Original Articles

A Three-Symbol Code for Organized Proteomes Based on Cyclical Imaging of Protein Locations

Walter Schubert*

Molecular Pattern Recognition Research Group, Institute of Medical Neurobiology, University of Magdeburg, Germany

Background: A major challenge in the post genomic era is to map and decipher the functional molecular networks of proteins directly in a cell or a tissue. This task requires technologies for the colocalization of random numbers of different molecular components (e.g. proteins) in one sample in one experiment.
Methods: Multi-epitope-ligand-"kartographie" (MELK) was developed as a microscopic imaging technology running cycles of iterative fluorescence tagging, imaging, and bleaching, to colocalize a large number of proteins in one sample (morphologically intact routinely fixed cells or tissue).
Results: In the present study, 18 different cell surface proteins were colocalized by MELK in cells and tissue sections in different compartments of the human immune system. From the resulting sets of multidimensional binary vectors the most prominent groups of protein–epitope arrangements were extracted and imaged as protein "toponome" maps providing direct insight in the higher order topological organization of immune compartments uncovering new tissue domains. The data sets suggest that protein networks, topologically organized in proteomes in situ, obey a unique protein-colocation and -anticolocation code describable by three symbols.
Conclusion: The technology has the potential to colocalize hundreds of proteins and other molecular components in one sample and may offer many applications in biology and medicine. © 2007 International Society for Analytical Cytology

Key terms: MELK/MELC; toponome; proteome; code; fluorescence microscopy

Principally, a cell or a tissue can establish or alter function by (i) regulation of proteins at the transcriptional or translational level; (ii) modification of proteins, e.g. glycosylation, limited proteolysis, folding, a.s.o.; (iii) spatial protein arrangement forming networks, and (iv) protein translocation and rearrangement. All current postgenomic technologies based on large-scale protein-expression profiling as provided, for instance, by microarrays (1), 2D gel electrophoresis, two hybrid technology (2–5), and others (6,7) require the destruction of cells to extract proteins, protein complexes, or RNA, or establish molecular interaction assays. By averaging over millions of cells, these methods address the first two items (i and ii), but cannot address the arrangement of proteins in intact individual cells or tissue, in which protein regulation, translocation, and spatial arrangement all are integrated into a highly specific, topologically structured protein web that finally exerts all biological functions. Reasonably, the ability to monitor all proteins in one cell in a single experiment has been formulated as a major goal for future functional genomics research (8), cytomics (9), and predictive and preventive medicine (10–12).

Cells assembled in tissues can modify their structural and functional activity to adapt to changing environmental demands or internal and external stress. This facility is of vital importance, and is often characteristically perturbed in disease processes. Sometimes only a few cells, or even a single cell are, of primary importance for histological diagnosis and disease classification (13,14). Understanding the organization of the proteomes of cells in tissues and, in particular, of such single key cells in disease processes in their natural environment in situ, is an important future challenge. A proteomic technology addressing such cells directly must be nondestructive and alignable with routine histologic examination working on light-microscopic scales. Although several methods have been described increasing the number of different parameters observable

*Correspondence to: Walter Schubert, Institute of Medical Neurobiology, Molecular Pattern Recognition Research Group, ZENIT Building, Medical Faculty, Otto-von-Guericke-University Magdeburg, Germany.
E-mail: walter.schubert@medizin.uni-magdeburg.de

Table 1
List of Protein Epitopes Localized by MELK

Antigen	Lokus link symbol	Lokus link name
CD4	[CD4]	CD4 antigen (p55)
CD8	[CD8A]	CD8 antigen, alpha polypeptide (p32)
HLADR	[HLA-DRB1 (ns)]	Major histocompability complex, class II, DR beta 1
HLADQ	[HLA-DQA1]	Major histocompability complex, class II, DQ alpha 1
CD3	[CD3E]	CD3E antigen, epsilonpolypeptide (TiT3 complex)
CD26	[DPP4]	Dipeptidylpeptidase 4 (CD26, adenosine deaminase complexing protein 2)
CD38	[CD38]	CD38 antigen (p45)
CD45ra	[PTPRC]	Protein tyrosine phosphatase, receptor type, C
CD16	[FCGR3A and FCGR3E]	Fc fragment of IgG, low affinity IIIa, receptor for (CD 16) and Fc fragment of IgG, low affinity IIIb
CD57	[CD57]	CD57 antigen
CD56	[NCAM1]	Neural cell adhesion molecule 1
CD7	[CD7]	CD7 antigen (p41)
CD62L	[SELL]	Selectin L (lymphocyte function-associated antigen 3)
CD71	[TFRC]	Transferrin receptor (p90, CD71)
CD11b	[ITGAM]	Integrin, alpha M (complement component receptor 3, alpha; also known as CD 11b (p170) macrophage antigen alpha polypeptide)
CD36	[CD36]	CD36 antigen (collagen type I receptor, thrombosponding receptor)
CD19	[CD19]	CD19 antigen
CD2	[CD2]	CD2 antigen (p50), sheep red blood cell receptor

The corresponding lokus link data base annotations are indicated.

independently in one sample (15–19) and quantitative tissue analysis has been reported (20,21), still far too few proteins can be visually examined in parallel on the level of single cells or single tissue-sections to map the contextual patterns of proteins forming molecular network architectures in structurally preserved samples.

Here, a fluorescence imaging technology termed multi-epitope-ligand-kartographie (MELK) is described using fluorescently labeled antibodies as tag libraries to colocalize a hitherto unaccessible large number of proteins in fixed cells and tissue sections with photonic resolution by using iterative rounds of protein localizations. While aspects of MELK have been described (22) and the principle feasibility to colocalize 17 antigens in one tissue section has been shown (23), the present study goes a step further. By analyzing several tissue compartments of the cellular immune system in parallel for 18 cell surface proteins, the basic features of MELK are described in detail showing that a three-symbol code can be derived from MELK data sets uncovering new tissue domains. This approach permits the mathematical description of the topological organization of (sub) proteomes in single cells or tissue sections, referred to as the toponome. The use of MELK to construct high dimensional toponome maps may have broad implications in the life sciences.

