S1. Publication Series *Residual Complexity and Bioactivity*

The present publication forms *Part 21* in a series of communications on *Residual Complexity and Bioactivity*.

From a **chemical perspective**, residual complexity (RC) refers to the subtle but significant convolution of major and minor chemical species in materials that originate from reaction mixtures, such as natural products. Because natural products are formed biosynthetically, they inherit a certain portion of side products from the metabolomic cocktail. This RC is frequently conserved in highly purified materials, even after an elaborate analytical separation scheme has been applied. The relationship between the (bio)synthetic cocktail and the products is perpetuated by the RC of the samples. In principle, RC affects all "pure" materials. RCs can be divided into two main groups: static RC describes the thermodynamically stable cases, whereas dynamic RC refers to situations where the impurity patterns change over time due to reactivity or other chemical change that occurs during the timeframe and under the conditions of the observation (e.g., a bioassay).

From a **biological perspective**, RC can have a major impact on bioactivity. Numerous forms of bioassays are widely used for the biological assessment (in vitro, ex vivo, in vivo) of bioactive agents. As many of the bioassays are mechanistically complex by nature, biological evaluation also can be residually complex. This adds a biological layer to the overall RC of bioactive agents and applies to various levels of chemical and biological complexity. Accordingly, both the chemical RC of the agent and the biological RC of the bioassay have to be considered when interpreting information about bioactivity.

As discussed in detail in Part 12 of the publication series (Journal of Natural Products, *75*: 1243-1255 (**2012**); see also below), the recognition and analysis of **RC can help establishing links between the observed biological activity and the underlying chemistry of bioactive agents**. The following table lists the preceding publications since 2008 which establish the publication series *Residual Complexity and Bioactivity*.

Part	**Reference**	**Brief Synopsis Regarding *Residual Complexity and Bioactivity***
Part 1	Jaki BU, Franzblau SG, Chadwick LR, Lankin DC, Zhang F, Wang Y, **Pauli GF** Purity Bioactivity Relationships – The Case of Anti-TB Active Ursolic Acid Journal of Natural Products *71*: 1742-1748 (**2008**) dx.doi.org/10.1021/np800329j	Demonstrates the relationship between purity, RC and anti-TB activity of different batches of a natural product; uses qHNMR methodology to establish quantitative relationships between purity/RC and activity.
Part 2	Chen S, Turner A, Jaki B, Nikolic D, van Breemen R, Friesen B, **Pauli GF** An Experimental Implementation of Chemical Subtraction Journal of Pharmaceutical and Biomedical Analysis *46*: 692-698 (**2008**)	Performs selective removal of a single, interfering phytoconstituent from a bioactive (*E. coli* anti-adherence) fraction and demonstrates the presence of RC in the removed ("subtracted") compound and its

	dx.doi.org/10.1016/j.jpba.2007.12.014	assessment by qNMR and MS methods.
Part 3	Schinkovitz A, Pro S, Main M, Chen SN, Jaki BU, Lankin DC, **Pauli GF** The Dynamic Nature of the Ligustilide Complex Journal of Natural Products *71*: 1606-1611 (**2008**) dx.doi.org/10.1021/np800137n	Shows how RC is generated and varies in purified samples of ligustilide, a designated bioactive marker present in *Angelica* and *Ligusticum* species; compares analytical methods suitable to assess dynamic RC.
Part 4	Gödecke T, Chen SN, Lankin D, Nikolic D, van Breemen R, **Pauli GF** Phytochemistry of Cimicifugic Acids and Associated Bases in *Cimicifuga racemosa* Root Extracts Phytochemical Analysis *20*: 120-131 (**2009**) dx.doi.org/10.1002/pca.1106	Establishes the new phytochemical methodology that leads to the LC-MS-driven discovery of trace amounts of N-Methyl-serotonin as serotonergic active principle in black cohosh; demonstrates the relevance of low-abundance constituents (RC) as potential bioactive markers for metabolomic mixtures such a botanical extracts.
Part 5	Chen SN, Lankin D, Chadwick L, Jaki B, **Pauli GF** Dynamic Residual Complexity of Natural Products by qHNMR: Solution Stability of Desmethylxanthohumol Planta Medica 75: 757-762 (**2009**) dx.doi.org/10.1055/s-0028-1112209	Exemplifies how the dynamic form of RC can lead to the generation of a highly potent phytoestrogen (8-PN) from the inactive precursor (DMX); institutes qHNMR methodology to assess RC in a time-resolved fashion, enabling correlation with bioassay outcome.
Part 6	**Pauli GF**, Friesen B, Goedecke T, Farnsworth N, Glodny B Occurrence of Progesterone and Related Animal Steroids in Two Higher Plants Journal of Natural Products *73*: 338-345 (**2010**) dx.doi.org/10.1021/np9007415	Unambiguously demonstrates the unexpected occurrence of the mammalian steroid, progesterone, in higher plants and shows that small amounts of this hormone as well as mammalian-like steroid metabolites (e.g., 3-O-sulfates) can form a small but integral part of the RC of plant metabolomes.
Part 7	Molina-Salinas G, Rivas-Galindo V, Said-Fernández S, Lankin D, Muñoz M, Joseph-Nathan P, **Pauli GF***, Waksman N* [*corresponding authors] Stereochemical Analysis of Leubethanol, an Anti-TB Active Serrulatane, from *Leucophyllum frutescens* Journal of Natural Products *74*: 1842-1850 (**2011**) dx.doi.org/10.1021/np2000667	Establishes the subtle but significant diastereomeric difference between elisabethanol, which had been isolated from a gorgonian organism, and leubethanol, the anti-TB active lead compound isolated from a plant; utilizes VCD for the determination of absolute stereochemistry and emphasizes ^{1}H iterative full spin analysis (HiFSA) as a dereplication tool and for the analysis of RC of natural products.
Part	Gödecke T, Napolitano J, Yao P, Nikolic D, Dietz B,	Establishes a qHNMR-based protocol

