Review

Journal of Molecular Recognition

Received: 26 August 2013, Accepted: 27 August 2013, Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jmr.2326

Systematic, spatial imaging of large multimolecular assemblies and the emerging principles of supramolecular order in biological systems

Walter Schubert[a,b,c]*

Understanding biological systems at the level of their relational (emergent) molecular properties in functional protein networks relies on imaging methods, able to spatially resolve a tissue or a cell as a giant, non-random, topologically defined collection of interacting supermolecules executing myriads of subcellular mechanisms. Here, the development and findings of parameter-unlimited functional super-resolution microscopy are described—a technology based on the fluorescence imaging cycler (IC) principle capable of co-mapping thousands of distinct biomolecular assemblies at high spatial resolution and differentiation (<40 nm distances). It is shown that the subcellular and transcellular features of such supermolecules can be described at the compositional and constitutional levels; that the spatial connection, relational stoichiometry, and topology of supermolecules generate hitherto unrecognized functional self-segmentation of biological tissues; that hierarchical features, common to thousands of simultaneously imaged supermolecules, can be identified; and how the resulting supramolecular order relates to spatial coding of cellular functionalities in biological systems. A large body of observations with IC molecular systems microscopy collected over 20 years have disclosed principles governed by a law of supramolecular segregation of cellular functionalities. This pervades phenomena, such as exceptional orderliness, functional selectivity, combinatorial and spatial periodicity, and hierarchical organization of large molecular systems, across all species investigated so far. This insight is based on the high degree of specificity, selectivity, and sensitivity of molecular recognition processes for fluorescence imaging beyond the spectral resolution limit, using probe libraries controlled by ICs. **© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.**

Additional supporting information may be found in the online version of this article at the publisher's web site.

Keywords: supermolecules; functional super-resolution; fluorescence imaging; toponome; imaging cycler; MELC; TIS; toponome imaging system

Above all is the question: how are cells and tissues organized in health and disease? The answer will not come by a molecule, not by sequence, but by a vision.

INTRODUCTION

It may be one of the most surprising facts in modern life sciences that practically all methods aiming at understanding molecular systems in organisms are based on the destruction of cells or tissues in order to analyze biomolecules on a large scale *ex vivo*. However, *in vivo*, most of these systems cannot exert successfully any specific biological functionality without following topological rules for their spatial interactions that connect them as large multimolecular assemblies within the structures of a cell or tissue. Therefore, it is not surprising that *ex vivo* approaches are related to, or held to be causal for, the inefficiency in developing successful molecular therapies against chronic diseases (Hutchinson and Kirck, 2011; Michor *et al*., 2011; Selkoe, 2011; World Alzheimer Report, 2011; Gatenby, 2012; Schubert, 2012b; McCarthy, 2013). Together with an increasing body of similar critical analyses, the recent detailed study and conclusions of Scannell *et al*. (2012) have evoked considerable scientific attention: "R&D (in the pharmaceutical industry) was more efficient several decades ago when many research activities that are today regarded as critical (for example the derivation of genomics-based drugs and high throughput screening) had not been invented." This is of course just a correlation and not necessarily a causal relationship. Similar early warnings were published already in 2001 by

* *Correspondence to: Walter Schubert, Molecular Pattern Recognition Research (MPRR) Group, Otto-von-Guericke University Magdeburg, ZENIT Building (65a), first floor, east wing, Leipziger Strasse 44, 39120 Magdeburg, Germany. E-mail: walter.schubert@med.ovgu.de*

a *W. Schubert*
Molecular pattern recognition research group, O-v-G-university Magdeburg, Germany

b *W. Schubert*
International faculty, Max-Planck (CAS-MPG) partner institute for computational biology, Shanghai, China

c *W. Schubert*
Human toponome project, TNL, Munich, Germany

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

others (Lehmann Brothers, 2001), and this reminds us that large molecular systems present in tissues have emergent properties, which cannot be understood solely by concepts such as "organisms are nothing but bags of genes" (Van Regenmortel, 2004).

In earlier studies, we have learned that the co-mapping of dozens to hundreds of molecular components in human tissue reveals highly ordered combinatorial molecular structures that are specific traits for different cell types, tissues, and diseases (Schubert, 1996, 2003; Schubert *et al*., 2006, 2008, 2009, 2012; Friedenberger *et al*., 2007; Bode *et al*., 2008; Bhattacharya *et al*., 2010). However, when the number of co-mapped molecular components, for example, proteins in membranes, falls below 15–49, this order is not seen anymore, and specificity/selectivity features disappear (Schubert *et al*., 2006; Schubert, 2010). This observation revealed that molecular systems *in vivo* generate specificities and selectivities of actions by large supramolecular assemblies, rather than by single molecular species. As shown by imaging cycler (IC) microscopy, such assemblies in cell surface membranes can be controlled by so-called lead membrane proteins that drive tumor cells into a status of metastasis/migration-competence through rearranging proteins in their cell surface membrane, rather than by up or down regulating their abundance (Schubert *et al*., 2006; Schubert, 2010). These observations show that the detailed study of such subcellular rearrangements of proteins and supramolecular assemblies by imaging can be essential for understanding how such molecular networks establish and maintain abnormal cell behavior in disease. Such a topologically determined subcellular rearrangement of proteins cannot be detected by large scale *ex vivo* expression profiling or genomics studies, and such mechanisms appear to be frequent, as suggested by studies in several human cancers and skin diseases (Schubert *et al*., 2009, 2006, 2012; Bhattacharya *et al*., 2010). Moreover, accumulating evidence indicates that mouse models of human diseases cannot replace direct functional insights derived from human tissues (Seok *et al*., 2013). The likely reason is that subcellular arrangements of protein networks in mice differ from those in humans. This places severe limitations on current investigations of molecular systems *in vivo*, for example, in cancer, autoimmunity, and Alzheimer's disease as well as for solving fundamental problems linked to the mechanisms of tissue-specific maintenance and homeostasis and its perturbation in disease.

The present review describes the ability to directly image large supramolecular assemblies that are the fundamental units for the spatial coding of function in eukaryotic cells and tissues. Direct large scale imaging of supramolecular assemblies of biomolecules in the same tissue section or cell is an essential component of a project to decode the toponome, this term being defined as the spatial network code of biomolecules in morphologically intact tissues and cells (Schubert, 2003). The resulting data provide a direct link between *in vivo* biological systems and the concept of emergentism required for decoding emergent properties quantitatively. Special reference is given to IC microscopy, which is an automated device that systematically runs cycles of molecular recognition processes for visualizing higher order functionalities at the level of supermolecules.

SUPRAMOLECULAR ASSEMBLIES AND THE TOPONOME

In his editorial article highlighting a series of studies addressing "supramolecular" research, Turro (2005) stated that this term "invokes a chemistry beyond the molecule": Although the term supramolecular assembly or supermolecule is now widely used in the chemical, biological, and materials literature, a precise definition is difficult to pin down. One way to obtain some insight into the term is to define a molecule as an assembly of atoms that is held together by relatively strong intramolecular bond, and to define a supermolecule as an assembly of molecules that is held together by relatively weak intramolecular bonds. This definition of supramolecular chemistry implies that the successful idea of structure that has been so successful for describing molecules will also be successful for describing supermolecules (end of original text from Turro, 2005). The present article focuses on supermolecules composed of large multi-protein assemblies, which can be imaged reproducibly in tissues and cells, because they are held together by interactions between their molecular components or by an external structure acting as a scaffold.

Within the context of the cell or tissue, the essential questions to be answered are the following: What do supermolecules, composed of many different protein species, look like?; how do they change under disease conditions?; and how can thousands of such supermolecules be imaged, or captured, as large spatial networks driving and controlling disease? Finally, can this type of extreme imaging be linked to genome sequencing?, and if successful, can the precise functional interrelationship be pinned down? To address these issues systematically, a research field analyzing the toponome has been established (for review, see Schubert *et al*., 2012; Schubert, 2013). It is driven by the belief that chronic diseases can only be understood when the different organizational levels of the cell (genome and toponome) can be directly observed within preserved structures of tissues.

The term "toponome" is defined as the spatial network code of proteins and other biomolecules (e.g., carbohydrates and nucleic acids) in morphologically intact cells and tissues (Schubert, 2003; Schubert *et al*., 2012). Aspects concerning the term toponome, described in Sections Supramolecular Assemblies and the Toponome and Structural and Functional Units of the Toponome, have been discussed previously (Schubert, 2013). This term is not another "omics" term, because it is derived from the ancient Greek nouns "τόπος" (place, position) and "νόμος" (law) and, hence, is purely descriptive (Schubert, 2003; Schubert *et al*., 2012). In a strict sense, the fusion of these two terms describes the biological reality of function that is encoded by rules of subcellular topologies, which are revealed by the direct visualization of toponome structures (Schubert *et al*., 2012). It indicates that a cell, which organizes this network of biomolecules, follows topological rules enabling coordinated interactions between its molecular components (Schubert, 2007, 2010). In a cell, or in the extracellular matrix, every single molecular component of such an interaction must be at the right time point at the right concentration and at the right subcellular location so that a specific molecular network can be formed. This interaction can take place either on the basis of strong or weak physical associations of biomolecules, as well as indirectly, by means of diffusible molecules binding to other biomolecules, such as proteins present at distinct locations. Hence, biomolecular networks are characterized by a non-random spatial context of their molecular elements. Consequently, any biomolecular network exerting a concrete cellular functionality obeys rules of topologically defined assemblies of biomolecules—a spatial code (toponome) enabling a directed action, such as a specific information flow across a pathway. The toponome contains the code of conduct exerting these functionalities. Toponomics is a discipline in

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

systems biology, cell biology, and histology, concerned with the study of the toponome of organisms. The human toponome project comprises the complete decoding of the human toponome of 20 000 different proteins for many cell types, tissues, and diseases (Cottingham, 2008) (www.huto.toposnoms.com). It is based on the IC principle (Schubert, 1996), variations of which are termed multi-epitope-ligand-cartography (Schubert, 2003; Schubert *et al*., 2006) or toponome imaging system (TIS) (Friedenberger *et al*., 2007), representing different stages of technology development.

Analyzing biomolecular systems directly in one particular tissue section by direct imaging of their molecular components implies that a multitude of distinct supramolecular assemblies, formed by the system *in vivo*, can be captured at any given subcellular data point and in hundreds/thousands of cells in their natural environment simultaneously. The resulting data sets, obtained by functional super-resolution microscopy TIS, contain the modes and rules necessary to quantify the relational (emergent) properties possessed only by the system as a whole (where the modes are the different geometric features of supermolecules and their spatial interrelationships).

As described in the following, the toponome obeys hierarchical rules of assembling proteins as large structures, which are predictive traits of biological functions, for example, controlling cell polarization (Schubert *et al*., 2006) and often has periodical characteristics (Schubert, contribution and discussions during the thematic research school on advances in systems and synthetic biology, Nice, France, 25–29 March 2013, details to be published elsewhere). These quantitative data are in line with many concepts of ontological emergence, in particular with Paul Humphreys' "emergence as a fusion" (fusion emergentism), synoptically discussed in O'Connor and Wong (2012).

IMAGING CYCLERS CAN BREAK THE SPECTRAL RESOLUTION LIMIT AND THE RESOLVING POWER OF EPIFLUORESCENCE MICROSCOPY

Historical aspects: why break the spectral resolution limit?

The first demonstration that the spectral resolution limit in fluorescence microscopy can be broken (Schubert, 1990) was followed by a detailed description of a corresponding automated method for the subcellular quantification of high-dimensional protein combinations (Schubert, 1996) allowing a systematic development of the IC principle. It led to many technological and conceptual developments and biological insights, some of which are featured in Figure 1.

