are subject to extensive intrinsic averaging and therefore in principle not capable of providing a definite description of NOM molecular and structural features beyond gross oversimplifications. However, massive signal overlaps leading to information projection in complex mixtures does not only apply to bulk measurements including UV, fluorescence and IR spectroscopy but also to more high-resolution techniques such as NMR spectroscopy and FTICR mass spectrometry.

Furthermore, NOM compositional and structural features reflecting the ecosystem characteristics from bulk to spectroscopy level indicate an overall restricted structural diversity determined by biochemistry rather than genuine statistical distributions. This is a result of the continual incorporation of biomolecules contributing towards a dynamic equilibrium of NOM synthesis and degradation. Initial processing of NOM will create chemical bonds between different groups of biomolecules[30] that are rare in biochemistry and hence less susceptible to enzymatic degradation (and difficult to be recognised in standard characterization).

Bio-orthogonal radical reactions (*e.g.* attack by hydroxyl radicals, lignin and tannin formation, and Fenton and quinone chemistry) initiate large scale production of novel molecules[31] and strongly attenuate biosignatures. Photochemistry recycles a high proportion of otherwise non-(bio)degradable molecules back to the biosphere. While organic matter transformations in an autumn forest soil transform biogenic leaf organic matter in days, subduction of marine sediments initiates organic matter removal and processing on millennial to geological time scales.[32] Adequate orthogonal and integrated combinations of high-resolution analytical methods are required to better understand the chemical structures of these complex organic mixtures and the processes involved in the context of global element cycling.[33] Overall, it is not expected that NOM structural diversity follows pure mathematical probabilities of structure generation. Furthermore, NOM structural diversity will incorporate core biochemical features, however, in a heavily modified manner that cannot be recognized by common analytical techniques because linkage information is either ill-defined or entirely inaccessible.

The abiome

The abiome brings us to all non-enzymatic, non-biological chemistry, such as that observed and described already in the primordial "soup". Our understanding of this ancient Earth chemistry is growing and is illustrated by non-biological systems of reactions that could have formed the network's core for converting carbon dioxide into organic compounds, as one of many seeds in the emergence of life.[34–36] Involved here are complex chemical processes of prebiotic chemistry, such as those found in extreme chemical environments like hydrothermal waters,[26] oceanic hydrothermal vents, deep-reaching tectonic faults, *etc*.[37] The classical surficial "*warm little pond*" concept introduced by Charles Darwin was most closely realized by the experiments conducted by Miller in 1953[38] or later by Wächtershäuser on catalytic surfaces.[39]

Many of these chemical reactions observed by simulations of early Earth conditions, such as hydrolysis, thermochemical sulfate reduction, photochemical, oxidation processes or various chemical conjugations, are found in all fields of life, the environment and food, to cite only a few.

To give a detailed example to illustrate the abiome, we choose the non-enzymatic browning reactions, also termed the Maillard reaction or advanced glycation. These are reactions between reducing sugars and amino compounds (*e.g.* amino acids, peptides, proteins), which lead to a heterogeneous and complex mixture of new compounds.[40,41] In heated foods, such Maillard reaction products (MRPs) are the main contributor to aroma and color formation.[42] Under physiological conditions, analogous reactions can lead to irreversible protein damage and cellular dysfunctions.[43]

The Maillard reaction is not a single type of chemical reaction but rather a superimposition of many simple chemical transformations that eventually can form a huge reaction network. Many of the reactions, independent of the precursor molecules, follow regular reaction patterns,[44] as shown in Fig. 5. Different amino precursors can flow through the shown pathways (Fig. 5A) in the same way. Typical reactions that are observed in the initial and intermediate stage of the reaction cascade are extended dehydration series, redox, and carbonyl cleavage reactions.[44,45] The amino precursor is mainly responsible for the molecular characteristics of the formed reaction products.[44] The almost infinite combining possibilities of precursor molecules, which are available for Maillard or Maillard-type reactions in complex systems (*e.g.* food products), and the same type of underlying chemical reactions ultimately lead to an extraordinary high chemical diversity. In food products, thousands of new compounds can be formed readily from only a few initial precursor molecules (Fig. 5B). As the reaction progresses, compounds with a higher degree of unsaturation and aromaticity are formed. This goes hand-in-hand with a continuous decrease in the available degrees of freedom for bond formation and compilation of atoms in the molecules (see also Fig. 6). Consequently, the compositional properties of reaction products formed from different precursor molecules more and more converge towards "end-products" of similar chemistry.[44] However, additional reactive moieties in the precursor molecules, such as thiols, may strongly influence the reactivity and form a large pool of mostly unpredictable reaction products (Fig. 5C). Nevertheless, an understanding of the entire reaction networks is
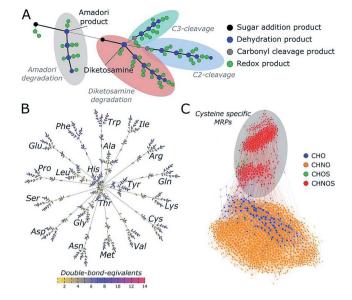
Fig. 5 (A) Schematic Maillard reaction cascade for one amino acid with one reacting sugar, (B) starting with the same sugar and the individual proteogenic amino acids, the series of compounds obtained have a high diversity with no compositional overlap and always the same reaction scaffold, (C) starting with four amino acids in mixtures and one sugar shows cross reactivity and complexity with specific reactivity of sulfur containing Cys.