METHODS
Antibodies and Reagents

Monoclonal antibodies (mAb's, Table 1) used in the present MELK procedures are highly standardized and widely used to label cell surface antigens. They were purchased from commercial suppliers (Beckmann-Coulter: CD2, CD3, CD4, CD7, CD8, CD16, CD19, CD36, CD38, CD45RA, CD56, CD57, CD71, HLADR; Becton Dickinson: CD26, CD62L, HLADQ) in conjugated form (PE- or FITC-conjugated). Propidium jodide, purchased from Sigma

(Germany) was used at a dilution of 0.1 µg/ml in PBS to stain nuclei. For the MELK procedures mAb's were diluted at 1 µg/ml in PBS and applied at 20°C for incubations.

Preparation of Cells and Tissue Sections

Frozen tissue blocks from a human palatine tonsil (surgical sample) and a human muscle tissue specimen (diagnostic muscle biopsy) were sectioned in a cryo-microtome. Five-micrometer thick cryosections were prepared according to earlier protocols (24): fixed in ice-cold acetone at −20°C for 20 min, air-dried, and then rehydrated in PBS (pH 7.4) at room temperature. Peripheral blood mononuclear cells (PBMC) were prepared as described (25). Briefly, PBMC were isolated from the blood of a healthy volunteer by a Ficoll isolation procedure, placed on an object slide, air-dried, and then fixed in ice-cold acetone and rehydrated at room temperature as described earlier. Both the PBMC and the tissue sections were incubated with normal goat serum, washed in PBS, and then placed on the object table of the imaging cycler microscope for further processing using MELK (see result section).

RESULTS
Imaging Cycler Set Up

The principle configuration of an instrument performing MELK cycles (see below) on one biological sample is given in Figure 1. The whole set up, termed imaging cycler (IC), follows the MELK protocols for iterative labeling of proteins or other molecular classes in one biological sample using dye-conjugated tags (antibodies or any other affinity tag) without any further secondary labeling step. By collecting nN signals, each signal comprising k bits per pixel, for each of n proteins and each of N pixels, or any other ligands, the IC device has a power of combinatorial discrimination (PCD) of knN. The set up consists of a pip-
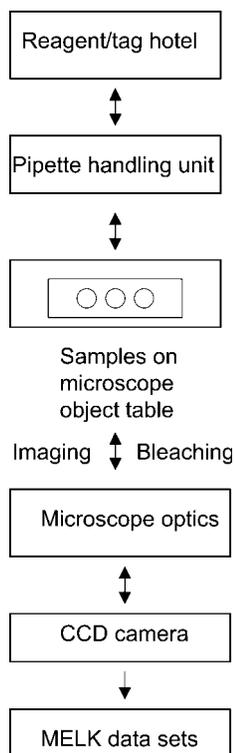
Fig. 1. Schematic illustration of the imaging cycler set up running MELK procedures.

ette handling unit, an antibody/tag, and reagent "hotel," a conventional bright field fluorescence microscope, and a liquid-nitrogen cooled charge-coupled device (CCD) camera for light detection of fluorescently labeled proteins/ligands in cells or tissue sections.

Principles of Multi-Epitope-Ligand-Kartograhie

The methodological principle of multi-epitope-ligand-kartograhie (MELK) is based upon the view that a cell or a tissue is essentially composed of highly conserved aqueous compartments and macromolecular complexes allowing at least medium-sized macromolecules to diffuse freely without significant hindrance. Provided that (i) this condition can be preserved by appropriate fixation methods, (ii) externally applied tag molecules (i.e. antibodies, lectins, a.s.o.) can penetrate cellular membranes; and (iii) the binding of tag molecules to their target structure within the cell is not hindered and can be detected optically, it should be possible to localize a very large number of protein species, or other molecular classes, in a cell by using appropriate tag libraries run by IC's. Given these requirements, a large scale tagging approach of a single cell or a single tissue section should be possible by applying the tag molecules, each of which is conjugated to the same specific dye, sequentially on the sample, provided that after each labeling and optical imaging step, the dye is bleached completely, so that the next tag can be applied, and so forth. The resulting repetitive incubation–imaging–bleaching cycles, permit the collection of large data sets

of protein distribution patterns with cellular or subcellular light microscope-resolution, depending on the objective and the numerical aperture used.

In experiments described below, monoclonal antibodies (mAbs) were directly conjugated to fluoresceine-iso-thio-cyanate (FITC) or phycoerythrin (PE) and assembled as a library. The repetitive use of mAb sets, each set containing two different FITC or PE-conjugated mAb's, and the repetitive cyclical spectral isolation of these fluorescent markers leads to a collection of images (Fig. 2, forward cycle images, runs A–C). Each image contains the fluorescence distribution pattern of one single protein epitope. Overlay of these images, based on precise geometric alignment, provides the possibility to consider, for each pixel, the list of its various fluorescence intensities expressed either in terms of a vector of grey values, or, after the introduction of appropriate threshold values for each signal, in terms of a 0,1-vector, also called a combinatorial molecular phenotype (CMP), indicating absence or presence of the protein in question at the pixel under consideration. Thus, each primary data set (the grey-value images) gives rise to a collection of pixel-CMPs. Each cell in a visual field can thus be compared quantitatively with other cells displaying similar or distinct CMPs. A collection of different CMPs represents a momentary snapshot of a cell's or a tissue's phenotype. By attaching CMPs to pixels of individual cells in a given biological sample, detailed somatotopic CMP maps, termed toponome maps, can be constructed. Toponome maps are considered to provide the basis for a mathematical and computational exploration of the contextual nature of proteins as a system in situ at the light-microscope level.