The only way to assess the combinatorial molecular structure of large molecular systems *in situ* is to co-map dozens, hundreds, or thousands of different proteins/biomolecules in one and the same morphologically intact fixed cell or tissue section. The only basic tool that can be potentially used for this task is fluorescence microscopy. However, to capture the high-dimensional combinatorial molecular organization of the toponome in one and the same subcellular structure, many more proteins must be co-mapped simultaneously in a cell or tissue than the number of wavelengths available within the spectrum of visible light (between approximately 300 and 700 nm). How can this limit be overcome? This question was answered experimentally, when—during my work in neuromuscular diagnostics (1984–1988)—I saw many patients with a disease called polymyositis characterized by T lymphocytes that invade skeletal muscle tissue, thereby displacing muscle fibers so that these fibers rupture, and the patients experience a progressive chronic

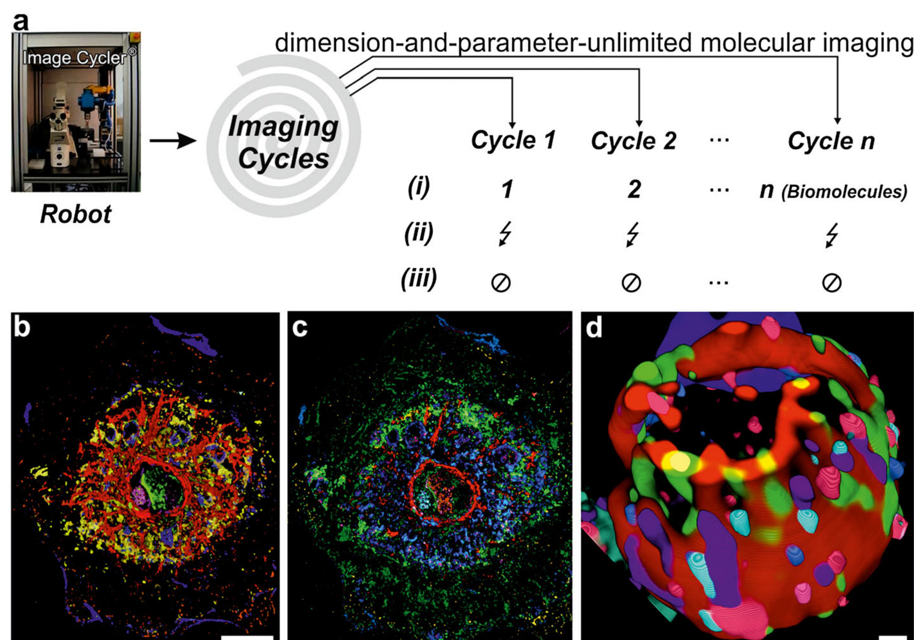


Figure 1. Principle of imaging cycler microscopy (a) and resulting featured toponome maps; (b) cover image from (Schubert *et al*., 2006): over 7000 protein clusters in a single human liver cell; (c) corresponding characteristic multi-protein assembly-toponome map shown in research highlight referring to (Schubert *et al*., 2006) (text of this highlight is found in Abbot A. Nature, 443, 609, 2006); (d) cover image from (Friedenberger *et al*., 2007) showing a cell surface protein cluster network of a single human peripheral blood T lymphocyte with multiple supermolecules differentially built by 27 distinct cell surface proteins: each color represents a unique protein combination (see also Figure 2). Corresponding rocking video at (http://www.nature.com/nprot/journal/v2/n9/ extref/ nprot.2007.320-S9.mov); deep insight-functional super-resolution video (journey across the cell toponome) to illustrate all organelle molecular imaging of the liver cell shown in (a and b) is given in (http://www.toposnomos.com/moviemaster.php?vid=leberzelle_video_black). Bars: b and c, 10 μm; d, 1 μm.

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

paraparesis of their proximal muscles. These observations suggested that a classical, cell-mediated attack against muscle tissue was not taking place, as generally assumed, but that a T cell invasion, driven by an abnormal set and arrangement of cell surface proteins on the T cell surface was exerting a strong, differential adhesive force for muscle tissue invasion. To address such a possibility for many different protein species on a single cell level, it seemed that dimension-unlimited molecular imaging using cyclical fluorescence tagging offered a solution for breaking the spectral limit of fluorescence microscopy (Schubert, 1990, 1996). The experimental basis of this theory relies on large probe libraries ($n = 20$, 50, 100, 1000, etc., different probes), in which every probe, binding specifically to a given moiety or biomolecule *in situ*, is conjugated to one and the same dye, for example, fluorescein-iso-thio-cyanate (FITC). Usage of more than one dye per cycle is also feasible. The technique then requires that a caged robot runs a series of automated and preprogrammed repetitive cycles of (i) probe-dye incubation on the object table (stage) of an epifluorescence microscope; (ii) imaging the resulting signal; and (iii) inactivation of this signal by means of soft fluorescence bleaching (or inactivation using an appropriate agent) (reviewed in Schubert *et al*., 2012). In other words, one specific fluorescence channel, for example, for FITC, of an epifluescence microscope is used to create N such cycles thereby creating N quasi-multi-channels per pixel. Each of these quasi-channels transmits only one particular biological information almost without noise, because only FITC is used for its transmission without being compromised by other fluorescence channels or signals. For example, 100 quasi-channels having been generated by 100 distinct repetitive cycles, together transmitting 100 different unique biological information per pixel (Figure 5(c)), will result in a maximum possible number of biologically informative quasi-channels summing up to 100 048 576 in one digital image. This number can be understood as the “biological reading frame” to quantify the transmittable amount of biological information. This process of collecting unique biological information per pixel can be repeated for a larger number of different optical planes across a cell or a tissue resulting in three dimensional (3D) images showing many distinct multi-protein assemblies. Figure 2 illustrates this technique using cyclical data acquisition for 3D toponome mapping. Additional material and experimental proof has been described in detail (Schubert *et al*., 2006). The biophysical laws underlying this method are based on mechanisms described by the so-called “Venice” and the “sunlight” hypotheses of dimension-unlimited and parameter-unlimited molecular imaging (Schubert, 2006, 2012a, 2013). An elaborate system of hydrous channels (compartments) present inside cells and tissues can be used for quasi-unlimited repetitive and controlled diffusion of dye-conjugated proteins (such as antibodies) for molecular recognition and labeling processes (“Venice” hypothesis), whereas these proteins, after their diffusion, are imaged at their binding sites by excitation at epifluorescent stray light conditions close to the characteristics of natural sunlight on the surface of the earth (100 000 lm/m^2) known not to inhibit or seriously alter protein structure and activities in cellular systems (“sunlight” hypothesis). This implies that a large number of imaging cycles, as described earlier over time (hours, days, or weeks) using epifluorescence excitation and bleaching at 488–530 nm wavelengths, do not alter the molecular structures of a tissue or a cell sample. This has been experimentally verified over many years using several generations of corresponding imaging robots (multi-epitope-ligand-cartography, and TIS) together referred to as imaging cyclers® (www.huto.toposnomos.com) (Schubert, 2006, 2007, 2012a; Schubert *et al*., 2006).

Historical aspects: *in situ* demonstration that myogenic cells are generated by transdifferentiation of adult endothelial cells. The foundation of molecular systems histology

A first new insight by using this method came—unexpectedly—from a cyclical nine-molecular parameter imaging of a human muscle tissue section. In this imaging procedure, combinatorial molecular features were seen allowing the conclusion that muscle stem cells can be directly formed through transdifferentiation of adult endothelial cells, when the latter cells leave the capillary network surrounding the muscle fibers and invade the basal lamina (BL) muscle cylinder during muscle regeneration. During this invasion, these endothelial cells undergo a transdifferentiation process leading to a mononuclear muscle phenotype, fusing with ruptured fibre ends to enter the process of local muscle fiber repair (Schubert, 1992). These data explained that Neural Cell Adhesion Molecule (NCAM)/Leu19-expressing cells inside the endomysial tube observed earlier by two-parameter laser scan microscopic imaging (Schubert *et al*., 1989) are actually cells that express both the endothelial and muscle regeneration stem cell phenotype. These data, which were the first to show that stem cells can be formed by an adult highly differentiated cell type in an area of tissue damage, were obtained through simultaneous observation of many different molecular cell surface parameters by cyclical imaging. This type of formation of myogenic cells was later confirmed by other methods (De Angelis *et al*., 1999; Zheng *et al*., 2007). Today, stem cell research is an established active field. Together, this historical example shows how the deepening of our view of cellular differentiation processes by applying hypothesis-free IC protein co-mapping leads to unexpected new insights into biological processes taking place *in situ/in vivo*, which then can initiate new directions in biological research. More important, after having explored the described relationships, we know today that—if one observes a mononuclear cell expressing both myogenic and endothelial markers inside the endomysial tube—these cells are derived from the local endothelial capillary network and are in a state of transdifferentiation to form a myogenic cell for muscle repair. Hence, the insight, gained from IC studies, has led to a supramolecular and subcellular feature that is predictive for (i) the origin; (ii) the functional status; and (iii) the fate of this cell, and such data are therefore functionally predictive. Similar conclusions can be drawn from the many high-dimensional IC studies that have been performed subsequently.

Further insights

Further biological validation of this principle was given in many studies addressing problems in biology and medicine (reviewed in Schubert *et al*., 2012). For example, toponome studies have discovered a new target protein in amyotrophic lateral sclerosis by hierarchical protein network analysis, a finding that has been confirmed by a mouse KO model. It has uncovered a lead target protein in tumor cells that controls cell polarization/metastasis (Schubert *et al*., 2006; Schubert, 2010), and it has found a new target protein that controls chronic neuropathic pain, a finding that has also been confirmed by an independent KO mouse model (Schubert *et al*., 2006). Current research surprisingly showed that toponome fingerprinting of peripheral human blood