8	Bolton J, van Breemen R, Chen SN, Lankin D, Farnsworth N, **Pauli GF** Integrated standardization concept for *Angelica* botanicals using quantitative NMR Fitoterapia *83*: 18-32 (**2012**) dx.doi.org/10.1016/j.fitote.2011.08.017	for the simultaneous quantitation of multiple marker compounds in the bioactive fraction ([anti-]estrogenicity, cytotoxicity) of *Angelica sinensis* botanicals; demonstrates the advanced role qHNMR can have in botanical standardization and evaluation of RC of the plant extracts.
Part 9	Qiu F, Imai A, McAlpine J, Lankin D, Burton I, Karakach T, Farnsworth N, Chen SN, **Pauli GF** Dereplication, Residual Complexity and Rational Naming - the Case of the *Actaea* Triterpenes Journal of Natural Products *75*: 432-443 (**2012**) dx.doi.org/10.1021/np200878s	Determines the RC of purified botanical reference materials of triterpenes from black cohosh; demonstrates the assessment of RC by computer-aided dereplication using classification binary trees (CBTs) to derive both structural information and quantitative measures for minor components contained in residually complex samples.
Part 10	Napolitano J, Gödecke T, Rodriguez Brasco MF, Jaki BU, Chen SN, Lankin DC, **Pauli GF** The Tandem of Full Spin Analysis and qHNMR for the Quality Control of Botanicals Exemplified with *Ginkgo biloba* Journal of Natural Products *75*: 238-248 (**2012**) dx.doi.org/10.1021/np200949v	Establishes ^{1}H iterative full spin analysis (HiFSA)as the basis of a qHNMR approach for the parallel quantitation of eight bioactive markers in *Ginkgo biloba*; exemplifies how multi-target standardization can be achieved without the need for identical calibrants in (residually) complex samples including reference materials, fractions, and extracts; addresses the role of RC in reference materials of calibrants.
Part 11	Qiu F, Friesen JB, McAlpine JB, **Pauli GF** NMR-based Design of Countercurrent Separation of *Ginkgo biloba* Terpene Lactones Journal of Chromatography A *1242*: 26-34 (**2012**) dx.doi.org/10.1016/j.chroma.2012.03.081	Introduces the use of qHNMR for both the design and the analysis of countercurrent separation (CS) of bioactive botanical markers; demonstrates the measurement of partition coefficients of target markers in mixtures; performs the evaluation of chromatographic orthogonality in CS; establishes quantitative links between predicted and measured chromatographic CS performance and the RC of the purified markers.
Part 12	**Pauli GF**, Chen SN, Friesen JB, McAlpine J, Jaki BU Analysis and Purification of Bioactive Natural Products - The AnaPurNa Study Journal of Natural Products, *75*: 1243-1255 (**2012**) dx.doi.org/10.1021/np300066q	Comprehensive meta-analysis of the literature (1999-2010) focusing on the role of analytical methodology in the purification and characterization of bioactive compounds from natural sources; addresses the role of RC in their purification and characterization