Performance of MELK

To run MELK on the biological samples used in this study, mAb's (Table 1) were used at a dilution of 1 µg/ml in PBS. These standardized mAb's recognize cell surface proteins, which are highly characterized, widely used in immunology, and known to interact in the cell surface membrane to form cell-to-cell and cell-to-matrix adhesion as well as signal transduction functions. A $40\times$ phase contrast water objective with a numerical aperture of 0.9 (Zeiss) was used to visualize a tissue region of interest by using PBS as optical immersion medium. To start the MELK, the phase contrast image of a selected visual field was registered in a first step, then followed by gentle removal of the PBS. By avoiding any displacement of the object and of the objective, the sample was incubated for 15 min "on stage" with a set of 2 mAb's directed against CD4 and CD8, conjugated to PE and FITC, respectively. After 10 washing steps with PBS "on stage" the fluorescence signals were registered selectively by using the following filter sets: BP465/FT510/BP515-545 for FITC and BP546/FT580/DEPIL575 $\pm$ 8 nm for PE. Signals were digitized by using a cooled CCD camera (photometrics). This was followed by soft fluorescence bleaching at the same excitation wavelengths for at least 10 min. The maximum bleaching effect was registered as a "post bleach end point image," and a second phase contrast image was registered
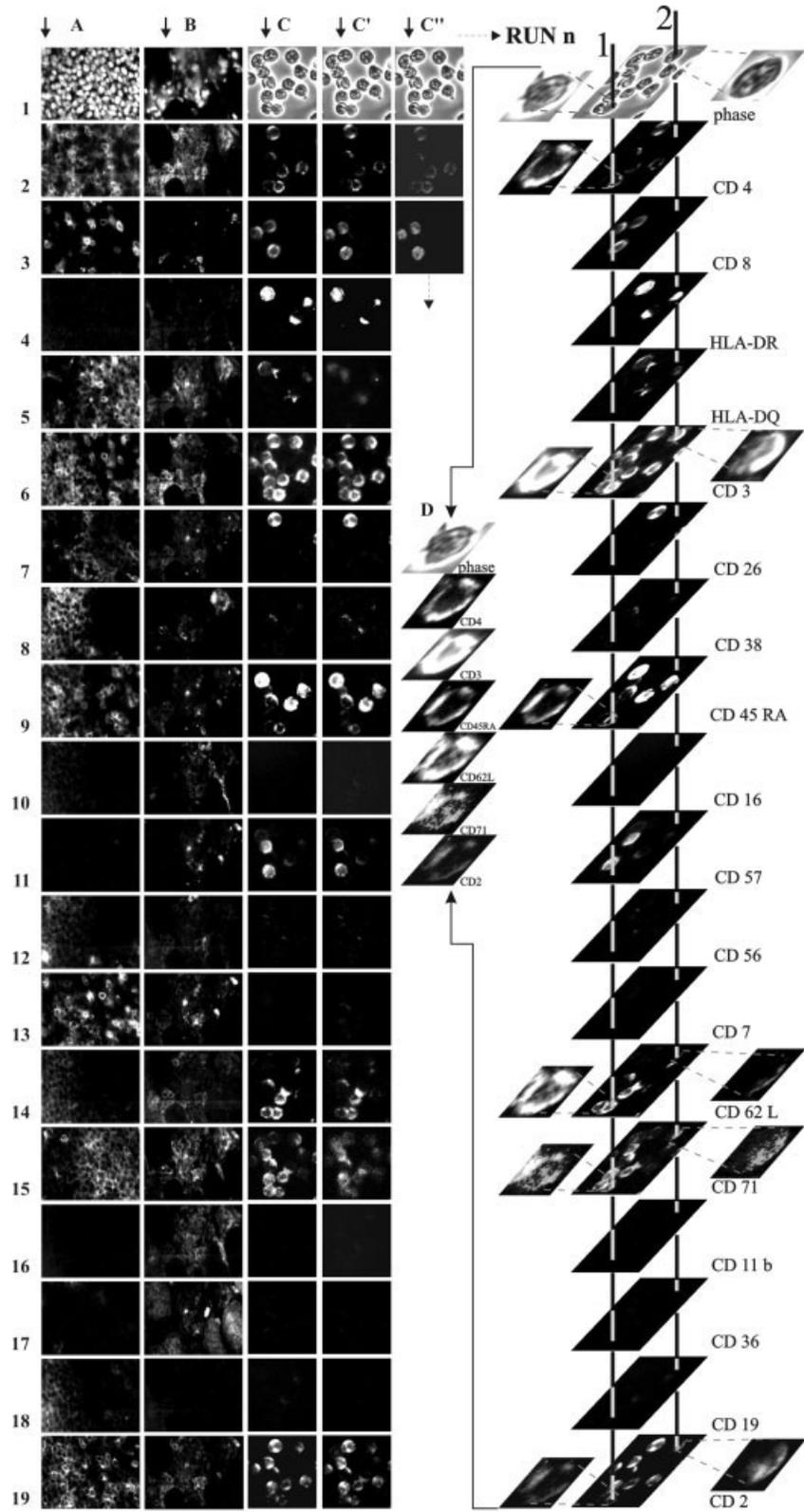
Fig. 2. Cyclical assessment of many protein signals in one sample. MELK data set illustrating the simultaneous localization of 18 different cell-surface differentiation marker proteins in three human immune compartments by using a library of mAbs (Table 1). (**A**) Human palatine tonsil, extrafollicular region; (**B**) Inflammatory infiltrate of the human skeletal muscle; (**C-C″**) PBMC, sequentially labeled by MELK using the same mAb library indicating reproducibility of epitope patterns. Altogether nine cycles, 2 mAbs per cycle, were run. (A)–(C), labeled epitopes are specified on the right. (**D**) Magnification of two differentially labeled PMBC. (A1) and (B1) propidium jodide nuclear stains. (C1–C1′) phase contrast images of PBMC. Note that CD36 is an antigen present in endothelial cells and in cross sectioned muscle fibers (B, 17) (26). Scale Bars: A/B, 50 μm; C, 20 μm.

to control the morphological integrity of the sample, thereby finalizing the first incubation–imaging–bleaching cycle of the MELK procedure (not shown). The next cycle was started in the same sequence of steps as described. Altogether nine cycles were run in the following sequence: CD4-PE/CD8-FITC; HLADR-PE/HLADQ-FITC; CD3-PE/CD26-

FITC; CD38-PE/CD45RA-FITC; CD16-PE/CD57-FITC; CD7-FITC; CD62L-PE/CD71-FITC; CD11b-PE/CD36-FITC; CD19-PE/CD2-FITC (Fig. 2).