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

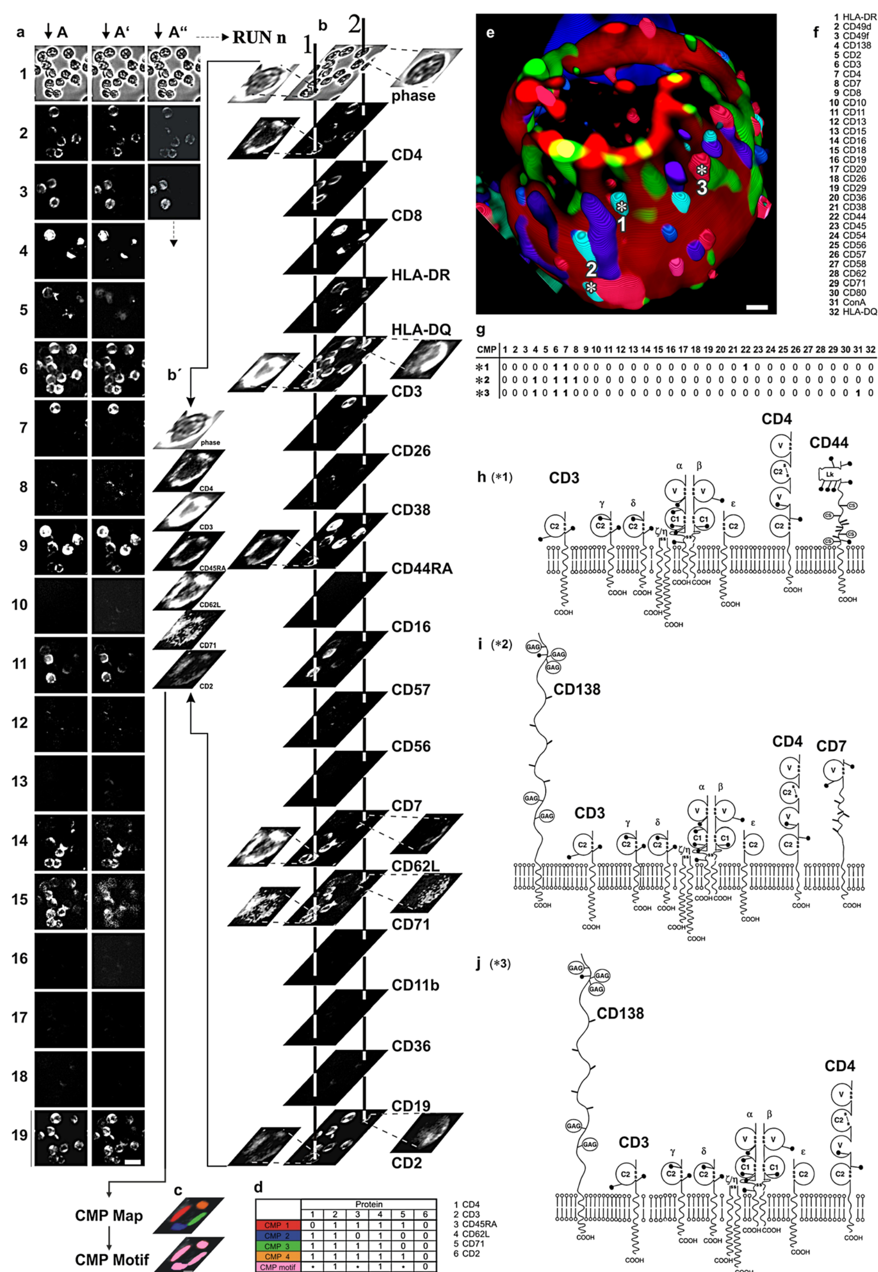


Figure 2. Cyclical assessment of many protein signals in one sample and spatial digital imaging of multi-protein clusters on the cell surface. An example of the cyclical multimolecular imaging procedure (by a robotic imaging cycler, IC) on peripheral blood mononuclear lymphocytes (PBML) is shown to obtain a two-dimensional (a–d) and a three-dimensioal toponome map of cell surface multi-protein clusters (e) including an example of their functional annotation (g–j). (a) Altogether, nine IC cycles, with two dye-conjugated probes/antibodies per cycle (FITC, phycoerythrin (PE) as dyes), were run to label 18 different cell surface proteins in one fixed cell sample. The labeled proteins are specified on the right of (b, vertical list). (A, A′, and A″): The same probe library was used to run three repetitive IC cycles on the identical cell sample (A through A″), whereas the sequence of cycles remained unchanged (so-called repetitive forward runs). Specificity of protein tagging at any location and lack of sterical hindrance of antibody binding during the IC procedure is verified by (i) aligning each signal horizontally (A through A″) and (ii) quantifying correlated signal intensities of the resulting pixel data set by means of mathematical methods: Note—by comparing A, A′, and A″ (repetitive forward IC runs)—that signal locations are identical from A through A″ (horizontal panel of images), whereas the signal intensities decline because of progressive saturation of the corresponding antibody binding epitopes. Such sets of routine IC validation procedures, involving repetitive forward, inverted, and permuted IC runs provide further evidence for quantitative precision of IC. It is part of the so-called logic high-end calibration procedure of IC. (b) Illustration of the process of toponome mapping by depicting two cells (b, vertical lines 1 and 2; cells magnified and aligned in b′). Overlay and alignment is used to set thresholds for each fluorescence signal (d, expert-based or automated) in order to identify regions of multi-protein clustering (characteristic multi-protein assemblies (CMPs)) and the corresponding CMP motif in (c), color decoding list in (d) (for details, see Schubert *et al.*, 2006). (e) 3D toponome map of a single CD4-PBML obtained by optical sectioning during TIS imaging: 32 TIS cycles were run at each out of 20 different optical planes across the cell probe/antibody library in (f). (g) CMPs 1–3 are extracted from several thousands of CMPs per this cell in total to illustrate which proteins (g, top line) are differentially associated as single protein clusters (CMPs) on the cell surface (e, asterisks 1–3 correspond to CMP 1–3 in g). (h–j) Illustration of the supramolecular assemblies (CMPs) revealed by IC in (e, asterisks 1–3) by using a spatial model of the single co-mapped transmembrane proteins (h–j, corresponding to asterisks 1–3 in e, respectively, with kind permission from Springer Verlag). Bar: e = 1 μm. For higher resolution of (h–j), see Supplementary Figure 1.

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

lymphocytes is able to detect progressive neurological disease 5 years prior to its clinical onset (Schubert W, to be published), and that a monogenetic disease can be successfully treated by conversion of the genotype specific toponome code into a normal one by using a small non-toxic drug (Ruonala M. *et al*., personal communication). Together, these technological, biological, and clinical validations show that the IC approach is able to unravel subcellular disease mechanisms and to find novel drug targets and biomarkers directly at the target sites of disease inside human tissue from biopsies and/or surgical material (Evans *et al*., 2012).

STRUCTURAL AND FUNCTIONAL UNITS OF THE TOPONOME

When large numbers of different protein species are co-localized in a compartment of a cell or tissue section, it is seen that every multi-protein assembly obeys a unique protein colocation and anti-colocation code, formally termed a combinatorial molecular phenotype, briefly, a characteristic multi-protein assemblies (CMPs) (Figure 3). Hence, any biologically determined large multimolecular assembly—a specific network of distinct biolomolecules—can be quantified as a specific CMP. Every data point in a cell or tissue expressing this particular CMP has an x/y/z coordinate, so that the molecular composition, subcellular topology, and frequency of any given CMP in a toponome data set can be exactly quantified. Moreover, the spatial relationship of distinct CMPs can also be quantified by using mathematical methods of topology to search for hidden spatial rules of functional linkages of different CMPs able to reveal a toponome code of biological systems. In this context, we understand the term topology as defined by JB Listing: the laws of connectivity, mutual position, and succession of points, lines, surfaces, and solid bodies, and their parts or aggregates in space, apart from their measure and proportion (Listing, 1847). An interesting but consistent finding in toponome analysis is that many different CMPs can have one, or more than one, biomolecule in common. These molecules are called lead molecules. Then, formally, a so-called CMP motif exists, which can be described by using a three-symbol code (Schubert, 2003; Schubert *et al*., 2006; Schubert, 2007): (i) lead molecule(s); (ii) molecules that are variably associated with the lead molecules; and (iii) absent molecules (anticolocated molecules). Together, the toponome is composed of subunits (functional entities) on different scales, referred to as CMPs and CMP motifs (Figure 3) (Schubert, 2003). Both the protein hierarchy and the lead proteins detectable within these structures can be a key to functional predictivity (e.g., prediction of cell behavior and function, see Historical aspects: in situ demonstration that myogenic cells are generated by transdifferentiation of adult endothelial cells. The foundation of molecular systems histology Section) and hence, therapeutic efficiency when an agent blocking a lead protein interferes therapeutically with an abnormal functionality. Lead proteins, detected by protein co-mapping on the surface of tumor cells can be first-order key target molecules, because they control both the subcellular topology and the function of large cell surface protein networks, as revealed by the toponomics-derived detection of the Aminopeptidase Polarization Control Network (APOCON). If the detected cell surface associated lead protein (CD13 = aminopeptidase N) is blocked or inhibited by a small molecule or an antibody, the corresponding protein network, which is arranged along the cell surface membrane as an array of CMPs, completely disassembles. This leads to the loss of the tumor cell's ability to enter the explore state from the spherical state. The tumor cell is then inhibited to migrate/metastasize (Schubert *et al*., 2006; Schubert, 2010) (Figure 4). This observation indicated that the detection of lead proteins *in situ* (rather than by *ex vivo* large scale expression profiling) can be an essential first step in developing efficient therapies, by using the hypothesis-free toponome decoding approach (for details, see A set of observations Section). Toponomics has thus the ability to detect relevant lead proteins in any type of human tissue, by using a variety of threshold-based and non-threshold-based methods to capture the high-dimensional combinatorial molecular code of protein systems in millions of subcellular data points simultaneously. For example, it was possible to detect more than 2000 different multi-protein assemblies in a single tissue section of prostate cancer (Schubert *et al*., 2009) and more than 5000 different multi-protein assemblies in colon cancer *in situ* (Bhattacharya *et al*., 2010). This has shown that cancer mechanisms are often restricted to subcellular protein rearrangements rather than to the up or down regulation of protein expression (Schubert *et al*., 2009, Schubert, 2013). More important, the toponome study of prostate cancer revealed a cell surface arrangement of supermolecules that is exclusively seen in cancer cells migrating inside the lumen of prostate acini that made it possible to annotate these structures as functional predictive traits for this type of metastasis (Schubert *et al*., 2009). The corresponding cell surface toponome has two enzymes as lead proteins, which are likely to be useful drug target candidates, selective for this type of metastasis. For a more detailed

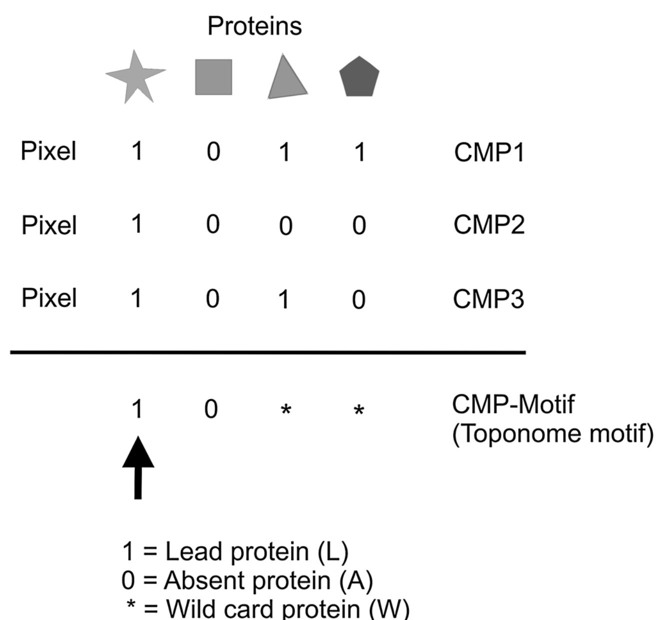


Figure 3. Schematic illustration of the topological and functional hierarchies of proteins within the toponome. In a biological system, such as a cell or tissue, an arbitrary number of distinct proteins (symbols in the top line) can form different combinatorial molecular phenotypes (characteristic multi-protein assembly (CMP) 1, CMP 2…) at one or several subcellular data points. They can have features in common, thereby forming a functional group termed CMP motif, with L, lead protein(s) (common to all CMPs of a CMP motif); A, absent protein(s) (absent in all CMPs of a CMP motif); W, wild card proteins (proteins that are variably associated with the (L), and the (A) proteins of a motif (Schubert, 2003, 2007; Schubert *et al*., 2006)).