		and discusses the impact of RC on the biological evaluation and validation of lead compounds.
Part 13	Napolitano J, Lankin D, Chen SN, **Pauli GF** Complete ^{1}H NMR Spectral Analysis of Ten Chemical Markers of *Ginkgo biloba* Magnetic Resonances in Chemistry *50*: 569-575 (**2012**) dx.doi.org/10.1002/mrc.3829	Establishes the methodology for the generation of unambiguous ^{1}H NMR fingerprints of bioactive markers, exemplified for terpene lactones and flavonoids from *Ginkgo biloba*; the fingerprints are suitable for both structural dereplication and qHNMR quantitation at various levels of RC, can be scaled to all existing NMR field strength and are independent of instrumentation.
Part 14	Riihinen K, Gödecke T, Pauli GF Purification of Berry Flavonoids by Long-bed Gel Permeation Chromatography Journal of Chromatography A, *1244*: 20-27 (**2012**) dx.doi.org/10.1016/j.chroma.2012.04.060	Establishes long-bed gel permeation chromatography (GPC) on Sephadex LH-20 as an efficient method for the purification of bioactive berry polyphenols; despite its capability to resolve closely related compounds, qHNMR analysis reveals an un expected degree of RC in the GPC fractions.
Part 15	Inui T, Wang Y, Pro S, Franzblau SG, **Pauli GF** Unbiased Evaluation of Bioactive Secondary Metabolites in Complex Matrices Fitoterapia *83*: 1218-1225 (**2012**) dx.doi.org/10.1016/j.fitote.2012.06.012	Establishes biochemometric methodology capable of identifying bioactive principles in crude metabolomic mixtures, as an alternative to bioassay-guided fractionation; establishes chemometric links between the bioassay and the preparative and analytical chemistry of (residually) complex natural products; exemplifies the concept for the anti-TB active principles of the ethnobotanical, *Oplopanax horridus*.
Part 16	Dong SH, Nikolic D, Simmler C, Qiu F, van Breemen RB, Soejarto DD, Pauli GF, Chen SN Diarylheptanoids from Dioscorea villosa (Wild Yam) Journal of Natural Products *75*: 2168-2177 (**2012**) dx.doi.org/10.1021/np300603z	Establishes a link between the ubiquitous RC of crude metabolomes of *Dioscorea villosa* and the approach of qHNMR-guided metabolomic mining of the minor diarylheptanoid constituents. These findings not only extend the utility of qHNMR applications, but also complement previous conclusions about the relevance of RC. In addition, the establishment of HiFSA profiles of the isolates lays the groundwork for the quantifiable assessment of the RC of

		Dioscorea preparations.
Part 17	Qiu F, Cai G, Jaki BU, Lankin DC, Franzblau SG, Pauli GF Quantitative Purity-Activity Relationships of Natural Products: The Case of Anti-Tuberculosis Active Triterpenes from Oplopanax horridus Journal of Natural Products, *76*: 413-419 (**2013**) http://dx.doi.org/10.1021/np3007809	Extends the previously developed concept of purity–activity relationships (PARs) for the quantitative evaluation of the effects of multiple minor components on the bioactivity of residually complex natural products. The anti-tuberculosis active triterpenes were selected as a case for the development of the quantitative PAR (QPAR) concept. The residual complexity of the purified triterpenes from the Alaskan ethnobotanical, *Oplopanax horridus*, studied by 1D- and 2D-NMR and identified as a combination of structurally related and unrelated impurities. Using biochemometric methodology, the qHNMR purity and anti-TB activity of successive chromatographic fractions of *O. horridus* triterpenes were correlated by linear regression analysis to generate a mathematical QPAR model. The QPAR concept enables a quantitative assessment when residually complexity affects the biological activity of natural products.
Part 18	Simmler C, Hajirahimkhan A, Lankin DC, Bolton J, Jones T, Soejarto D, Chen SN, Pauli GF Dynamic Residual Complexity of the Isoliquiritigenin-Liquiritigenin Interconversion During Bioassay Journal of Agricultural and Food Chemistry, *61*: 2146-2157 (**2013**) http://dx.doi.org/10.1021/jf304445p	Demonstrates the relevance of the dynamic residual complexity of chalcones and flavanones under in vitro bioassay conditions by time-resolved investigation of the isomerization of the chalcone, isoliquiritigenin, and the flavanone, liquiritigenin in cell culture media, with and without cells. Isomerization occurs and leads to equilibrium mixtures with enantiomeric excess of the R-flavanone. The findings underscore the importance of choice and definition of in situ conditions in biological test systems as they affect chemical reactivity and, thus, residual complexity of the test agents.
Part 19	Gödecke T, Napolitano JG, Rodriguez Brasco MF, Chen SN, Jaki BU, Lankin DC, Pauli GF Validation of a Generic qHNMR Method for Natural	Introduces a validated protocol for quantitative ^{1}H NMR (qHNMR) analysis of complex samples with relatively high dynamic range. Comprehensive

	Products Analysis Phytochemical Analysis, *24*: in press (**2013**)	coverage of acquisition and processing parameters as well as software, and specific attention to the requirements of natural products makes the protocol suitable for a wide range of analytes and applications, from the quality control of highly complex crude botanical extracts to the assessment of the purity and residual complexity of reference compounds.
Part 20	Napolitano JG, Lankin D, Graf T, Friesen JB, Chen SN, McAlpine JB, Oberlies, NH, Pauli GF HiFSA Fingerprinting of Isomers with Near Identical NMR Spectra: The Silybin/Isosilybin Case The Journal of Organic Chemistry, 78: in press (**2013**) http://dx.doi.org/10.1021/jo302720h	Demonstrates how the "hidden" complexity of regio- and diastereoisomers with near-identical NMR spectra can be distinguished and unambiguously assigned by quantum mechanical driven, ^{1}H iterative Full Spin Analysis (HiFSA). The method is illustrated with four the four main flavonolignans von *Silybum marianum* ([iso]silybins A/B) which have posed an analytical challenge for nearly six decades. The highly reproducible HiFSA ^{1}H NMR fingerprints allow distinction of the near-identical our isomers at ^{1}H frequencies from 300 to 900 MHz, as well as their parallel quantification, even in difficult to characterize mixtures. The methodology opens new opportunities to explore hidden diversity in the chemical space of organic molecules.