key to manipulating such complex reaction cascades towards the formation of desired reaction products, while at the same time, the formation of unwanted intermediates and products is avoided. The high chemical diversity observed in the Maillard reaction and also the structural similarity (*e.g.* the formation of isomeric species by carbonyl migration mechanisms) require the highest resolution in several analytical dimensions (*i.e.* separation science, mass spectrometry and exact mass analysis, spectroscopy) to describe the chemistry even in the simplest model systems. Maillard is a great showcase of chemical diversity and complexity in general, and is found in thermal processed foods.[46] To cite here the opposite tendency in clinging to essentials, reducing complexity, "*Note by Note cooking*", as an application of molecular gastronomy, is about reducing complex foods to their elementary compounds: "*Dishes are made entirely from pure compounds or mixtures of pure compounds. No meat, fish, fruits or vegetables are used in the recipes. The aim is not to re-create foods which already exist but to create new foods and potentially new flavours. The shapes, colours, textures, consistency, odours, temperatures and trigeminal stimulation can all be designed by the chef*".[47]

The biome

Life in a nutshell – biology is (bio)chemistry as seen from a molecular perspective. Biotas are islands of high chemical organization at different scales within both theoretical chemical space and their abiotically processed environment. Subcellular organization – compartmentalization – is a living organism's central tool to
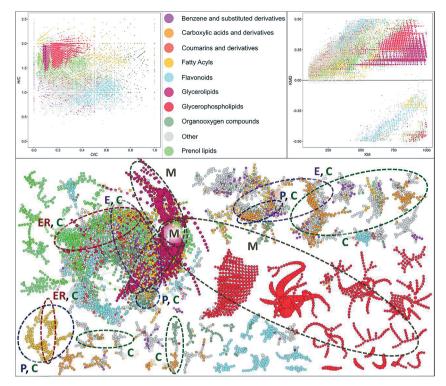
Fig. 6 Upper left: A van Krevelen (VK) diagram visualization of 9118 CHNOPS metabolites from HMDB. All features are colored according to their compound classes (the nine most abundant compound classes are colored in bright colors; all the others are shown in gray). The scatter dot sizes are scaled relative to the number of isomers per formula, as listed in HMDB. Upper right: A Kendrick mass defect plot of the same data. Bottom: A mass difference network of 6654 HMDB molecular formulas (see the main text for more details). The colors are according to the legend in the upper center. The node size is proportional to the count of isomers in HMDB. The data set was mapped against the Recon 2.2 model of the human metabolism. The regions with high annotation densities are cellular compartments C (cytoplasm), E (extracellular), ER (endoplasmic reticulum), P (peroxisome), and M (membrane).

separate its chemical space in ways that support biosynthetic processes. There could be no establishment of electrochemical gradients – the basis for energy production and conduction of neural impulses – without lipidic membrane bi-layers, whose building blocks are largely hydrophobic lipids. The basic currency, by which energy is generated and stored, spans highly hydrophilic sugars, carboxylic acids as well as ATP and NADH, whose action needs to be orchestrated both spatially and temporally, and a large number of metabolic intermediates with intermediate log P and varying pK_a values need to be guided towards their appropriate reaction partners to maximize their contribution to an organism's central aim, homeostasis. On the other hand, building blocks of bio-membranes may need to be made 'ready for transport' across hydrophilic regions of a cell by means of conjugation to carnitines, which serve as ligand structures for corresponding receptors at mitochondrial membranes. The key language of all

these processes is chemistry, and its (bio-)chemical syntax follows rules that are currently too complex for scientists to fit into a single, global model on cells, organs, organisms or their ecology. The most straightforward way that systems biology models this syntax of biochemistry is based on biochemical (metabolic) pathways, signaling pathways and other networks.