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

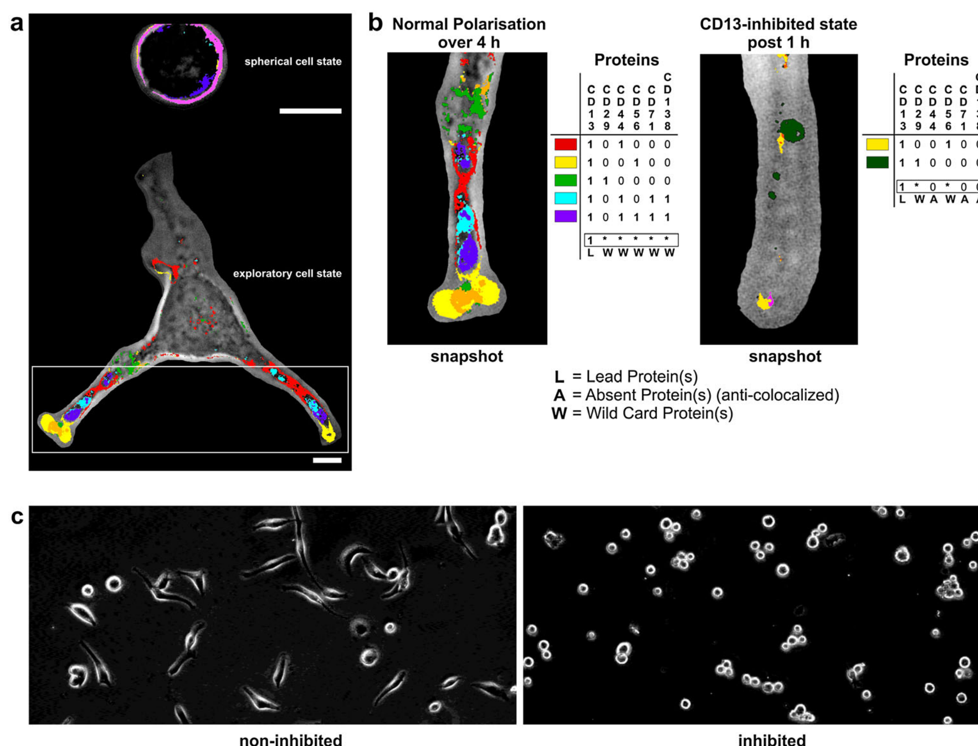


Figure 4. Identification of a molecular network on the cell surface of a single rhabdomyosarcoma tumor cell. (a) Two cell states are seen: spherical state (top) and exploratory state (bottom) that develop from the spherical cell state. On the cell surface of the lower cell (exploratory state), different colors are displayed, which are arranged as an array from the tip of the corresponding cell extensions toward the cell body: yellow, dark blue, light blue, red, and green. These colors indicate different protein clusters, which were mapped by TIS [6]. Note that the relative topology and sequential arrangement of these clusters are precisely corresponding in both lower cell extensions. (b) Left upper part: Detail of one of the cell extensions shown in (a, bottom left) together with the color coding map for the corresponding protein clusters (characteristic multi-protein assemblies (CMPs)). It is evident that these protein clusters have in common the proteolytic enzyme CD13 as lead protein, whereas the other shown proteins are variably associated with the lead protein in the different clusters (present or absent). The common feature of all clusters can be expressed as a three-symbol code [4]: LWWWWW; when the lead protein is inhibited, a disintegration and total loss of most of the protein clusters are observed (b, upper right), and a new common motif occurs: LWAWAA. (c) Functionally, the latter inhibitory motif leads to complete loss of the ability of the cells to enter the exploratory state (compare: left part of c: non-inhibited cells rapidly undergoing cell polarization; right part of c: inhibited CD13: cells remain in the spherical cell state) (Schubert *et al*., 2006; Schubert, 2010). (d) Mean fluorescence intensities of 23 molecules/proteins investigated in the spherical cell of (a, top) and the polarized cell (a, bottom): note very similar, if not identical protein signal intensities/abundances. (e) Total number of CMPs formed by these molecules, giving unique CMP clusters for the two cell states with very little overlap. Bar: a = 2 μm.

outline concerning similar functionally predictive features of toponome structures, see A set of observations Section.

WHAT IS FUNCTIONAL SUPER-RESOLUTION?

A fundamental aspect of the TIS IC technology is its power of combinatorial molecular discrimination (PCMD) per subcellular data point. For example, if a TIS measuring procedure comprises the co-mapping of 100 distinct protein and carbohydrate moieties, the resulting PCMD is 2^{100}, with the prerequisite that every fluorescence signal of any given moiety out of these 100 distinct moieties is registered as being present or absent relative to a threshold (automatically determined or determined by expertises, that is, one bit per protein) (Schubert, 2003; Schubert *et al*., 2006). If the signals are registered without any threshold by using an approach termed similarity mapping (SIM) (Schubert *et al*., 2012), the resulting PCMD is 256^{100} per data point (Schubert *et al*., 2012) (Figure 5(c)). This TIS-SIM approach is entirely performed in real time, allowing the histologist to use it like an electronic microscope in parallel with normal bright field microscopy in routine histological diagnostics to detect pixels, which express the identical protein profiles by being highlighted, while the manually controlled cursor is moving across the tissue (Schubert *et al*., 2012) (for real time videos, see www.huto.toposnomos.com). Figure 5(c) gives an example showing the high PCMD of 256^{100} per data point discriminating between the three layers of the BL of the skin, having a diameter of 120 nm, as known from transmission electron microscopy (Figure 5(a)). This dimension is rather stable for all basal laminae across all human tissues, as well as for the ultrastructurally distinguishable three BL layers: lamina lucida, lamina densa, and lamina fibroreticularis. As shown in Figure 5(c), TIS-SIM can resolve all these three layers by co-mapping 100 distinct proteins at these sites, whereas the resolution of traditional triple fluorescence is four to five times lower at these sites (Figure 5(b)) as compared with transmission electron microscopy (Figure 5(a)). This indicates that ICs overcome both the spectral limit and the resolving power of traditional fluorescence microscopy. In the present example, the resolving power of about 40 nm is sufficient to discriminate the three layers of BL. For note, the BL is the very structure that has developed during evolutionary biology to separate tissues of ectodermal and mesodermal origins, which are a prerequisite for multicellular organisms being able

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

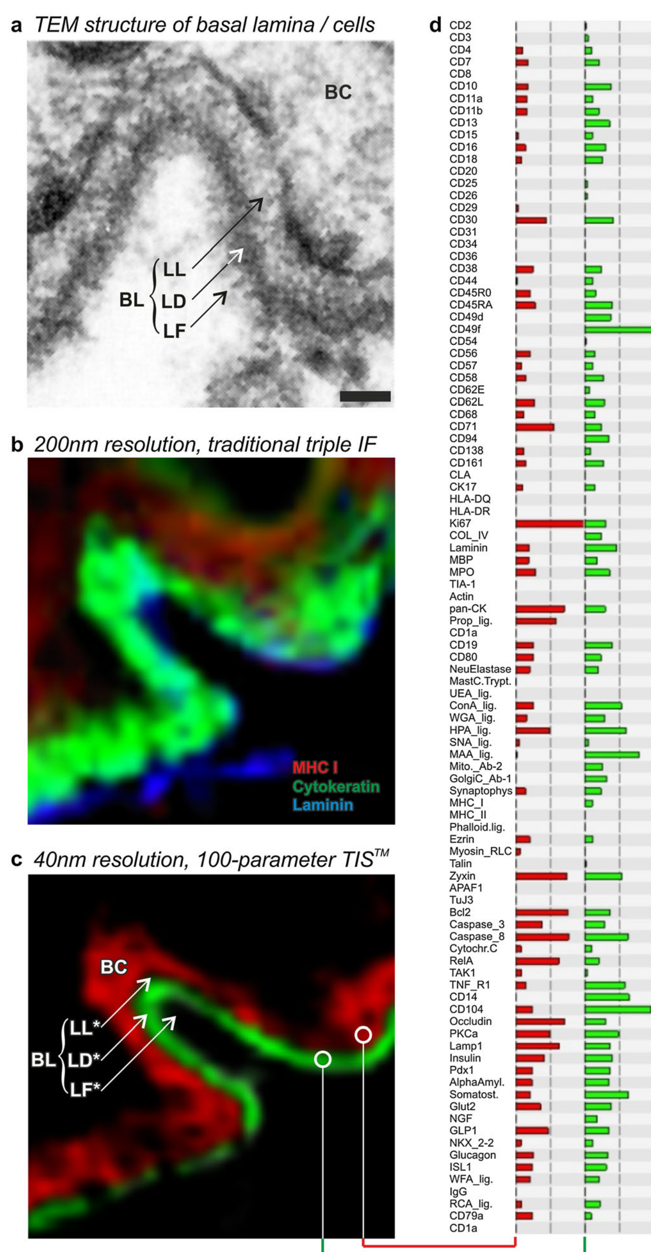


Figure 5. Dermoepithelial junction in human tissue: Visualization of the basal lamina densa as a giant extracellular matrix supermolecule (c, green protein profile). Compare low-resolution traditional triple fluorescence microscopy (b) with 100-parameter TIS image (c) and with corresponding transmission electron micrograph (TEM) (a) indicating that only the functional super-resolution of TIS microscopy (c) can capture the basal lamina densa (LD) as a giant supermolecule at power-glass-like resolution able to distinguish the lamina densa (LD) from the lamina lucida (LL) as well as from the basal ceratinocyte (BC) layer and the lamina fibroreticularis (LF). The protein profile (relational molecular stoichiometry) of the corresponding 100 component supermolecules associated with the LD is seen in (d, cluster 2, green), which is different from that expressed in the BC layer (d, cluster 1). Bar: 50 nm (a). Note that images in (b and c) are simultaneously taken from a 5-μm thick diagnostic frozen human skin tissue section (biopsy material); for (c) similarity mapping (SIM) was used as TIS-SIM non-threshold-based real time imaging (Schubert *et al*., 2012; real time applications in: www.huto.toposomos.com). Note: some molecules are part of cells interacting with the BL.

to develop specialized tissues. One consequence is that most of the separated cell types are not able to cross the BL barrier by migration. This is no longer the case in many cancers, and the breakdown of many selective functionalities of the BL as a boundary surface between blood vessels and tissues often contributes to the pathogenesis of autoimmune diseases, and other chronic diseases, such as asthma and diabetes, arteriosclerosis, or neurodegenerative diseases (Alzheimer's or Down Syndrome having Alzheimer's pathology) (Fassbender *et al*., 2000). In these diseases, the BL is penetrated by immune cells at characteristic sites contributing to, or even being causal for, the disease-specific tissue processes of chronic inflammation and degeneration leading to progressive organ dysfunction. Hence, knowledge and IC decoding of the complete multimolecular assembly code (the toponome code) of the BL is essential for understanding the fundamental biological processes taking place in cancerogenesis and other chronic diseases and during their maintenance, when cells of ectodermal and mesodermal origin escape this code and transmigrate. The use of sequential series of 100-component probe libraries (Figure 5(c)) can result in so-called TIS hypercycles to decode the BL toponomes composed of several thousand biomolecules.

Together, the information content of TIS IC data sets is orders of magnitudes greater than that of *ex vivo* large scale molecular profiling or low content fluorescence imaging involving very few parameters (Schubert *et al*., 2012). This is underscored also by many 3D toponome imaging data, for example, by the toponome of peripheral blood lymphocytes (Figures 2(e) and 1(c)).

Conceptually, and within the context of fluorescence imaging techniques, the IC microscopy principle is a molecular systems microscopy, which aims at understanding and quantifying the supramolecular organization of large biomolecular systems directly *in vivo/in situ*. This primary goal is achieved by molecular parameter-unlimited imaging in one and the same sample to reveal the higher order of molecular organization and the corresponding molecular combinatorics (Schubert, 1996; Schubert *et al*., 2006). This differs from other fluorescence imaging techniques, such as interferometric photoactivated localization microscopy, stochastic optical reconstruction microscopy, or stimulated emission depletion, whose focus is optical super-resolution of one or a few biomolecules (Hell, 2003; Betzig *et al*., 2006; Rust *et al*., 2006). The so-called high-content/high-throughput screening microscopy (Starkuviene and Pepperkok, 2007), which often uses three fluorescence parameters per imaging procedure, also falls in this latter group, which might be labeled as a component localisation microscopy technique. These techniques often use secondary antibodies to improve sensitivity of detection.

One particular aspect that is unique and basic for the IC microscopy is that this methodology relies on conjugated antibodies (or any other affinity reagent), avoiding artifacts that can occur when secondary antibodies are used. This has the important advantage (according to the previously defined goal of this technique) that such direct *in situ* epitope recognition and binding reactions can be cyclically repeated many times until the saturation of a given epitope, present within a supramolecular complex, is reached, which is defined as an endpoint according to the sensitivity of the measuring apparatus applied. This can be performed for any number of successive cyclical imaging procedures of other epitopes expressed by the same protein or by other proteins present in a supramolecular assembly, and this can be repeated for any number of distinct epitopes/proteins for a large number of cycles per one IC run

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

on the same tissue. The resulting series of epitope signal images are then subjected to exact quantitative alignments and analyses (for details, see Schubert, 2003; Schubert *et al*., 2006, including supplementary material). This whole procedure provides an (often ignored) additive effect resulting from the collective staining of individual antigenic specificities that are recognized within a multimolecular complex (Van Regenmortel, personal communication). This overcomes the problems related to the intrinsic polyspecificity of antibodies, which always gives rise to cross-reactivities when only few labeled antibodies are used. These potential cross-reactions are different for each antibody used, with the result that non-specific staining reactions are diluted out by the IC process and therefore masked. This mechanism explains the high selectivity, specificity, and sensitivity of the IC approach. The same mechanism explains why polyclonal antisera give more specific staining than monoclonal antibodies (Richards *et al*., 1975; Van Regenmortel, 2012).