Image Processing and Construction of Toponome Maps

The images of each MELK procedure were aligned pixel-wise based on matching the phase contrast images of each cycle. For each image a threshold was set to binarize each fluorescence signal [0 = below threshold; 1 = above threshold]. Thresholds were generally set below the lowest fluorescence signal and above the background signal of individual images. These criteria coincide largely with threshold criteria determined by an independent automated threshold setting method involving combinatorial geometry and statistics (27) (not shown). The binarized images were superimposed, and the resulting binary CMPs were saved as a list. Fractions of this list—generally the 30 most frequent CMPs (in this study)—were visualized as an assembly of color-coded geometric objects in the images defined as toponome maps.

Mapping Mononuclear Leukocytes in Multiple Immune Compartments

To illustrate the working of MELK, protein labeling cycles were run on samples from different human immune compartments and then aligned (Figs. 2A–2C). First, peripheral blood mononuclear cells (PBMC) from a healthy donor were analyzed. Briefly, PBMC isolated from the blood and fixed on cover slips by routine procedures (25) were subjected to the MELK procedure. In the present example, altogether nine cycles with 2 mAbs per cycle were performed showing highly heterogeneous cell surface protein–epitope distribution patterns (Fig. 2C) with fluorescence intensities ranging from high to low (Fig. 2C, 2–18). Moreover, some epitopes were undetectable reflecting the fact that in the blood of healthy donors, PBMC may not coexpress all the 18 cell-surface receptors.

To examine the specificity and reproducibility (robustness) of the single labeling patterns obtained during a MELK procedure, MELK was performed on the same sample repetitively as internal control (Figs. 2C′ and 2C″). After RUN 1 (Fig. 2C), the same mAb library was applied twice to the same sample at the same mAb concentrations (RUN 2 and RUN 3, respectively). As illustrated by direct comparison of the corresponding cellular epitope distribution patterns (Figs. 2C–2C″; horizontal collection of images), the cell-surface locations of the corresponding epitope signals in all three runs are identical, although the intensities of the signals may decline from RUN 1 to RUN 3. This is due to the progressive mAb saturation of the epitope-binding sites remaining free in the cells after the preceding runs. This illustrates that the labeling procedures of cycles 1 through 9 do not alter the localization characteristics of the same mAbs applied in the consecutive cylces 10 through 19, respectively (corresponding to images 2–19 in Run 1 and 2–19 in Run 2, respectively). To demonstrate this in detail, the CD4 signal was depicted as

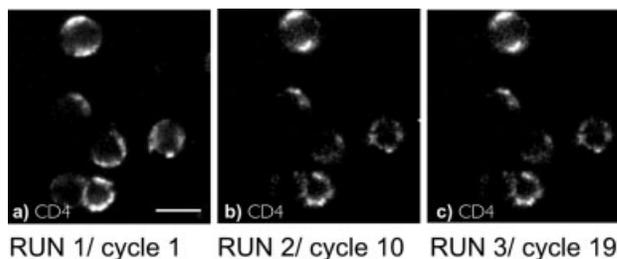RUN 1/ cycle 1   RUN 2/ cycle 10   RUN 3/ cycle 19

FIG. 3. Reproducibility of individual protein signals after multiple labeling cycles illustrates robustness of MELK. Micrograph illustrating the CD4 signal reproducibility through three sequential MELK runs on the same cells (**a–c** corresponding to cycle 1, cycle 10, and cycle 19, respectively). Note that the CD 4 images are depicted from Figure 3 (C2–C″2): the fluorescence intensities per cell slightly decline in the second and the third runs most likely due to a decrease of the number of remaining free nontagged epitopes after the preceding MELK runs. Note that the subcellular locations of the signals are reproducible despite the preceding rounds of mAb tagging procedures. Scale bar: 10 μm.

an example: Figure 3 gives the CD4 cell-surface fluorescences of the RUN 1 through RUN 3, each of which was registered using identical optical parameters including the digital integration time of the CCD camera (2″). While the pattern of relative intensities of the CD4 signal remains the same, the slight decline of the CD4 fluorescence intensity may reflect the decrease of remaining free epitopes for mAb binding. This indicates that the CD4 mAb recognizes reproducibly the same subcellular sites irrespective of the preceding cycles of labeling with the same and other mAbs. Similar results were obtained, when the primary sequence of antibodies was randomly permuted (not shown).

To examine the working of MELK on tissue sections, the aforementioned mAb library was applied to map mononuclear leukocytes (ML) within two different immune tissue compartments (human palatine tonsil and inflamed muscle). The characteristic histologic feature of these compartments is the functional assembly of ML as lymphoid tissue structures: (i) the extrafollicular, subepithelial site of the human palatine tonsil (PT); (ii) an inflammatory infiltrate of ML in the human skeletal muscle located in the endomysium. The latter is the hallmark of polymyositis (PM) (14). PT is a lymphoepithelial secondary immune organ belonging to the integrated mucosal immune system of the pharynx (28) to protect the body from microbial invasion (29). PM is a chronic inflammatory human muscle disease, mediated by invasion, and accumulation of T lymphocytes penetrating the basal lamina of muscle fibres (14). The working of MELK applied to acetone-fixed, 5-μm-thick cryo tissue sections of these two lymphoid compartments is illustrated in Figs. 2A and 2B, showing a unique distribution pattern for each of the 18 protein–epitopes in these tissue types, indicating selectivity and specificity of the MELK procedures.