The complexity of a living being's organization requires instrumental analytics and scientific methods that can detect, identify, quantify and characterize metabolic building blocks. The best that these intricate analytical methods can do is to either get microscopic, detailed information on the organization of a small number of instances or a macroscopic, generic representation of many instances. Likewise, structural elucidation and quantification of metabolites (generally considered as "metabolomics") must merely cover those metabolites with the highest affinity towards an analytical system (NMR spectroscopy and mass spectrometry are biased towards entirely different subsets of metabolites). A wider range of metabolically relevant features can be detected with ultrahigh resolution mass spectrometry (UHR-MS), but the information obtained is neither quantitative nor complete in terms of structure elucidation. The task to understand and ultimately control complex systems by means of instrumental analysis must therefore always suffer from over-generalization along any of the analytically relevant dimensions (time, sensitivity, costs; see Fig. 8).[48] Fig. 6 shows a possible representation of human metabolism through the eyes of a UHR mass spectrometer with infinite resolution and perfect sensitivity. The upper panels in Fig. 6 show traditional representations of UHR-MS feature spaces, *i.e.* a van Krevelen diagram[49,50] (which requires molecular formulas to be assigned to *m/z* features) and a Kendrick mass defect plot[51] (KMD plots are based on accurate experimental *m/z* values).

All representations in Fig. 6 show 9118 unique molecular formulas mined from HMDB 4.0 (released 07/2018). The nine most frequent compound classes are colored, and the others are in gray. The dot and node sizes in the plots are proportional to the number of isomers listed for each represented molecular formula. Fig. 6 shows that metabolites with acetyl-CoA as a fundamental building block donor are the most readily discerned on a van Krevelen diagram and a KMD plot (the colour legend is similar for the whole figure). The corresponding glycerolipids and glycerophospholipids further show the largest numbers of known isomers per molecular formula, which substantiates their prominence: mass spectrometry-based methods for metabolome analysis are biased towards lipids, which are very efficiently ionized by electrospray ionization systems. Their retention time to structure relationships in LC-MS are straightforward and they can even be discerned in shotgun MS on a coarse compound class level. Their rules of formation further facilitate the prediction of structures, leading to an over-representation in metabolic databases.

While flavonoids can be well distinguished from the above mentioned two lipid classes in VK diagrams and KMD plots, most other compound classes appear to be intricately entangled. Better visual and conceptual discrimination of mass spectrometric data can be achieved by Kendrick-analogous mass difference networks (MDiNs),[52,53] as displayed in the lower part of Fig. 6. Molecular formulas with their monoisotopic exact masses are nodes and corresponding differences in element counts and their corresponding exact mass differences are edges in the MDiNs. The presented MDiN was built from the same data as the VK diagram and

KMD plot, and formulas were connected if their corresponding mass differences matched changes in CH_2, H_2 or O. All connections are purely compositional by nature, but they are bound to encompass true chemical reactions. An MDiN on MS data can therefore be seen as a draft metabolic network, *i.e.* a generic framework to mine the syntax of biochemistry from. The MDiN presented here shows the largest 50 sub-networks that result from the above described MDiN reconstruction. It covers 6654 (76%) of HMDB molecular formulas belonging to CHNOPS compositional space. The largest graph component (an isolated sub-network that is not connected to other sub-networks) covers 68% of all corresponding HMDB features. This MDiN articulates the special role of glycerolipids and glycerophospholipids as indicated by their large node sizes, their clear separation and highly ordered net-like patterns. All other compound classes appear entangled intricately once more, yet they show a clearer visual separation on the graph than on the VK diagram and KMD plot. A further mapping of a genome scale metabolic network model of human metabolism (Recon 2.2,[54] extracted from Metexplore,[55] containing 1018 unique molecular formulas) against this HMDB-MDiN led to 769 hits (8.8% of HMDB; 76% of Recon 2.2) and allowed the pinpointing of regions on the graph that were majorly populated by metabolites characteristic of certain cellular compartments. This example shows that the projection of both structural chemical space and even cellular compartments populate distinct regions in compositional space. While compositional space alone does not resolve information on a scale that allowed for mechanistic insights into all processes perturbing a complex biological system, its usefulness as a recommendation system – a coarse guide through complexity – turns out to be apparent.