Several reviews and editorials on toponomics have been published, and techniques applying the cyclical toponomics principle of imaging have been reported to address a variety of biological problems in cytometry and cell biology (Bedner *et al*., 2001; Ecker and Tarnok, 2005; Abott, 2006; Ecker *et al*., 2006; Laffers *et al*., 2006; Mittag *et al*., 2006; Murphy, 2006; Cottingham, 2008; Sage, 2009; Eyerich *et al*., 2010; Micheva and Bruchez, 2012). A variety of cellular and disease mechanisms, stem cell features, and target molecules have been identified by cyclical imaging, and several toponomic studies have analyzed skin allergies and the binding sites and activity of a therapeutic agent for inflammatory skin disease (Ademmer *et al*., 1998; Ebert *et al*., 1998; Haars *et al*., 2000; Bonnekoh *et al*., 2008, 2007a, 2007b, 2006; Coste *et al*., 2008; Alex *et al*., 2009; Berndt *et al*., 2010; Ruetze *et al*., 2010; Eckhardt *et al*., 2013; Ostalecki *et al*., 2013). Other aspects of focussed toponome research have been reviewed recently (Gieseler 2013a, 2013b; Hillert, 2013a, 2013b; Khan and Waddington, 2013, Krusche, 2013) including its mathematical description (Dress, 2013a, 2013b; Dress and Epstein, 2013).

SUPRAMOLECULAR SELF-SEGMENTATION OF TISSUES

TIS-SIM detection of skin lamina densa as a giant supermolecule

Toponome imaging system-similarity mapping data show that the lamina densa is a giant supermolecule unique to this BL layer and different from the suprabasal epidermis and the infrabasal dermis (Figure 5). This is shown by the fact that only the pixels belonging to the lamina densa structures are highlighted in the same color and display the same protein profile alongside the whole lateral extension of this band-like structure (Figure 5(d), green profile). The pixel protein profile highlighting the lamina densa is specific and selective for this site, whereas the protein profile for the suprabasal keratinocyte layer is clearly different (Figure 5(c), red profile). Moreover, the highlighted pixels (Figure 5, green profile) extend across several hundred microns, and the protein profile as well as the relative abundance of the co-mapped proteins is not altered across this site. This indicates that there must be a highly controlled stoichiometry of the co-mapped proteins and of associated carbohydrate structures and that the detected structure is a giant supermolecule expressed as a specific trait of the lamina densa. This has several implications for understanding systems of biomolecules found in human tissues *in vivo/in situ* as follows: (i) TIS-SIM or generally speaking ICs are able to detect supermolecules inside cells and tissues at high functional and structural resolution; and (ii) this allows researchers to uncover practically all existing modes and rules of topological and functional organization of protein networks for all biomolecules across human tissues and diseases in sequential arrays of IC hypercycles. (iii) The IC approach can be complemented with genome sequencing. Hence, genome analysis can be combined with the analysis of the genome's downstream "operational level," found in the toponome with its own spatial coding rules. This reveals the interrelationship between genome and toponome structures and functionalities in individual tissues and cells. The complete skin data set obtained by the IC approach used in a published TIS-SIM analysis (Schubert *et al*., 2012) is available (Schubert *et al*., 2006, supplementary video http://www.nature.com/nbt/journal/v24/n10/extref/nbt1250-S14.mov).

Supramolecular order and stochastic disorder of the T cell surface—insight into scales of organization

We have also imaged the supramolecular cell surface structure of human T lymphocytes in the blood, which is regarded as a fluid tissue. These investigations have shown that in both (CD4+ and CD8+) T cell phenotypes, there are CMPs of cell surface proteins belonging to the T cell receptor (TCR) complex, whereas many other proteins do not cluster, as shown by co-mapping of 32 distinct proteins (Figure 2(e and g)). Moreover, we detected clear-cut differences in the frequency and spatial distribution of these CMPs on the cell surface. Whereas there is a dense CMP cluster-to-cluster network on the surface of CD4+ T cells (Figure 2 (e)), CD8+ cells lack this density and display only a few CMPs with no or only a few CMP cluster-to-cluster interfaces (Schubert *et al*., 2012). Both these T cell phenotypes appear to have a "built-in-program" allowing their cell surface biomolecules to cluster differently and giving rise to small variations in the composition of their supramolecular assemblies in the cell surface membrane.

These data represent the first observations of the phenotype-specific supramolecular order of TCR associated proteins in peripheral T cells (i.e., their molecular face). By contrast, imaging single TCR associated molecules reveals highly stochastic and rapid lateral movements of the corresponding molecules without any recognizable clustering (James *et al*., 2011). These observations are not necessarily incompatible and provide new insights in the molecular organization of the T cell surface present at different levels of integrated hierarchies, as suggested by the following thought experiment.

The cell surface cluster of differentiation antigens CD2, CD4, CD45, and CD3 are proteins that have been shown to undergo rapid stochastic movements inside the cell surface membrane without any topological clustering (James *et al*., 2011), although their individual movements could not be traced simultaneously because of methodological restrictions (James *et al*., 2011). This may explain why clustering of these several different protein species was not observed. The absence of any monomolecular clustering (i.e., clustering of many molecules of the same species) therefore does not contradict supramolecular clustering of different protein species as shown schematically in Figure 6. These events may be overlooked, when only one or two different protein species are imaged (Figure 6(a–c); compare with Figure 6 (d)). This interpretation is supported by the fact that the same proteins CD2, CD4, CD45, and CD3 known to undergo rapid stochastic movements in the cell surface membrane

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

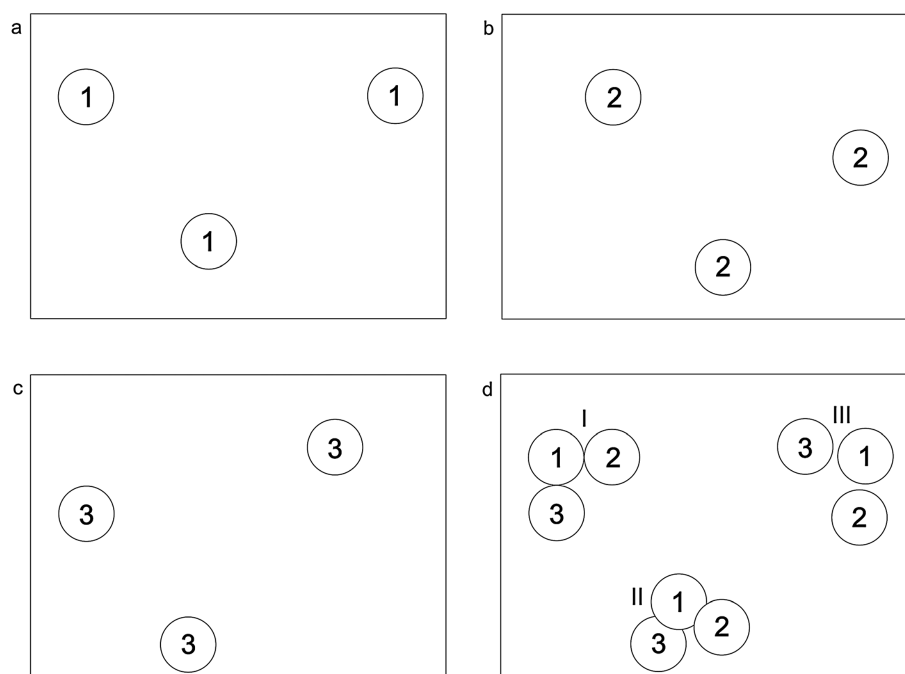


Figure 6. Scheme of stochastic disorder and supramolecular order in the same cell surface membrane. (a–c) Snapshots of three distinct cell surface proteins during stochastic lateral movements in the same cell surface membrane: no clustering is seen. (d) Simultaneous imaging reveals two distinct multi-protein assemblies (1 + 2 + 3) indicating differential adhesion of these proteins at positions (I) and (II) but not at position (III), generating differential patterns of supramolecular order (Figure 2(e)), which is not seen by few parameter imaging. Compare with power law of supramolecular order in Figure 7.

(James *et al*., 2011) indeed have been shown to co-cluster on the cell surface of peripheral blood T cells (Figure 2(c)), thereby forming at least four distinct CMPs as membrane domains having one CMP motif in common (Figure 2(c and d)) (Schubert *et al*., 2006). Such correlations open up new possibilities to unravel the systems biology of cell surface membranes by comparing different microscopic and cell preparation methods.

This means that formation of multi-protein assemblies (Friedenberger *et al*., 2007; Schubert *et al*., 2012) (Figure 2(c–e)) and stochastic rapid movements of the same proteins are probably not incompatible but are essential for allowing an emergent system of biomolecules to work properly on different scales of organization: Extremely rapid movements of cell surface proteins by lateral diffusion ensure that almost any protein species is present everywhere at any time. Because these proteins are likely to exert intermolecular binding forces when encountering each other during their movements and because these forces may be different in different molecular classes, it is likely that proteins with the highest intermolecular binding forces will form the most frequent supramolecular assemblies. Accordingly, supramolecular assemblies of proteins with weaker binding forces will be less frequently seen. In other words, the quantitative TIS images containing exact information on the frequency of supramolecular assemblies (CMPs) at the T cell surface (Friedenberger *et al*., 2007) may provide direct information on a relational (i.e., not absolute) scale of differential binding forces of the proteins under investigation.

Functional super-resolution studies of T cell surfaces in human blood (see Figure 2(e)) are important for analyzing homeostasis, immunosurveillance, and pre-clinical detection of organ-selective invasion in chronic diseases. When this cell surface resolution is not applied, and/or the number of co-mapped proteins on the T cell surface falls below 15, this higher-order information implodes as shown in our observations in human skin (Figure 7). This has turned out to be a rule of thumb for cell surface membranes, when the higher-order molecular organization is addressed.

EMERGING TOPONOME SPACE

A set of observations

A toponome space is the moduli space of toponomes: In algebraic geometry, moduli spaces are understood as spaces of parameters or objects. It can be mapped on a scale ranging from subcellular compartments to cell types, tissue types, organisms,

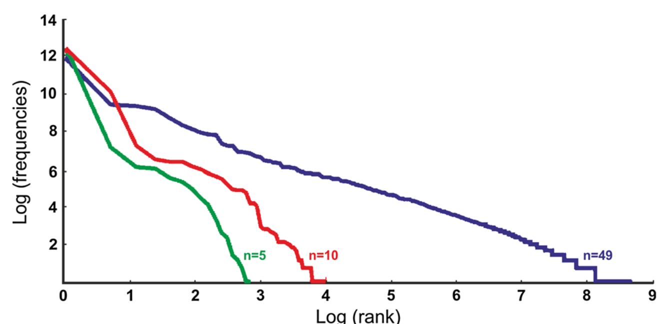


Figure 7. The toponome is a highly organized molecular system, as revealed by power law: Zipf's plot of the relationship between the rank and the frequency of characteristic multi-protein assembly (CMP) motifs in human skin. Data were obtained for 49 simultaneously localized molecules (blue), or subsets of these molecules (red), or five molecules (green). The (log–log) linearity seen with data for 49 molecules is progressively less apparent as fewer molecules are examined: Only linearity indicates high degree of functional organization requiring dimension-unlimited microscopy for its detection (Schubert *et al*., 2006, 2012).

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

and species. A review of toponome data by the end of the year 2007 performed in our lab was based on a toponome analysis of 6.787×10^{18} data points in human cells and tissues. It revealed a total of 12 984 650 distinct cell type-specific, tissue type-specific, or disease-specific CMP (toponome) motifs, of which 120 258 motifs were found to be the most prevalent. This collection of large supramolecular assemblies represent the major modes and rules present in the healthy and diseased human toponome. Subsequent analyses in prostate cancer and colon cancer provided insight into the diversity of CMPs (moduli space) of large supramolecular assemblies of cell surface and surface associated proteins (Schubert *et al*., 2009; Bhattacharya *et al*., 2010). An example of a prostate cancer CMP data set comprising more than 2000 distinct CMPs present in a single tissue section is available (Schubert *et al*., 2009, supplementary material available at: http://pubs.acs.org/ doi/suppl/10.1021/pr800944f). Similar data sets have been published (Friedenberger *et al*., 2007; Oeltze *et al*., 2011). They show (i) the 30 most frequent CMPs (out of 9646 CMPs in total of simultaneously measured peripheral blood mononuclear cells) (http://www.nature.com/nprot/journal/v2/n9/extref/nprot.2007.320-S8.mov); and (ii) various real time visualizations of CMPs in 3D (http://www.computer.org/csdl/trans/tg/2011/12/ ttg2011121882-abs.html).