Formalization of MELK Data Sets as Geometric Objects

Deriving CMPs from binarization of original MELK data sets (Figs. 2A and 2B) was chosen as one possible formal-

Table 2
*Schematic Illustration of the Topological Hierarchies
of Proteins within the Toponome*

	Proteins				
	★	■	▲	⬠	
Pixel	1	0	1	1	CMP1
Pixel	1	0	0	0	CMP2
Pixel	1	0	1	0	CMP3
	1	0	*	*	CMP–Motif (Toponome motif)
	↑				

Lead Protein (L); O = absent protein (A); * = wild card proteins (W)

Hierarchies are identified by setting a threshold for each protein signal (1 = present, above threshold, 0 = absent, below threshold; = 1 bit for each protein signal per pixel). Four proteins are presented as symbols (upper line). The interrelationship of different CMPs and a CMP motif (group of different CMPs having a lead protein in common) is illustrated. CMP motifs, characterize a functional state of a cell or a tissue, in which, formally, the lead protein is hierarchically dominant.

ism to assign a label to a local protein ensemble in biological structures. This type of formalism is a relatively simple way of describing the cocompartmentalization and arrangement of proteins in subcellular and supracellular contexts at once, although the information content of the primary grey value images is strongly reduced. Consequently, the resulting CMPs represent an approximation of the relative abundances of proteins within a structural data point (pixel). The identification of the combinatorial organization of proteins per data point or sets of data points in situ as CMPs requires "holistic" images containing the relevant information of the whole MELK data set. Such images were constructed by associating one color with the most prominent CMPs, coloring each pixel with that color that displays this CMP. Several CMPs may be grouped as CMP motifs. By definition, a CMP motif is the formal description of a cluster of CMPs that have several features in common: (i) one or more proteins that are present in all CMPs of the motif, the lead proteins (Ls); (ii) inversely correlated (anticolocalized) proteins, not present in any of the CMPs of the motif (absent = A), and (iii) variably associated or "wild card" proteins, e.g. proteins that may not be present in all CMPs of the cluster (W). A typical CMP-motif would then be denoted by a sequence of 1s, 0s, and *s, or Ls, As, and Ws, indicating the lead, the inversely correlated, and the wild card proteins, respectively (schematically illustrated in Table 2).

Toponome Maps of Lymphoid Tissue: Direct Labels of an Organized Cell Surface Proteome Fraction

The visualization of all, or fractions of all, detected CMPs derived from the MELK data set shown in Figs. 2A and 2B reveals a set of multidimensional binary vectors represented as a toponome map (Fig. 4). It provides a comprehensive view of protein cocompartmentalization structures in the given visual field. Since the protein cocompartmentalization machinery in cells and tissues determines structure and function, toponome maps dis-

sect a cell or a tissue into topologically confined functional entities, or domains. The toponome maps shown in Figure 4 are direct representations of the protein cocompartmentalization of the 18 cell-surface proteins in lymphoid tissue structures under investigation. One can easily identify several CMPs, CMP clusters, and CMP motifs that are signatures of different cell types/phenotypes delineating different spatially separated domains in the lymphoid tissue: Figure 4, segments I–VI (Seg I–Seg VI). Interestingly, the extrafollicular tonsil region (Figs. 4A–4C) and the lymphoid infiltrate in muscle (Figs. 4D and 4E) are clearly distinct by the presence of different CMPs, CMP clusters, and CMP motifs, indicating tissue specific different groupings of the same proteins.

Modes of the Toponome Described by Three Letters

CMP motifs can be described by a simple three letter code schematically outlined in Table 2. Concerning the formal description of topologically separated tissue domains, or toponome units (Seg I–Seg VI), the definition of lead proteins (Ls) is extended to single CMPs denoting a tissue region (e.g. Seg IV, V, and VI in Fig. 4): expressing the 1s as Ls in the corresponding CMPs. Hence the binary codes given in the boxes of Fig. 4 for Seg I–Seg VI can be translated into a L, A, W-code (Fig. 5A), expressing the topologically confined pixel groups of these segments as graphs of a specific sequence of As, Ls, and Ws, in which each of these letters denotes a specific protein (1–18). This formalism readily shows that Seg I is isomorph to Seg VI (Seg I ≈ Seg VI), as these segments share protein 12 (CD38) as a lead protein (L), while all the other segments are not isomorph. It may be an interesting further notion that Ls can turn into Ws, and As can turn into Ls, as shown for example for L12 → W12 (CD38) in (Seg I/Seg II), and A15 → L15 (HLADQ) in (Seg I/Seg III), respectively. To further indicate similar important interrelationships by appropriate graphical representations of the binarized protein signals, the three-letter graphs (Fig. 5A) are symbolized as a simple tree (Fig. 5B), in which only the major nodes linking the different graphs, are indicated. In this regard, the node [L12/Seg I → L12/Seg VI] is an element of the isomorphic graphs Seg I and Seg VI (Seg I ≈ Seg VI), while all the other nodes [A5/Seg I → L5/Seg IV; A7/Seg I → L7/Seg V; L12/Seg I → W12/Seg II; A15/Seg I → L15/Seg III] represent elements that switch from one letter to another reflecting changes in abundance of this particular protein relative to other proteins in the corresponding tissue regions (Fig. 4).