Analyzing the metabolome certainly is the greatest challenge in modern life science omics. While in genomics, transcriptomics or proteomics the analysis relies on mathematical analysis of repetitive subunits, this is far from being the case in metabolomics. Metabolites are measured from biological samples and we have many application examples related to body fluids, such as urine, saliva, blood plasma, and tissue samples.[56,57] Even simple breathing (exhaled breath condensates) can carry information about the state of health.[58] The methods of choice for targeted or non-targeted analysis are presented in Fig. 7 as a "systems chemical analytics" toolbox.

Towards systems chemical analytics

We have presented natural complex mixtures as originating from biotic and abiotic chemical processes and defined the complex dimensions in the scale of chemical evolution. Chemistry is everywhere and there is no analytical approach to fully characterize the chemistry of a complex natural system as a whole. To quote Goethe at this point: "*One only sees what one looks for. One only looks for what one knows*". Converted to our topic, with any available technology "*one only sees what one looks for with the focus of the analytical system one uses*". Targeted analysis may be appropriate to verify a hypothesis, and non-targeted more holistic approaches may be better suited to generate a new hypothesis in addition within the frame of the analytical observations.

Novel analytical technologies enable a partial description of chemical diversity and a description of biogeosystems at different scales, and some examples are
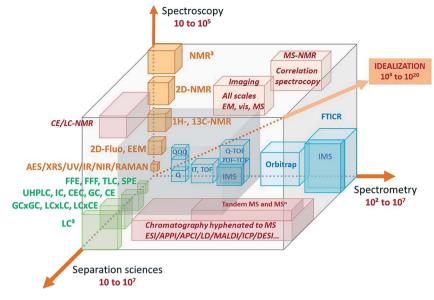
Fig. 7 A schematic representation of the systems chemical analytics space involving dimensions of spectrometry, spectroscopy and separation sciences with increasing resolution. Hyphenations such as LC-NMR, MS-NMR or LC-MS are given on the corresponding side plane projections. Additional resolution or orthogonality is obtained in mass spectrometry with various ionization methods, mass fragmentations and/or ion mobility. The gray cube in the centre covers the analytical methods used in routine applications.

presented in this Faraday Discussion. To name only a few, mass spectrometry is well represented as an orthogonal detection method following chromatography (gas and liquid chromatography (GC and LC)) or in its ultrahigh resolution mode (FTICR-MS) as a direct injection method. Ionization sources, such as electrospray- or photo-ionization (ESI, APPI), ion mobility spectrometry (IMS) and a deuterium exchange approach (specificity on H/D exchange of the analytes), are also presented for the front end of mass spectrometry as possible additional selectivity dimensions. Chromatography is presented as a sample preparation technology (solid phase extraction (SPE), supercritical extraction (SFE)) and as an analytical separation tool (GC, size exclusion chromatography (SEC), reverse phase or hydrophilic interaction chromatography (HILIC)). In these examples, the mode of detection includes spectroscopic (UV/DAD, NMR, FTIR) and mass spectrometric techniques with additional dimensions of information achieved by tandem MS. NMR methods are well represented as well in this volume with novel two or three dimensional setups, combined for an efficient description of structural information out of complex mixtures. Mathematics and statistics offer important modern tools for understanding data and provide context in terms of significant chemical and/or biological information. A few manuscripts also focus specifically on chemometrics as a universal tool to mine correlated information (*e.g.* bioactivity) in multidimensional analytical spectra and data.

For investigations of complex natural mixtures, a systems approach using different bioanalytical approaches[59] with the highest orthogonality appears the most appropriate to reveal the most significant detail. We would call this

approach "systems chemical analytics" or SCA. Systems chemical analytics involves selective separation – spectral and/or spectrometric technologies in various resolution setups complemented with extensive mathematical big-data mining (multivariate statistics, neural networks and artificial intelligence, machine learning, *etc.*). An exhaustive representation of possible SCA elements and combinations is illustrated in Fig. 7.