Together, these data collections suggest that the toponome space is large but not arbitrarily large.

During an analysis of rhabdomyosarcoma (RB) cells, we were able to identify a clear-cut order and hierarchical rules of supramolecular organization of the cell surface membrane. In more than 120 experiments, we found that these cells always used the same higher order of organization under the same experimental conditions and never switched to other patterns (Schubert *et al*., 2006).

To generate a cell surface protein network code enabling the cell to self-organize during tumor morphogenesis, the RB cell organizes many proteins into specific supramolecular protein assemblies (CMPs) arranged along the RB cell extensions, always in the same topological sequence (reminiscent of an array) (Figure 4(a), bottom). This sequence occurs only when spherical RB cells (Figure 4(a), top) elongate to form three cell extensions (exploratory state) preceding the migratory state. This state is reached when one of the extensions is retracted, and a long axis is formed. Analysis of the CMPs shows that aminopeptidase N (APN = CD13) is present in all CMPs, wheres all other proteins are variably associated with CD13 (Figure 4(b), left-hand side). The protein present in all CMPs is defined as lead protein, the variably associated proteins being “wild card” proteins. Following the lead protein hypothesis (Schubert, 2003), inhibition of the lead protein CD13 by antibody SJ1D1 or small molecule RB3014 results in the disassembly of the cell surface CMPs (Figure 4(b), right-hand side) with the consequence that RB cells become unable to execute the transition from the spherical to the exploratory state (Figure 4(c), compare inhibited and non-inhibited states). If the lead protein (CD13) is inhibited during the exploratory state, the RB cells retract all extensions and revert to the spherical state (Schubert *et al*., 2006). Hence, the proteolytic activity of CD13 is essential for the organization both of the intramembrane location and the molecular composition of the identified supermolecules (CMPs). CD13 is an essential protein both for the formation and maintenance of the cell polarization exploratory state. Because the formation of cell extensions depends on the directed polymerization and action of the actin cytoskeleton, one can conclude that the proteolytic activity of CD13 in the cell surface membrane dictates the directed actions of the actin apparatus in the cell. Thus, the actin network is switched on or off, depending on whether CD13 in the cell surface membrane is proteolytically active or not (Schubert *et al*., 2006). This network was labeled APOCON, because APN (CD13) controls the topology and function of this network (see Structural And Functional Units Of The Toponome Section). This gives rise to a model in which the APOCON is resolved as two interacting networks: (i) a planning network (PN) and (ii) an executive network (EN). The planning network is present in the initial spherical state of the cell, where it consists of at least one topologically circumscribed node containing the proteins CD13 (APN), CD44, CD56, CD71, and CD138 (Schubert, 2010). If the functional action of this node is inhibited by blocking APN proteolysis, no cell extension is formed, because the EN (actin actions) is not activated, although the actin binding proteins CD138 and CD44 are present in the membrane. This allows us to conclude that the presence of CD138 and CD44 as cytoskeleton-organizing membrane proteins (Barclay *et al*., 1993) within this node is not sufficient for activating the actin polymerization apparatus of the EN. The cytoskeleton-organizing operations of CD138/CD44 are obviously controlled by membrane APN proteolysis: if APN is not proteolytically active, no supramolecular organization around CD138 and CD44 occurs, and the latter proteins do not activate the intracellular actin apparatus. Moreover, it can be speculated that the whole supramolecular protein cluster of this node including CD56, and CD71, besides CD138/CD13/CD44, is required to initiate the EN, most likely by specific constraints that this cluster may exert on the conformation/activation of CD138/CD44 (for details, see Schubert *et al*., 2006; Schubert, 2010).

Thus, toponome mapping and experimental inhibition of the lead protein(s) are able to reveal functional protein networks and their internal hierarchies. Moreover, the morphogenetic machinery of RB cells generating solid tumor tissue depends on the same protein networks (as previously mentioned), as revealed by CD13 (APN) inhibition (not shown). In agreement with this observation, similar protein networks have been observed in prostate and colon cancer (Schubert *et al*., 2009; Bhattacharya *et al*., 2010).

When a lead protein is found within a hierarchically built protein network of cell surface membranes as a specific feature of a disease, such an observation supports the lead protein hypothesis, which views it as a first-order drug-target candidate probably driving the corresponding pathogenic network and thus the supramolecular disease phenotype. This interpretation is one of the reasons for systematically decoding the human toponome at the target sites in diseased tissue sections by IC and combining this approach with genome sequencing.

Discriminatory power on different scales in disparate systems

Using a metaphor addressing levels of organization in systems and the discriminatory power of subsystems (see seminar “systems biology” for the 12th school year scholars, 2008; and lecture series for bachelor students in computer visualistics “histological and microscopic image information”, OvG University Magdeburg, Germany, Faculty of Informatics, years 2000–2013): it has been suggested that the hierarchical sequential organization of CMPs in RB cells is reminiscent of a written language system in which the basic components (letters) are arranged as higher-order structures (words), which are further arranged into a topologically meaningful sequence of words to

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

form two sentences with different meanings in two different books (Figure 8(a,b)): When the letters are counted and compared and plotted as frequency profiles, almost no difference between these sentences is seen (Figure 8(c)) ("Rowling's Harry Potter and Shakespeare's Hamlet are the same"). However, when words are compared, many of them allow a distinction to be made (Figure 8(d and e)), whereas only the complete topological arrangement of words as sentences uncovers the higher-order meaning ("Rowling's Harry Potter and Shakespeare's Hamlet are not the same") (Figure 8(a and b)). Similarly, when the abundance of proteins as the basic components of the system expressed in the spherical and the exploratory RB cell states (Figure 8(f and g), respectively) are compared as profiles (Figure 8 (h)), these profiles are very similar, if not identical (corresponding to the letter profile in Figure 8(c)). This allows no distinction of the RB cell's spherical and exploratory states (Figure 8(f and g), respectively), from which the protein abundances were derived (similar to the profiles of letters) (compare Figure 8(c)). By contrast, the CMPs of RB cells show many specific differences between the spherical and explore states (Figure 8(i and j), respectively) comparable with the word level (Figure 8(d and e)). However, only the topological arrangement of CMPs as spatial arrays (Figure 8(g)) gives the specific and higher-order cell surface code, which brings about the exploratory state (Figure 8(g), magnified in Figure 9b).

Hence, CMPs and their relational topologies as interlocked structures are the key to decode the cellular function of hierarchically organized supermolecules in terms of their compositional topologies, here referred to as CMPs. Notably, current observations suggest that the simple topological rearrangement of only one or two proteins can be sufficient to drive a CMP network into a pathological functionality, because such rearrangements are able to fundamentally modify the emergent behavior of the corresponding CMP system as a whole. Because any CMP detected by TIS imaging is resolved as a function of a given subcellular topology in the x/y/z direction with corresponding coordinate numbers of the tissue, the important higher-order spatial relationship of thousands of such CMPs can be mapped. This uncovers the pattern of their relational topologies for decoding large spatial molecular networks as a whole. For this purpose, mathematical methods of combinatorial geometry and statistics can be applied to quantify topologies following the definition of Listing, as described earlier (Structural and Functional Units of the Toponome Section). A useful general measure for highly organized molecular system is Zipf's law, which applies when more than about 20 distinct proteins are co-mapped (Figure 7) (Schubert *et al*., 2006). Further information on *ex vivo* big data science and *in vivo* toponome decoding approaches are given in the supplementary text.

Principles of order: supramolecular segregation, periodicity, and predictivity

The large body of observations made with the IC analysis of biological systems (see A Set of Observations Section) is based on studies of mice, rats, and humans (summarized in Schubert, 2013). This reveals the rules that govern supramolecular order in these species, in particular in biological membranes (unit membranes and matrix membranes) and allow the supramolecular segregation of cellular functionalities.

(i) Any set of distinct biomolecular species (a_1, a_2, $a_3 \ldots a_k$), or subset thereof, is segregated by cells into a finite number of supermolecules, each of which is composed of at least two distinct biomolecular species. (ii) Every supermolecule is segregated into a given subregion of a biological structure, (e.g., a membrane), thereby forming a functional domain, whereas any biomolecule

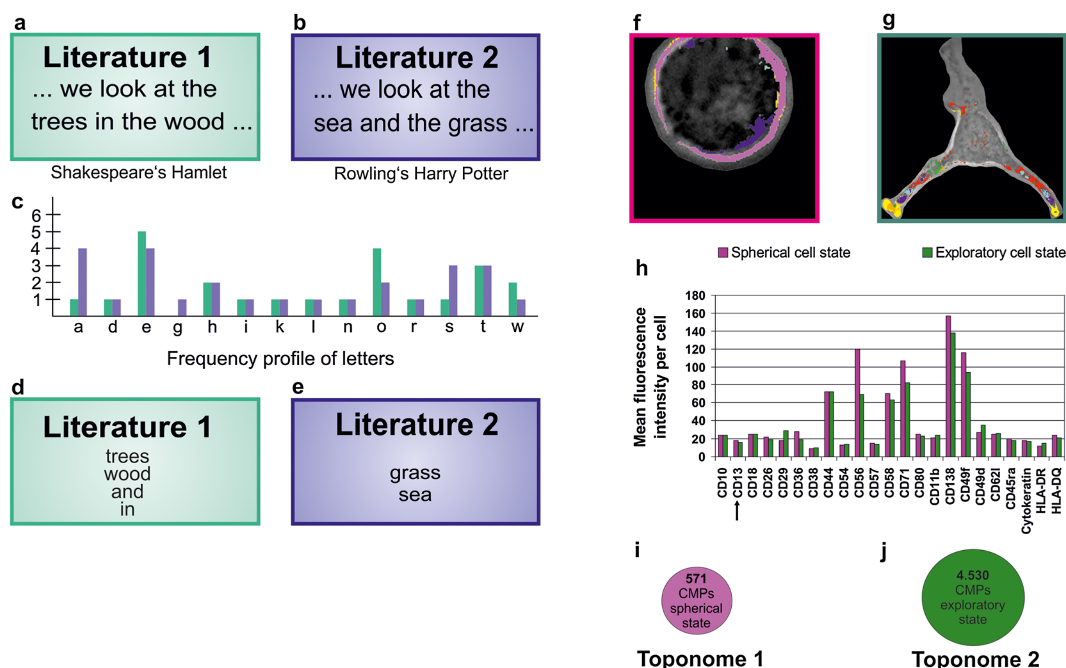


Figure 8. Three integrated levels of systems organization for comparison in toponome and written language. Note the highest discriminatory power between different literatures (a and b) and different cell state specific toponomes (f and g), when topological arrangements of words (a and b) and characteristic multi-protein assemblies (CMPs) (f and g) are compared (see also Figure 9(b)). Lowest discriminatory power is seen in (c) and (h), when the smallest systems components, the letters (c) and the biolomolecules (h), are quantified as profiles. Medium discriminatory power is seen, when words (d and e) and CMPs (i and j) are compared within the corresponding systems (a and b) and (f and g), respectively.

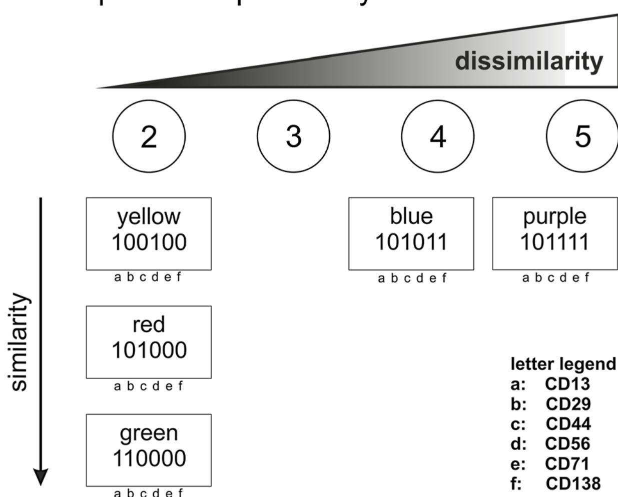
© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

which is part of this supermolecule can be more widely distributed. (iii) At least two distinct, spatially adjoined supermolecules are required to dictate or execute a given cellular functionality.