DISCUSSION

The characteristic property of the MELK technology, presented here, is its independence of the limitations of multispectral fluorescence-isolation procedures. Thereby, it provides the degree of freedom necessary to explore the organization of proteomes (or subproteomes) in morphologically intact cells and tissues on a broad scale: It works at photonic resolution; is independent of the
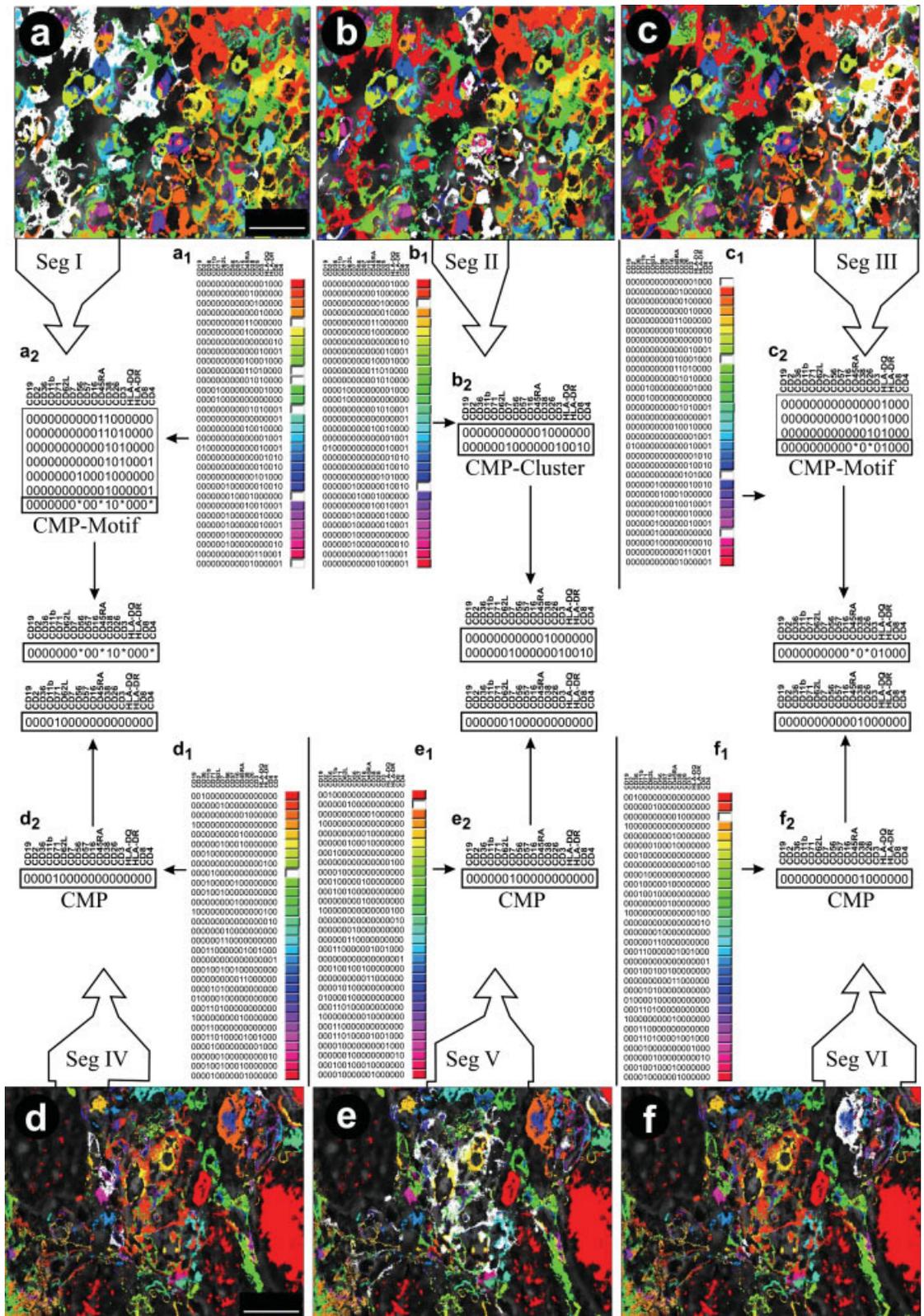
Fig. 4. Toponome maps of lymphoid tissue uncovering spatially separated tissue domains. Lymphoid tissue structures of the tonsil (**a–f**) and the inflamed muscle (**d–f**) are organized into three different spatially separated domains, or segments, as defined by single CMPs, CMP clusters, or CMP motifs (white regions in the images a–f, denoted as segments, Seg I–VI, respectively). These toponome maps were constructed from MELK data shown in Figures 2A and 2B. The 30 most frequent CMPs indicated in different colors have been localized simultaneously (lists a1–f1). Lined boxes: specific groups of pixels high-lighting different segments of the tissue structures (Seg I–Seg VI). These toponome units (a2–f2) have been extracted from the lists (a1–f1) to link the topo-nome binary code with the corresponding tissue structure as geometric objects (white regions in a–f). Top horizontal line of every CMP list (a1–f1): specifi-cation of the epitopes. Scale bars: 20 μm.

a Proteins 1 - 18

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Seg I	A	A	A	A	A	A	A	W	A	A	W	L	A	W	A	A	A	W
Seg II	A	A	A	A	A	A	W	A	A	A	A	W	A	W	A	A	W	A
Seg III	A	A	A	A	A	A	A	A	A	A	W	A	W	A	L	A	A	A
Seg IV	A	A	A	A	L	A	A	A	A	A	A	A	A	A	A	A	A	A
Seg V	A	A	A	A	A	A	L	A	A	A	A	A	A	A	A	A	A	A
Seg VI	A	A	A	A	A	A	A	A	A	A	A	L	A	A	A	A	A	A