Bringing back metabolomes or NOM as examples of complex natural (super) mixtures, one can think about various complementary approaches that only together assemble a (still coarse) frame of their chemical shape. The mathematical capacity to distinguish different molecular structures of a given molecular composition (Fig. 3) is considerably beyond current awareness and far above the capacity of any single analytical method currently available. Hence, only a combination of several methods[16] can provide a competitive degree of information to successfully cope with real world complexity. Separation sciences, to cite only chromatography or electrophoresis,[60] are highly effective for achieving complexity reduction by the physical separation of molecules based on a wide range of discrimination criteria (size, hydrophobicity, charge, interaction, affinity, *etc.*). High-field NMR spectroscopy provides the capability for quantitative and non-destructive *de novo* determination of chemical environments from any polydisperse and molecularly heterogeneous environmental samples. Quantitative relationships between a number of spins and area (1D NMR spectroscopy) and volume (2D NMR spectroscopy) of NMR resonances operate in the absence of differential NMR spectroscopy. This key feature implies the use of NMR spectroscopy as a quantitative reference for complementary structure-selective analytical methods, like mass spectrometry (which detects gas phase ions and is subject to ionization selectivity) in the case of complex mixtures and fluorescence spectroscopy (which selectively detects fluorescent chemical environments of sp^2-hybridized carbon). To contrast the diverging selectivity of analytical methods: NMR spectroscopy is particularly informative in the description of aliphatic chemical environments which are based on sp^3-hybridized carbon. These are inactive in fluorescence spectroscopy, and the difference in the size of the aliphatic groups will cause rather inconspicuous mass shifts in the FTICR mass spectra: more expansive aliphatic systems will result in higher mass molecules, with somewhat larger H/C and smaller O/C elemental ratios. This characteristic is, however, insufficient to allow reliable conclusions about chemical structures. NMR spectroscopy enables distinction of the size of aliphatic units and also allows for in-depth assessment of the intrinsic chemical environments, like open chain and cyclic arrangements of carbon. The chemical diversity in aliphatic networks as well as typical NMR transverse relaxation rates decrease in the order $C_q > CH > CH_2 > CH_3$ leading to more comprehensible NMR-based constraints for the chemical environments of methyl groups than for carbon deep within branching networks.[61] Mass spectrometry and especially its ultrahigh resolution variants, such as FTICR mass spectrometry,[62] are perfectly suited for the classification of heteroatom-containing molecules directly from complex mixtures, but mathematical elaboration is required to reveal molecular compositions. Traditional spectra (intensity *versus* mass) show useful trends of mass evolution but the intrinsic molecular complexity is caused by isotope-specific mass defects and resides in the internal decimal places, which are not apparent from full width spectra. Various projections of the compositional space

separate integer and decimal mass numbers in two dimensions; examples include van Krevelen diagrams, mass-edited H/C ratios, and Kendrick mass defect analysis including KMD/z* diagrams, but also average carbon oxidation state *vs.* mass and other diagrams provide useful relationships between contiguous homologous series and classes of compounds. This kind of data analysis can be advanced to create multidimensional mass difference networks as described above.

As illustrated in Fig. 8, GC-MS and LC-MS are well established in the field of bioanalysis and they benefit from continuously growing databases. In the quest to optimize resolution (for molecular annotation), sensitivity (for trace quantifications) and robustness (for automation), further modern developments involve comprehensive 2D chromatography (LC×LC, GC×GC) with hyphenations to MS in the highest resolution with automated fragmentation strategies. Ideally, one seeks a technology that enables, in a few steps and short time, the integration of the data with the data analyzed in the past, using the growing knowledge over time. Quantitative approaches, such as NMR spectroscopy, enable such data integration with future analysis without complicated transfer functions and the technology can be partially automated. Rapid direct injection ultrahigh resolution mass spectrometry (UHR-MS), such as FTICR-MS and Orbitrap, is a growing field that is combined as a rapid screen (a few minutes) with LC/MS analysis for compound identification. Costs are often rapidly a limiting issue, and resolution limitations should also set the frame of possibilities or provide orientation in the choice of the right orthogonal analytical tools.[63]
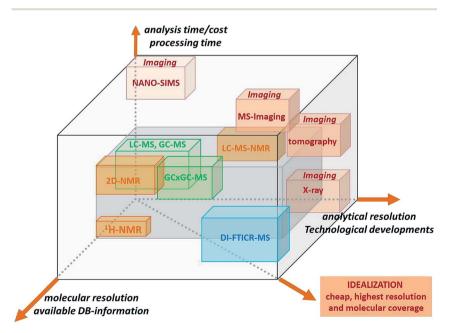

Fig. 8 Schematic representation of the systems chemical analytics as pragmatically and conceptually adapted from project management. Ideally, one aims for limiting costs by short analysis and processing times while maximising resolution in both dimensions (molecular and analytical resolution).