Any set of supermolecules as defined earlier can be formally sorted as temporal instances of similarities/dissimilarities (*formal, compositional periodicity*), where the highest degree of similarity is present, when the numbers of distinct biomolecules as components of a supermolecule are the same, whereas at least one of these biomolecules, coexisting in a given supermolecule, differs from any other supermolecule of this class (Figure 9, vertical classes 1–3). The degree of dissimilarity is given by the number of distinct biomolecules composing one supermolecule (Figure 9, horizontal line). Hence, a supermolecule with two coexisting biomolecules is different from a supermolecule displaying four distinct biomolecules, and so on. Note that each of the supermolecules is annotated as CMP.

The second type of periodicity, as a specific instance of subcellular topological order, is given by the following rules.

Within an executing network of distinct supermolecules, the spatial order of the supermolecules, classified in the formal periodical system (Figure 9(a)), follows specific functional and spatial rules of the network itself (Figure 9b) (corresponding to the order of words in sentences, see Figure 8(a–b)), which are independent of the similarity rules governing the supermolecules themselves (specific case: *spatial periodicity*, see Figure 9(b)). Hence, the formal periodicity (Figure 9(a)) of several supermolecules (CMPs) does not allow to predict the spatial arrangement of these supermolecules, whereas only simultaneous imaging of these supermolecules can detect their spatial arrangement: Once the latter is detected and functionally decoded, will the imaging detection on its own be sufficient to predict that this spatial arrangement will drive for instance the tumor cells into metastasis competence, as demonstrated in many experiments. Hence, corresponding pattern recognition has important implications in predictive medicine. For example, a large set of thousands of CMPs (supermolecules) (http://pubs.acs.org/doi/suppl/10.1021/pr800944f)—when visualized as a colored mosaic—gives the molecular picture of prostate cancer, having been featured as "molecular face" (Sage, 2009). This leads to the identification of metastasis competent intraductal prostate cancer cells and simultaneoulsy identifies the lead proteins as molecular targets (Schubert *et al*., 2009).



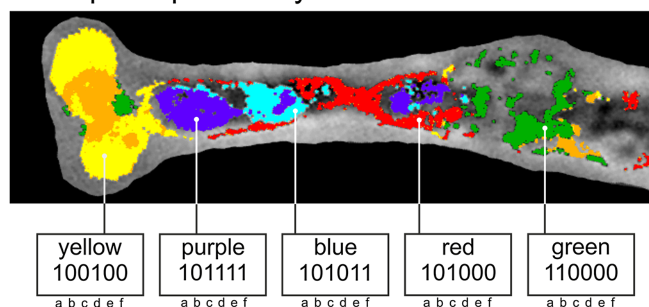


Figure 9. Example of compositional (a) and spatial periodicity (b) of supermolecules (characteristic multi-protein assemblies (CMPs)) in a tumor cell surface, governed by the law of supramolecular segregation. Note that the spatial periodicity (b) is not dictated by the order given in the formal compositional CMP periodicity (a), indicating a high degree of freedom for cells to compose CMPs to much larger, biologically "meaningful", albeit non-random, CMP arrangements.

SUMMARY

The emerging toponome space indicates that natural selection acting on variations of cellular functionalities is built on highly organized spatial segregation of supermolecules. The resulting functional patterns produce specific supramolecular regularities in health and disease. Many defined chronic diseases appear to be driven by disease-specific new supramolecular arrangements (Schubert, 2013) that often generate large supracellular machines, not contradicting the general law of supramolecular segregation.

To understand the many physiological and pathological conditions in morphologically intact tissues, "mapping togetherness," as featured in Figure 1(b) (Abott, 2006), correctly addresses the main tasks of toponomics: Finding on the large scale of biomolecules, cells, and tissues, which biomolecules (in particular proteins) belong together as supermolecules (CMPs), when co-mapped at high subcellular resolution; how the topological hierarchies, as revealed by CMP and CMP motifs, are related to coding functional/dysfunctional cellular actions; and how pathological motifs drive disease mechanisms and how they can be efficiently modified/destroyed by therapeutic intervention, avoiding side effects. The current concept of toponomics follows the existing insight, that systematic mapping of toponome units—CMPs and CMP motifs—leads to a collection of what can be called a toponome dictionary, a collection of functional entities implicated in specific subcellular mechanisms. It is hoped that the increasing number of detected entities will provide an insight into a system of modes and rules (a "grammar") that encompasses all functional actions in cells and tissues. We find it interesting that toponome studies can include simultaneously networks of proteins, carbohydrate moieties, and nucleic acid sequences in order to reveal the intranuclear 3D toponome of single cells (www.huto.toposnomos.com), with the option of mapping the whole 3D organization of the genome on a single cell level. A step toward this goal has been presented recently (Schubert *et al*., 2008). Another interesting possibility is to combine computer modeling of supramolecular assemblies (Heinrich *et al*., 2010) with corresponding toponome data. Such strategies must involve large

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

libraries of antibodies (Dübel *et al*., 2010), cell biological assays, biobanks (Wichmann *et al*., 2011), and individual clinical tissues (Schubert *et al*., 2013).

Some challenging questions remain: Are there functional toponome entities, similar to the "seme" in semiotics, that are small, not further subdividable "meaningful" combinations of few biomolecules that are shared by many CMPs? If this is the case, a small universal unit, co-existing with others, could make up a spatial structure that enables the cellular performance of only one particular functionality. Different combinations of such seme-like elements could also encode other functionalities. Finding such key elements would be fundamental for understanding cellular behavior in terms of selective and specific toponome activities. If such basic rules exist, do they operate in particular locations or are they present everywhere in the cell? Do such rules make it possible to understand genomic alterations? Can a toponome be autonomous and can it be a sensor for external environmental influences that drives cellular disease processes in the absence of genomic mutations? Does the integrated cell have a deep structure and is there a fundamental code beyond the genome that controls all functionalities, including learning?

Acknowledgments

This reseach was supported by the Klaus Tschira Foundation (KTS), the Deutsche Forschungsgemeinschaft (DFG Schu627/10-1), the BMBF (grants CELLECT, NBL3, NGFN2, and NGFNplus), the DFG-Innovationskolleg (INK15) as well as the EU project IMAGINT (Health-F5-2011-259881), and the human toponome project (www.huto.toposnomos.com). I thank ToposNomos Ltd. for providing access to the Imaging Cycler® reference lab and Andreas Krusche and Reyk Hillert for helping with the figures and formatting the manuscript.

REFERENCES

Abott A. 2006. Mapping togetherness (research highlight, referring to Schubert et al. 2006). *Nature* **443**: 608–609.

Ademmer K, Ebert M, Müller-Ostermeyer F, Friess H, Büchler MW, Schubert W, Malfertheiner P. 1998. Effector T lymphocyte subsets in human pancreatic cancer: detection of CD8 + CD18+ cells and CD8 + CD103+ cells by multi-epitope imaging. *Clin. Exp. Immunol.* **112**(1): 21–26.

Alex P, Gucek M, Li X. 2009. Applications of proteomics in the study of inflammatory bowel diseases: current status and future directions with available technologies. *Inflamm. Bowel Dis.* **15**(4): 616–629.

Barclay AN, Birkeland ML, Brown MH, Beyers AD, Davis SJ, Somoza Ch, Williams AF. 1993. *The Leukocyte Antigen Facts Book*. Academic Press: London.

Bedner E, Du L, Traganos F, Darzynkiewicz Z. 2001. Caffeine dissociates complexes between DNA and intercalating dyes: application for bleaching fluorochrome-stained cells for their subsequent restaining and analysis by laser scanning cytometry. *Cytometry* **43**(1): 38–45.

Berndt U, Philipsen L, Bartsch S, Hu Y, Röcken C, Bertram W, Hämmerle M, Rösch T, Sturm A. 2010. Comparative multi-epitope-ligand-cartography reveals essential immunological alterations in Barrett's metaplasia and esophageal adenocarcinoma. *Mol. Cancer* **9**: 177.

Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**(5793): 1642–1645.

Bhattacharya S, Mathew G, Ruban E, Epstein DB, Krusche A, Hillert R, Schubert W, Khan M. 2010. Toponome imaging system: *in situ* protein network mapping in normal and cancerous colon from the same patient reveals more than five-thousand cancer specific protein clusters and their subcellular annotation by using a three-symbol code. *J. Proteome Res.* **9**(12): 6112–6125.

Bode M, Irmler M, Friedenberger M, May C, Jung K, Stephan C, Meyer HE, Lach C, Hillert R, Krusche A, Beckers J, Marcus K, Schubert W. 2008. Interlocking transcriptomics, proteomics and toponomics technologies for brain tissue analysis in murine hippocamapus. *Proteomics* **8**(6): 1170–8.

Bonnekoh B, Malykh Y, Böckelmann R, Bartsch S, Pommer AJ, Gollnick H. 2006. Profiling lymphocyte subpopulations in peripheral blood under efalizumab treatment of psoriasis by multi epitope ligand cartography (MELC) robot microscopy. *Eur. J. Dermatol.* **16**(6): 623–635.

Bonnekoh B, Pommer AJ, Böckelmann R, Hofmeister H, Philipsen L, Gollnick H. 2007a. Topo-proteomic *in situ* analysis of psoriatic plaque under efalizumab treatment. *Skin Pharmacol. Physiol.* **20**(5): 237–252.

Bonnekoh B, Böckelmann R, Pommer AJ, Malykh Y, Philipsen L, Gollnick H. 2007b. The CD11a binding site of efalizumab in psoriatic skin tissue as analyzed by multi-epitope ligand cartography robot technology. Introduction of a novel biological drug-binding biochip assay. *Skin Pharmacol. Physiol.* **20**(2): 96–111.

Bonnekoh B, Pommer AJ, Böckelmann R, Philipsen L, Hofmeister H, Gollnick H. 2008. In-situ-topoproteome analysis of cutaneous lymphomas: perspectives of assistance for dermatohistologic diagnostics by multi epitope ligand cartography (MELC). *J. Dtsch. Dermatol. Ges.* **6**(12): 1038–51.

Coste O, Brenneis C, Linke B, Pierre S, Maeurer C, Becker W, Schmidt H, Gao W, Geisslinger G, Scholich K. 2008. Sphingosine 1-phosphate modulates spinal nociceptive processing. *J. Biol. Chem.* **283**(47): 32442–32451.

Cottingham K. 2008. Human toponome project. *J. Proteome Res.* **7**(5): 1806.

De Angelis L, Berghella L, Coletta M, Lattanzi L, Zanchi M, Cusella-De Angelis MG, Ponzetto C, Cossu G. 1999. Skeletal myogenic progenitors originating from embryonic dorsal aorta coexpress endothelial and myogenic markers and contribute to postnatal muscle growth and regeneration. *J. Cell Biol.* **147**(4): 869–878.

Dress A. 2013a. Information Theory and Toponomics. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 1028–1030.

Dress A. 2013b. Subset Surprisology and Toponomics. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 2025–2027.

Dress A, Epstein D. 2013. Topology and Toponomics. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 2183–2185.

Dübel S, Stoevesandt O, Taussig MJ, Hust M. 2010. Generating recombinant antibodies to the complete human proteome. *Trends Biotechnol.* **28**(7): 333–339.

Ebert MP, Ademmer K, Müller-Ostermeyer F, Friess H, Büchler MW, Schubert W, Malfertheiner P. 1998. CD8 + CD103+ T cells analogous to intestinal intraepithelial lymphocytes infiltrate the pancreas in chronic pancreatitis. *Am. J. Gastroenterol.* **93**(11): 2141–2147.

Ecker RC, Tarnok A. 2005. Cytomics goes 3D: toward tissomics. *Cytometry A* **65**(1): 1–3.