FIG. 5. A three symbol-code describing toponome units as graphs. The segments (Seg I–Seg VI) signifying subregions in the tissue sections (Fig. 4) are described by a three-letter code. L = lead proteins; W = wild card proteins; A = absent proteins (for definitions see Table 2). The resulting combinatorial graphs for Seg I–Seg VI are derived from the sequences of 1s, 0s,*s outlined in the lined boxes (Fig. 4) indicating toponome units confined to their corresponding tissue subregions (groups of white pixels). (**a**) List of graphs giving the sequence of Ls, As, and Ws corresponding to proteins 1–18. (**b**) Line graph illustrating the nodes (small boxes) by which the six graphs are linked: small arrows indicate to which graph the corresponding letters belong. Numbers 1–18: 1 = CD19, 2 = CD2, 3 = CD36, 4 = CD11b, 5 = CD71, 6 = CD62L, 7 = CD7, 8 = CD56, 9 = CD57, 10 = CD16, 11 = CD45RA, 12 = CD38, 13 = CD26, 14 = CD3, 15 = HLADQ, 16 = HLADR, 17 = CD8, 18 = CD4.
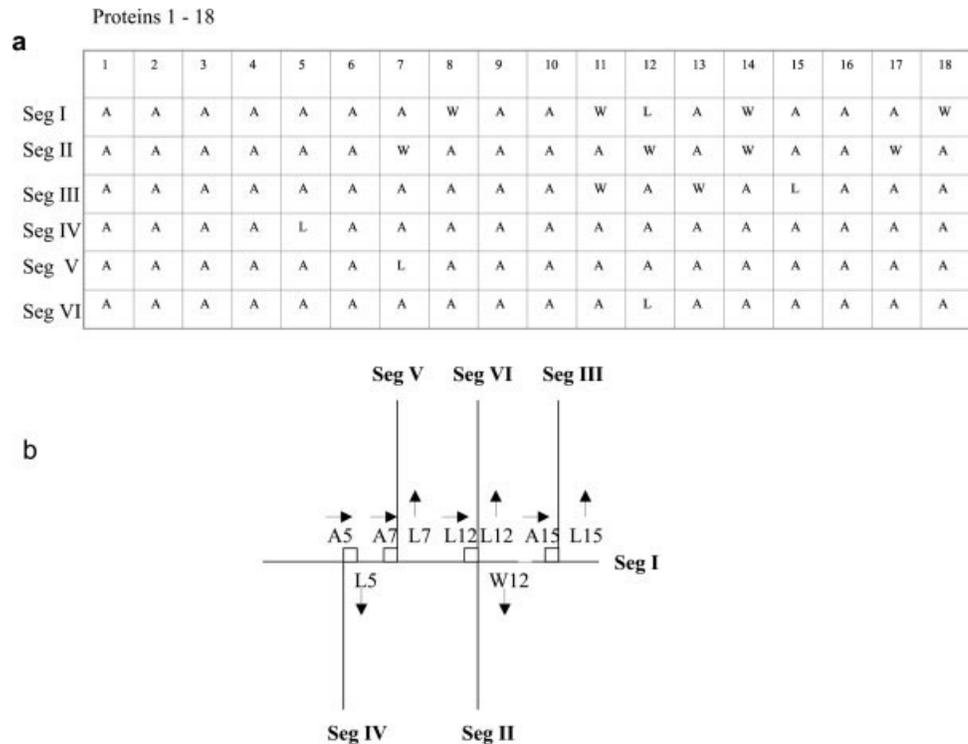
kind and number of parameters (i.e. proteins) to be observed independently; works, as shown here, on routinely fixed cells and tissues; allows for a direct alignment of CMPs with known biological structures necessary for the integration of cell-biological and histo(patho)logical interpretations; and covers the whole spectrum from the localization of single proteins (or other molecular classes) to higher-order contextual molecular patterns in one experiment.

Careful calibration procedures are required to establish antibody- or any other affinity tag-libraries for MELK. The procedure involves: (i) characterization of each antibody/tag selectively (both biochemically and histologically); (ii) finding its optimal concentration for incubations related to tissue type of interest, fixation method used, fluorescence signal range, and dynamic range of the CCD camera; (iii) combining the so-selected antibodies/tags in a MELK procedure comparing their signal patterns qualitatively and quantitatively with those obtained by single staining experiments (above i); (iv) identifying or excluding significant sterical hindrances of antibody/tag-epitope binding by using repetitive MELK runs (Figs. 2C–C″), optionally including, inversions and permutations of the MELK runs, e.g. changing the sequences of antibody labeling procedures; (v) potentially, as a result of (iv): changing the sequence of antibody labeling steps to correct for possible mutual inhibitions. If this regimen is carefully followed, it appears to be possible to colocalize at least 100 different tissue components successfully. Moreover, multiple mAbs against different epitopes of one given protein can be used to substantiate the identity of this protein in question. A tissue section, which is fixed in acetone, as shown here, or in paraformaldehyde (not shown), and which is progressively labeled by MELK at 20°C, remains well preserved for at least 2 weeks. The present MELK runs comprised less than 1 day on one instrument. The morphological integrity of the tissue over the whole measuring period can—in addition to visual control—be quantified by (i) registering one phase contrast image per cycle, (ii) generating a last cycle image by using the first cycle antibody, and (iii) then match these signals to compare the grey value data for correspondences.

The toponome maps of lymphoid tissue (Fig. 4) suggest that the differential combination (local relative abundances) of the 18 cell-surface proteins under investigation plays a crucial role in the hierarchical organization of lymphocytes into functionally different compartments and tissue domains, and similar combinatorial rules may govern lymphocyte homing (30). The resulting CMPs and CMP motifs give the relevant coding structures at a higher level of abstraction.

The present data sets suggest a hierarchical order of proteins and of their combinatorial patterns (CMPs and CMP motifs). The resulting lead proteins (Ls) may be of importance as putative elements controlling the functionality/structural integrity of molecular networks. Although analysis of physical protein–protein interactions is beyond the scope of the present study, it is interesting to consider that lead proteins may represent the interaction hubs typical for scale-free networks (31,32). Supporting this view, MELK has already been used to identify FcγRIII receptor as a lead protein and potential target in amyotrophic lateral sclerosis (33), a prediction that has been confirmed by the phenotype of an Fcγ receptor knock-out mouse (34).

The entirety of CMP motifs is considered to represent the total functional plan of a cell or a tissue. The corresponding geometric features, assembled as toponome maps, are likely to focus our attention on hierarchical patterns of proteins that are associated with given functional states of cells or tissue (e.g. in disease) escaping detection by protein-profiling techniques averaging protein abundances over millions of cells, or by tissue-destructing methods like laser capture microdissection (35) and scanning mass spectrometry (36,37). On the other hand, MELK combined with these technologies might provide a challenging future option.

Together, the present findings indicate that MELK is a powerful way to analyze protein organization in cells or tissue, describable by a three-symbol code. This makes structurally bound protein systems amenable to mathematical analysis using methods of combinatorial statistics and geometry. The present data suggest that every protein network in a cell or a tissue obeys a unique protein colocalization and anticolocalization code, a hypothesis that can now be systematically addressed using biological model systems by selectively inhibiting or deleting lead- and wild card proteins and the resulting consequences on CMP motifs and cell/tissue functionalities observed. Another challenging option will be to search for toponome "fingerprints" distinguishing diseases at an early time point of pathogenesis.