The rapid development of analytical technologies enables even more detailed and sensitive descriptions of complex natural systems, which is perfectly in line with our holistic goals in understanding complex processes and discovering new ones. Systems chemical analytics thus embodies the best combination of approaches for complexity deconstruction and systems simplification. On the other hand, in a pragmatic approach, even in the frame of endless analytical possibilities (in costs, information and resolution), one should never forget the goals within each individual project/experiment to adapt the systems chemical analytics accordingly.

Conflicts of interest

There are no conflicts to declare.

References

1 P. Ball, *Nature*, 2008, **453**, 160–162.
2 D. Bonchev and W. A. Seitz, The concept of complexity in chemistry, in *Concepts in Chemistry: A Contemporary Challenge*, ed. D. H. Rouvray, Chemistry Publisher: Wiley Editors, 1997, ch. 11, pp. 353–381.
3 P. Ferreira, *Tracing complexity theory, ESD 83: Research seminar in Engineering systems*, Massachusetts Institute of Technology MIT, 2001.
4 P. Galanter, Complexism and the role of evolutionary art, in *The art of artificial evolution; a handbook on evolutionary Art and Music*, ed. J. Romero and P. Machado, Springer, Berlin, 2018, pp. 311–332.
5 S. E. Phelan, *Syst. Pract. Action Res.*, 1999, **12**(3), 238.
6 J. McIntyre, *Global Food History*, 2019, **5**(1–2), 1–4.
7 C. Roullier-Gall, L. Boutegrabet, R. D. Gougeon and P. Schmitt-Kopplin, *Food Chem.*, 2014, **152**, 100–107.
8 C. Roullier-Gall, M. Lucio, L. Noret, P. Schmitt-Kopplin and R. D. Gougeon, *PLoS One*, 2014, **9**(5), e97615.
9 R. D. Gougeon, M. Lucio, M. Frommberger, D. Peyron, D. Chassagne, H. Alexandre, F. Feuillat, A. Voilley, P. Cayot, I. Gebefügi, N. Hertkorn and Ph. Schmitt-Kopplin, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**(23), 9174–9179.
10 G. M. Whitesides and R. F. Ismagilov, *Science*, 1999, **284**, 89–92.
11 https://www.acs.org/content/acs/en/education/whatischemistry/everywhere.html.
12 S. A. Kauffman, D. P. Jelenfi and G. Vattay, The clock of chemical evolution, arXiv: 1806.06716 [q-bio.PE].
13 A. Shevchenko and K. Simons, *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 593–598.
14 M. Witting and P. Schmitt-Kopplin, *Arch. Biochem. Biophys.*, 2016, **589**, 27–37.
15 N. Hertkorn, M. Frommberger, M. Witt, B. P. Koch, P. Schmitt-Kopplin and E. M. Perdue, *Anal. Chem.*, 2008, **80**, 8908–8919.
16 N. Hertkorn, C. Ruecker, M. Meringer, R. Gugisch, M. Frommberger, E. M. Perdue, M. Witt and P. Schmitt-Kopplin, *Anal. Bioanal. Chem.*, 2007, **389**, 1311–1327.
17 J. Liu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. O. O'Meara, T. Che, E. Algaa, K. Tolmachova, A. A. Tolmachev, B. K. Shoichet, B. L. Roth and J. J. Irwin, *Nature*, 2019, **566**, 224–229.

18 R. Tecon, S. Mitri, D. Ciccarese, D. Or, J. R. van der Meer and D. Johnson, *mSystems*, 2019, **4**(1), e00265-18.

19 T. C. G. Bosch and D. J. Miller, *The Holobionte imperative: Perspectives from Early emerging animals*, Springer, 2016, pp. 1–155.

20 J. Guigue, M. Harir, O. Mathieu, M. Lucio, L. Ranjard, J. Levêque and P. Schmitt-Kopplin, *Biogeochemistry*, 2016, **128**(3), 307–326.

21 M. Gonsior, J. Valle, P. Schmitt-Kopplin, N. Hertkorn, D. Bastviken, J. Luek, M. Harir, W. Bastos and A. Enrich-Prast, *Biogeosciences*, 2016, **13**(14), 4279–4290.