Ecker RC, Rogojanu R, Streit M, Oesterreicher K, Steiner GE. 2006. An improved method for discrimination of cell populations in tissue sections using microscopy-based multicolor tissue cytometry. *Cytometry A* **69**(3): 119–123.

Eckhardt J, Ostalecki C, Kuczera K, Schuler G, Pommer AJ, Lechmann M. 2013. Murine whole-organ immune cell populations revealed by multi-epitope-ligand cartography. *J. Histochem. Cytochem.* **61**(2): 125–133.

Evans RG, Naidu B, Rajpoot NM, Epstein D, Khan M. 2012. Toponome imaging system: multiplex biomarkers in oncology. *Trends Mol. Med.* **18**(12): 723–31.

Eyerich K, Böckelmann R, Pommer AJ, Foerster S, Hofmeister H, Huss-Marp J, Cavani A, Behrendt H, Ring J, Gollnick H, Bonnekoh B, Traidl-Hoffmann C. 2010. Comparative *in situ* topoproteome analysis reveals differences in patch test-induced eczema: cytotoxicity-dominated nickel versus pleiotrope pollen reaction. *Exp. Dermatol.* **19**(6): 511–517.

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

Fassbender K, Masters C, Beyreuther K. 2000. Alzheimer's disease: an inflammatory disease? *Neurobiol. Aging* **21**(3): 433–436.

Friedenberger M, Bode M, Krusche A, Schubert W. 2007. Fluorescence detection of protein clusters in individual cells and tissue sections by using toponome imaging system (TIS): sample preparation and measuring procedures. *Nat. Protoc.* **2**(9): 2285–94. (front cover story).

Gatenby R. 2012. Perspective: finding cancer's first principles. *Nature* **491**(7425): S55.

Gieseler A. 2013a. Cell Membrane Toponomics. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 364–366.

Gieseler A. 2013b. Synaptic toponome. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 2036–2038.

Haars R, Schneider A, Bode M, Schubert W. 2000. Secretion and differential localization of the proteolytic cleavage products Abeta40 and Abeta42 of the Alzheimer amyloid precursor protein in human fetal myogenic cells. *Eur. J. Cell Biol.* **79**(6): 400–406.

Heinrich S, Salo-Ahen OMH, Huang B, Rippmann FF, Cruziani G, Wade RC. 2010. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.* **23**(2): 209–219.

Hell SW. 2003. Toward fluorescence nanoscopy. *Nat. Biotechnol.* **21**(11): 1347–1355.

Hillert R. 2013a. Combinatorial molecular phenotypes (CMPs). In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 440–441.

Hillert R. 2013b. Toponome analysis. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 2188–2191.

Hutchinson L, Kirck R. 2011. High drug attrition rates - where are we going wrong? *Nat. Rev. Clin. Oncol.* **8**(4): 189–190.

James JR, McColl J, Oliveira MI, Dunne PD, Huang E, Jansson A, Nilsson P, Sleep DL, Gonçalves CM, Morgan SH, Felce JH, Mahen R, Fernandes RA, Carmo AM, Klenerman D, Davis SJ. 2011. The T cell receptor triggering apparatus is composed of monovalent or monomeric proteins. *J. Biol. Chem.* **286**(37): 31993–2001.

Khan M, Waddington C. 2013. Clinical aspects of the toponome imaging system (TIS). In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 412–414.

Krusche A. 2013. TIS robot. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 2172–2174.

Laffers W, Mittag A, Lenz D, Tárnok A, Gerstner AO. 2006. Iterative restaining as a pivotal tool for n-color immunophenotyping by slide-based cytometry. *Cytometry A* **69**(3): 127–130.

Lehmann Brothers. 2001. *The Fruits of Genomics*. Lehman Brothers: New York.

Listing JB. 1847. *Vorstudien zur Topologie*. Studien: Göttinger; 811–875.

McCarthy N (editor in chief, no authors listed). 2013. Editorial: thoughts for a new year. *Nat. Rev. Cancer* **13**(1): 1.

Micheva KD, Bruchez MP. 2012. The gain in brain: novel imaging techniques and multiplexed proteomic imaging of brain tissue ultrastructure. *Curr. Opin. Neurobiol.* **22**(1): 94–100.

Michor F, Liphardt J, Ferrari M, Widom J. 2011. What does physics have to do with cancer? *Nat. Rev. Cancer* **11**(9): 657–670.

Mittag A, Lenz D, Gerstner AO, Tárnok A. 2006. Hyperchromatic cytometry principles for cytomics using slide based cytometry. *Cytometry A* **69**(7): 691–703.

Murphy RF. 2006. Putting proteins on the map. *Nat. Biotechnol.* **24**(10): 1223–1224.

O'Connor T, Wong HY. 2012. Emergent properties. In *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*, Zalta EN (ed). (http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/)

Oeltze S, Freiler W, Hillert R, Doleisch H, Preim B, Schubert W. 2011. Interactive, graph-based visual analysis of high-dimensional, multiparameter fluorescence microscopy data in toponomics. *IEEE Trans Vis Comput Graph*. **17**(12): 1882–91.

Ostalecki C, Konrad A, Thurau E, Schuler G, Croner RS, Pommer AJ, Stürzl MA. 2013. Combined multi-gene analysis at the RNA and protein levels in single FFPE tissue sections. *Exp. Mol. Pathol.* **95**(1): 1–6.

Richards FF, Konigsberg WH, Rosenstein RW, Varga JM. 1975. On the specificity of antibodies. *Science* **187**(4172): 130–137.

Ruetze M, Gallinat S, Wenck H, Deppert W, Knott A. 2010. *In situ* localization of epidermal stem cells using a novel multi epitope ligand cartography approach. *Integr. Biol. (Camb).* **2**(5-6): 241–249.

Rust MJ, Bates M, Zhuang X. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**(10): 793–795.

Sage L. 2009. The molecular face of prostate cancer. *J. Proteome Res.* **8**(6): 2616. (editorial to Schubert et al 2009)

Scannell JW, Blanckley A, Boldon H, Warrington B. 2012. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11**(3): 191–200.

Schubert W. 1990. Multiple antigen-mapping microscopy of human tissue. In *Excerpta Medica*, Burger G, Oberholzer M, Vooijs GP (eds). Elsevier: Amsterdam. *Adv Anal Cell Pathol*; 97–98

Schubert W. 1992. Antigenic determinants of T lymphocyte alpha beta receptor and other leukocyte surface proteins as differential markers of skeletal muscle regeneration: detection of spatially and timely restricted patterns by MAM microscopy. *Eur. J. Cell Biol.* **58**(2): 395–410.

Schubert W. 1996. Automated device and method for measuring and identifying molecules or fragments thereof. European patent EP 01810428B1 (priority 29.5.1996, issued 2004).

Schubert W. 2003. Topological proteomics, toponomics, MELK technology. *Adv. Biochem. Eng. Biotechnol.* **83**: 198–209.

Schubert W. 2006. Toponomanalyse. In *Bioanalytik*, 2rd edn, Lottspeich F, Engels JW (eds). Spektrum Akademischer Verlag, Elsevier: Heidelberg; 1035–1048.

Schubert W. 2007. A three-symbol code for organized proteomes based on cyclical imaging of protein locations. *Cytometry A* **71**(6): 352–60.

Schubert W. 2010. On the origin of cell function encoded in the toponome. *J. Biotechnol.* **149**(4): 252–9.

Schubert W. 2012a. Toponomanalyse. In *Bioanalytik*, 3rd edn, Lottspeich F, Engels JW (eds). Springer Spektrum: Berlin, Heidelberg; 1140–1151

Schubert W. 2012b. Treatment of chronic diseases. Is there a logic of failure? Dagstuhl Seminar "Structure Discovery in Biology: Motifs, Networks & Phylogenies", Dagstuhl castle, Germany, 15-20 July 2012.

Schubert W. 2013. Toponomics. In *Encyclopedia of Systems Biology*, Dubitzky W, Wolkenhauer O, Cho KH, Yokota H (eds). Springer Science + Business Media: Berlin, Heidelberg; 2191–2212.

Schubert W, Zimmermann K, Cramer M, Starzinski-Powitz A. 1989. Lymphocyte antigen Leu-19 as a molecular marker of regeneration in human skeletal muscle. *Proc. Natl. Acad. Sci. U. S. A.* **86**(1): 307–311.

Schubert W, Bonnekoh B, Pommer AJ, Philipsen L, Boeckelmann R, Malykh Y, Gollnick H, Friedenberger M, Bode M, Dress AW. 2006. Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nat. Biotechnol.* **24**(10): 1270–1278. (front cover story)

Schubert W, Friedenberger M, Bode M, Krusche A, Hillert R. 2008. Functional architecture of the cell nucleus: towards comprehensive toponome reference maps of apoptosis. *Biochim. Biophys. Acta* **1783**(11): 2080–2088.

Schubert W, Gieseler A, Krusche A, Hillert R. 2009. Toponome mapping in prostate cancer: detection of 2000 protein clusters in a single tissue section and cell type specific annotation by using a three-symbol code. *J. Proteome Res.* **8**(6): 2696–707.

Schubert W, Gieseler A, Krusche A, Serocka P, Hillert R. 2012. Next-generation biomarkers based on 100-parameter functional super-resolution microscopy TIS. *N. Biotechnol.* **29**(5): 599–610.

Schubert W, de Wit NCJ, Walden P. 2013. Systems biology of cancer. In *Molecular Biology of Cancer*, 2nd edn, Pelengaris S, Khan M (eds). John Wiley & Sons: New York. (in press); 552–582.

Selkoe DJ. 2011. Resolving controversies on path to Alzheimer's therapeutics. *Nat. Med.* **17**(9): 1060–1065.

Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, Richards DR, McDonald-Smith GP, Gao H, Hennessy L, Finnerty CC, López CM, Honari S, Moore EE, Minei JP, Cuschieri J, Bankey PE, Johnson JL, Sperry J, Nathens AB, Billiar TR, West MA, Jeschke MG, Klein MB, Gamelli RL, Gibran NS, Brownstein BH, Miller-Graziano C, Calvano SE, Mason PH, Cobb JP, Rahme LG, Lowry SF, Maier RV, Moldawer LL, Herndon DN, Davis RW, Xiao W, Tompkins RG. 2013. Genome responses in mouse models poorly mimic human inflammatory diseases. *Proc. Natl. Acad. Sci. U. S. A.* **110**(9): 3507–12. (www.pnas.org/cgi/doi/10.1073/pnas.1222878110)

Starkuviene V, Pepperkok R. 2007. The potential of high-content highthroughput microscopy in drug discovery. *Br. J. Pharmacol.* **152**(1): 62–71.

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.

Turro NJ. 2005. Supramolecular structure and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **102**(31): 10765–10777.

Van Regenmortel MH. 2004. Biological complexity emerges from the ashes of genetic reductionism. *J. Mol. Recognit.* **17**(3): 145–148.

Van Regenmortel MH. 2012. Basic research in HIV vaccinology is hampered by reductionist thinking. *Front. Immunol.* **3**: 194.

Wichmann HE, Kuhn KA, Waldenberger M, Schmelcher D, Schuffenhauer S, Meitinger T, Wurst SH, Lamla G, Fortier I, Burton PR, Peltonen L, Perola M, Metspalu A, Riegman P, Landegren U, Taussig MJ, Litton JE, Fransson MN, Eder J, Cambon-Thomsen A, Bovenberg J, Dagher G, van Ommen GJ, Griffith M, Yuille M, Zatloukal K. 2011. Comprehensive catalog of European biobanks. *Nat. Biotechnol.* **29**(9): 795–797.

World Alzheimer Report. 2011. http://www.alz.co.uk/research/world-report-2011 [06 May 2013]

Zheng B, Cao B, Crisan M, Sun B, Li G, Logar A, Yap S, Pollett JB, Drowley L, Cassino T, Gharaibeh B, Deasy BM, Huard J, Péault B. 2007. Prospective identification of myogenic endothelial cells in human skeletal muscle. *Nat. Biotechnol.* **25**(9): 1025–1034.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

© 2013 The Authors. Journal of Molecular Recognition published by John Wiley & Sons, Ltd.