LITERATURE CITED

1. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M. Global analysis of protein activities using protein chips. Science 2001;293:2101–2105.
2. Fromout-Racine M, Rain J, Legrain P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat Genet 1997;16:277–282.
3. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothenberg JM. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 2000; 403:623–627.
4. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 2002;415:180–183.
5. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, et al. A protein interaction map of *Drosophila melanogaster*. Science 2003;302:1727–1736.
6. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, et al. Functional organisation of the yeast proteome by systematic analysis of protein complexes. Nature 2002;415:141–147.
7. Kim JK, Bhinge AA, Morgan XC, Vishwanath RI. Mapping DNA-protein interactions in large genomes by sequence tag analysis of genome enrichment. Nat Methods 2005;2:47–53.
8. Collins FS, Green ES, Guttmacher AE, Guyer MS. A vision for the future of genomics research. Nature 2003;422:835–847.
9. Valet G. Cytomics, the human cytome project and systems biology: Top-down resolution of the molecular biocomplexity of organisms by single cell analysis. Cell Prolif 2005;38:171–174.
10. Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. Science 2004; 306:640–643.
11. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. Science 2005;308:523–529.
12. Valet GK, Tarnok A. Cytomics in predictive medicine. Cytometry B Clin Cytom 2003;53B:1–3.
13. Douma S, Van Laar T, Zevenhoven J, Meuwissen R, Van Garderen E, Peeper DS. Suppression of anoikis and induction of metastasis by the neurotrophic factor TrkB. Nature 2004;430:1034–1040.
14. Schubert W, Masters CL, Beyreuther K. APP+ T lymphocytes selectively sorted to endomysial tubes in polymyositis displace NCAM-expressing muscle fibers. Eur J Cell Biol 1993;62:333–342.
15. De Rosa SC, Herzenberg LA, Roederer M. 11-color, 13-parameter flow cytometry: Identification of human naive T cells by phenotype, function, and T cell receptor diversity. Nat Med 2001;7:245–248.
16. Alivisatos AP. The use of nanocrystals in biological detection. Nat Biotechnol 2004;22:47–52.
17. Dickinson ME, Simbuerger E, Zimmermann B, Waters CW, Fraser SE. Multiphoton excitation spectra in biological samples. J Biomed Opt 2003;8:329–338.
18. Mittag A, Lenz D, Gerstner AO, Sack U, Steinbrecher M, Koksch M, Raffael A, Bocsi J, Tarnok A. Polychromatic (eight-color) slide-based cytometry for the phenotyping of leukocyte, NK, and NKT subsets. Cytometry A 2005;65A:103–115.
19. Ecker RC, Steiner GE. Microscopy-based multicolor tissue cytometry at the single-cell level. Cytometry A 2004;59A:182–190.
20. Kriete A, Boyce K. Automated tissue analysis—a bioinformatics perspective. Methods Inf Med 2005;44:32–37.
21. Gerstner AOH, Trumpfheller C, Racz P, Osmancik P, Tenner-Racz K, Tárnok A. Quantitative histology by multicolor slides-based cytometry. Cytometry A 2004;59A:210–219.
22. Schubert W. Topological proteomics, toponomics. MELK Technology. Adv Biochem Eng Biotechnol 2003;8:189–211.
23. Schubert W. Multiple antigen-mapping microscopy of human tissue. In: Burger G, Oberholzer M, Voojs GP, editors. Advances in analytical cellular pathology. Amsterdam: Excerpta Medica; 1990. pp 97–98.
24. Schubert W, Zimmermann K, Cramer M, Starzinski-Powitz A. Lymphocyte antigen Leul9 as a molecular marker of regeneration in human skeletal muscle. Proc Natl Acad Sci USA 1989;86:307–311.
25. Schubert W, Schwan H. Detection by 4-parameter microscopic imaging and increase of rare mononuclear blood leukocyte types expressing the FcγRIII receptor for immunoglobulin G in human sporadic amyotrophic lateral sclerosis (ALS). Neurosci Lett 1995;198:29–32.
26. Schubert W. Triple immunofluorescence confocal laser scanning microscopy: Spatial correlation of novel cellular differentiation markers in human muscle biopsies. Eur J Cell Biol 1991;55:272–285.
27. Dress A, Lokot T, Pustyl'nikov LD, Schubert W. Poisson numbers and poisson distributions in subset surprisology. Ann Combinatorics 2004; 8:473–485.
28. Nave H, Gebert A, Pabst R. Morphology and immunology of the human palatine tonsil. Anat Embryol 2001;204:367–373.
29. Ogra PL. Mucosal immune response in the ear, nose and throat. Pediatr Infect Dis J 2000;19:4–8.
30. Butcher EC, Picker LJ. Lymphocyte homing and homeostasis. Science 1996;272:60–66.
31. Bray D. Molecular networks: The top–down view. Science 2003;301: 1864–1865.
32. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M. Evidence for dynamically organised modularity in the yeast protein–protein interaction network. Nature 2004;430:88–93.
33. Schubert W. Use of substances with immunomodulating activity for the treatment of amyotrophic lateral sclerosis. U.S. Patents 6,638,506; 6,649,165;6,638,515 (1999).
34. Mohamed HA, Mosier DR, Zou LL, Siklos L, Alexianu ME, Engelhardt JI, Beers DR, Le WD, Appel SH. Immunoglobulin Fcγ receptor promotes immunoglobulin uptake, immunoglobulin-mediated calcium increase, and neurotransmitter release in motor neurons. J Neurosci Res 2002;69:110–116.
35. Benoyahn D, Akavia UD, Solcher, R, Shur I. Gene expression in skeletal tissues: Application of laser capture microdissection. J Microsc 2005;220:1–8.
36. Pacholsky ML, Winograd N. Imaging with mass spectrometry. Chem Rev 1999;99:2977–3005.
37. Stoeckli M, Chaurand P, Hallahan DE, Caprioli RM. Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. Nat Med 2001;7:439–496.