22 N. Hertkorn, M. Harir, K. M. Cawley, P. Schmitt-Kopplin and R. Jaffe, *Biogeosciences*, 2016, **13**, 2257–2277.

23 N. Hertkorn, M. Harir, B. P. Koch, B. Michalke and P. Schmitt-Kopplin, *Biogeosciences*, 2013, **10**, 1583–1624.

24 K. Ksionzek, O. Lechtenfeld, S. L. McCallister, P. Schmitt-Kopplin, J. K. Geuer, W. Geibert and B. P. Koch, *Science*, 2016, **354**(6311), 456–459.

25 D. Schulze-Makuch, D. Wagner, S. P. Kounaves, K. Mangelsdorf, K. G. Devine, J. P. de Vera, Ph. Schmitt-Kopplin, H. P. Grossart, V. Parro, M. Kaupenjohann, A. Galy, B. Schneider, A. Airo, J. Frosler, A. F. Davila, F. L. Arens, L. Caceres, F. Sols Cornejo, D. Carrizo, L. Dartnell, J. DiRuggiero, M. Flury, L. Ganzert, M. Gessner, P. Grathwohl, L. Guan, J. Heinz, M. Hessa, F. Kepplerb, D. Maus, C. P. McKay, R. U. Meckenstock, W. Montgomery, E. A. Oberline, A. J. Probst, J. S. Saenz, T. Sattlerb, J. Schirmack, M. A. Sephton, M. Schloter, J. Uhl, B. Valenzuela, G. Vestergaard, L. Wormerd and P. Zamorano, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**(11), 2670–2675.

26 M. Gonsior, N. Hertkorn, N. Hinman, S. E. M Dvorski, M. Harir, W. J. Cooper and P. Schmitt-Kopplin, *Sci. Rep.*, 2018, **8**(1), 14155.

27 V. La Cono, G. Bortoluzzi, E. Messina, G. La Spada, F. Smedile, L. Giuliano, M. Borghini, Ch. Stumpp, P. Schmitt-Kopplin, M. Harir, W. K. O'Neill, J. E. Hallsworth and M. Yakimov, *Sci. Rep.*, 2019, **9**(1), 1679.

28 R. U. Meckenstock, F. von Netzer, C. Stumpp, T. Lueders, A. M. Himmelberg, N. Hertkorn, P. Schmitt-Kopplin, M. Harir, R. Hosein, S. Haque and D. Schulze-Makuch, *Science*, 2014, **345**(6197), 673–676.

29 P. Schmitt-Kopplin, A. Gelencser, E. Dabek-Zlotozyrnská, G. Kiss, N. Hertkorn, M. Harir, Y. Hong and I. Gebefügi, *Anal. Chem.*, 2010, **82**, 8017–8026.

30 G. A. McKee and P. G. Hatcher, *Geochim. Cosmochim. Acta*, 2010, **74**, 6436–6450.

31 D. C. Waggoner and P. G. Hatcher, *Org. Geochem.*, 2017, **113**, 315–325.

32 P. B. Kelemen and C. E. Manning, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E3997–E4006.

33 Y. Li, M. Harir, M. Lucio, M. Gonsior, B. P. Koch, P. Schmitt-Kopplin and N. Hertkorn, *Water Res.*, 2016, **106**, 477–487.

34 A. Ruf, L. L. S. d'Hendecourt and P. Schmitt-Kopplin, *Life*, 2018, **8**, 18.

35 T. Böttcher, *J. Mol. Evol.*, 2018, **86**, 1–10.

36 K. Muchowska, S. J. Varma and J. Moran, *Nature*, 2019, **569**, 104–107.

37 B. Trias, B. Ménez, P. le Campion, Y. Zivanovic, L. Lecourt, A. Lecoeuvrel, P. Schmitt-Kopplin, J. Uhl, S. R. Gíslason, H. A. Alfreðsson, K. G. Mesfin, S. Ó. Snæbjörnsdóttir, E. S. Aradóttir, I. Gunnarsson, J. M. Matter, M. Stute, E. H. Oelkers and E. Gérard, *Nat. Commun.*, 2017, **8**, 1063.

38 S. L. Miller, *Science*, 1953, **117**, 528.

39 G. Wächtershäuser, *Syst. Appl. Microbiol.*, 1988, **10**, 207.

40 J. E. Hodge, *J. Agric. Food Chem.*, 1953, **1**, 928–943.

41 F. Ledl and E. Schleicher, *Angew. Chem., Int. Ed. Engl.*, 1990, **29**, 565–594.

42 M. Hellwig and T. Henle, *Angew. Chem., Int. Ed.*, 2014, **53**, 10316–10329.

43 T. P. Labuza, V. M. Monnier, J. Baynes and J. O'Brien, *Maillard Reactions in Chemistry, Food and Health*, Elsevier, 1st edn, 1998.

44 D. Hemmler, C. Roullier-Gall, J. W. Marshall, M. Rychlik, A. J. Taylor and P. Schmitt-Kopplin, *Sci. Rep.*, 2018, **8**, 16879.

45 D. Hemmler, C. Roullier-Gall, J. W. Marshall, M. Rychlik, A. J. Taylor and P. Schmitt-Kopplin, *Sci. Rep.*, 2017, **7**, 3227.

46 H. This, *Notes Académiques de l'Académie d'agriculture de France (N2AF)*, 2017, vol. 1, pp. 1–12.

47 H. This, *Flavour*, 2013, **2**, 1.

48 S. Forcisi, F. Moritz, B. Kanawati, D. Tziotis, R. Lehmann and P. Schmitt-Kopplin, *J. Chromatogr. A*, 2013, **1292**, 51–65.

49 D. W. Van Krevelen, *Fuel*, 1950, **29**, 269–284.

50 S. Kim, R. W. Kramer and P. G. Hatcher, *Anal. Chem.*, 2003, **75**, 5336–5344.

51 E. Kendrick, *Anal. Chem.*, 1963, **35**, 2146–2154.

52 D. Tziotis, N. Hertkorn and P. Schmitt-Kopplin, *Eur. J. Mass Spectrom.*, 2011, **17**, 415–421.

53 F. Moritz, M. Kaling, J. P. Schnitzler and P. Schmitt-Kopplin, *Plant, Cell Environ.*, 2017, **40**, 1057–1073.

54 N. Swainston, K. Smallbone, H. Hefzi, P. D. Dobson, J. Brewer, M. Hanscho, D. C. Zielinski, K. S. Ang, N. J. Gardiner, J. M. Gutierrez, S. Kyriakopoulos, M. Lakshmanan, S. Li, J. K. Liu, V. S. Martinez, C. A. Orellana, L.-E. Quek, A. Thomas, J. Zanghellini, N. Borth, D. Y. Lee, L. K. Nielsen, D. B. Kell, N. E. Lewis and P. Mendes, *Metabolomics*, 2016, **12**, 109.

55 L. Cottret, C. Frainay, M. Chazalviel, F. Cabanettes, Y. Gloaguen, E. Camenen, B. Merlet, J. C. Portais, S. Heux, N. Poupin, F. Vinson and F. Jourdan, *Nucleic Acids Res.*, 2018, **46**(W1), 495–502.

56 S. Forcisi, F. Moritz, M. Lucio, R. Lehmann, N. Stefan and P. Schmitt-Kopplin, *Anal. Chem.*, 2015, **87**(17), 8917–8924.

57 K. Neth, M. Lucio, A. Walker, B. Kanawati, J. Zorn, P. Schmitt-Kopplin and B. Michalke, *Chem. Res. Toxicol.*, 2015, **28**, 1434–1442.

58 F. Moritz, M. Janicka, A. Zygler, S. Forcisi, A. Kot-Wasik, J. Kot, I. Gebefügi, J. Namiesnik and P. Schmitt-Kopplin, *J. Breath Res.*, 2015, **9**, 027105.

59 *Bioanalytics*, ed. F. Lottspeich and J. Engels, Wiley-VCH, 2018, pp. 1–1110.

60 *Capillary Electrophoresis, Methods in Molecular Biology 1483*, ed. P. Schmitt-Kopplin, Springer, 2nd edn, 2017.

61 N. Hertkorn, M. Harir and P. Schmitt-Kopplin, *Magn. Reson. Chem.*, 2015, **53**, 754–768.

62 *Fundamentals and Applications of Fourier Transform Mass Spectrometry*, ed. B. Kanawati and P. Schmitt-Kopplin, Elsevier, 2019, pp. 1–726.

63 *Metabolomics in Practice, Successful Strategies to Generate and Analyze Metabolic Data*, ed. M. Lämmerhofer and W. Weckwerth, Wiley-VCH, Weinheim, 2013.