Published in final edited form as:
Anal Chem. 2020 May 05; 92(9): 6587–6597. doi:10.1021/acs.analchem.0c00349.

ToF-SIMS and machine learning for single-pixel molecular discrimination of an acrylate polymer microarray

Wil Gardner[1,2,3], **Andrew L. Hook**[4], **Morgan R. Alexander**[4], **Davide Ballabio**[5], **Suzanne M. Cutts**[2], **Benjamin W. Muir**[3], **Paul J. Pigram**[1,*]

[1]Centre for Materials and Surface Science and Department of Chemistry and Physics, La Trobe University, Melbourne, Victoria, Australia

[2]La Trobe Institute for Molecular Sciences, La Trobe University, Melbourne, Victoria, Australia

[3]CSIRO Manufacturing, Clayton, Victoria, Australia

[4]Advanced Materials and Healthcare Technologies, School of Pharmacy, University of Nottingham, Nottingham NG7 2RD, UK

[5]Milano Chemometrics and QSAR Research Group, Department of Earth and Environmental Sciences, University of Milano-Bicocca, P.zza della Scienza 1, 20126, Milano, Italy

Abstract

Combinatorial approaches to materials discovery offer promising potential for the rapid development of novel polymer systems. Polymer microarrays enable the high-throughput comparison of material physical and chemical properties – such as surface chemistry and properties like cell attachment or protein adsorption – in order to identify correlations that can progress materials development. A challenge for this approach is to accurately discriminate between highly similar polymer chemistries or identify heterogeneities within individual polymer spots. Time-of-flight secondary ion mass spectrometry (ToF-SIMS) offers unique potential in this regard, capable of describing the chemistry associated with the outermost layer of a sample with high spatial resolution and chemical sensitivity. However, this comes at the cost of generating large scale, complex hyperspectral imaging datasets. We have demonstrated previously that machine learning is a powerful tool for interpreting ToF-SIMS images, describing a method for color-tagging the output of a self-organizing map (SOM). This reduces the entire hyperspectral data set to a single reconstructed color similarity map, in which the spectral similarity between pixels is represented by color similarity in the map. Here, we apply the same methodology to a ToF-SIMS image of a printed polymer microarray for the first time. We report complete, single-pixel molecular discrimination of the 70 unique homopolymer spots on the array, while also identifying intra-spot heterogeneities thought to be related to intermixing of the polymer and the pHEMA coating. In this way, we show that the SOM can identify layers of similarity and clusters in the data, both with respect to polymer backbone structures and their individual side groups.

[*]**Corresponding Author**: p.pigram@latrobe.edu.au.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Finally, we relate the output of the SOM analysis with fluorescence data from polymer-protein adsorption studies, highlighting how polymer performance can be visualized within the context of the global topology of the data set.

Time-of-flight secondary ion mass spectrometry (ToF-SIMS) is an attractive technique for studying surface chemistry at the molecular scale, with a depth resolution of a few nanometers and excellent spatial resolution. ToF-SIMS has a broad range of applications for both organic and inorganic materials, including, chemical mapping of cells and tissue[1–9], studying stress corrosion cracking in metal alloys and flexible optoelectronic devices[10, 11] and characterizing chemically similar polymeric materials. [12–18] The data produced by contemporary ToF-SIMS instruments is inherently hyperspectral, characterized by both *x* and *y* spatial information, with a mass spectrum associated with each pixel. When spatial information is not relevant to the analyst, the data can be summed to produce a single 1D spectrum containing the cumulative mass spectral information from across the analysis area. Alternatively, spatial coordinates can be used to produce 2D chemical maps of the surface, in which the distribution of intensities from a single peak, mass segment or group of peaks is visualized as an image. The latter approach can provide unique insights into how particular components are distributed on the surface. For example, in the field of bioimaging, a recent study investigated the distribution of zinc in the hippocampi of healthy and brain injured rats. [8] Other recent work has focused on imaging nanoparticles in cells and tissues, both organic[5] and inorganic. [9, 19]

The size and complexity of ToF-SIMS imaging data are significant challenges impeding complete and robust interpretation. A single ToF-SIMS spectrum can contain hundreds or sometimes thousands of identifiable mass peaks. This equates to the same number of individual molecular ion maps. Each map will typically contain tens of thousands of pixels, each of which could potentially contain analytically important information. Since it is not feasible to compare hundreds of images, or to analyze thousands of individual spectra, the use of multivariate analysis (MVA) is essential and has proven invaluable in the accurate interpretation of ToF-SIMS hyperspectral images.Our previous work[20–25] has focused on applying ToF-SIMS and MVA in the study of printed polymer microarrays. Hook et al. [26] provide a comprehensive review of the development, principles, applications and advantages of polymer microarray systems for high-throughput materials discovery. Briefly, polymer microarrays are prepared by printing a series of polymer spots onto glass slides, with diameters typically in the range of hundreds of micrometers. [20, 21], [24, 26] Such arrays have been successfully used to discover novel materials with useful biological properties, such as improved stem cell attachment[27–29] or resistance to bacterial biofilm formation. [30, 31] Once printed, the polymers can be subjected to a suite of characterization techniques, such as XPS, Raman spectroscopy, ToF-SIMS and water contact angle (WCA) measurements. [26] The microarray thereby provides a platform for the high-throughput characterization of polymer materials. Beyond materials discovery, the large number of material-biological interactions that can be assessed using this sample format makes it possible to identify structure-function relationships by correlating properties such as surface chemistry with performance; for example protein adsorption or polymer wettability. ToF-SIMS, in combination with MVA, has been particularly valuable in this regard. [22–24]

Initially, MVA was used as a tool for evaluating the surface chemistries of different polymer and copolymer materials, based on their summed 1D ToF-SIMS spectra. For example, Urquhart et al. [22] used principal components analysis (PCA) to analyze ToF-SIMS spectra from a 576 spot copolymer microarray. The study demonstrated important insights into the similarities between many of the polymers, highlighting specific fragments that were identified as commonalities. The authors also demonstrated the use of partial least-squares (PLS) regression analysis for correlating ToF-SIMS spectra with surface wettability, via WCA measurements. Similar results were obtained by Celiz et al. [20] who, in addition to predicting the WCA of polymers based on their surface chemistry, identified correlations between ToF-SIMS data and stem cell adhesion. Analogous findings are reported by Yang et al. [23] regarding the adhesion of human embryoid body cells to acrylate polymers with differing surface chemistries.

In addition to analyzing 1D spectra, recent research has focused on using MVA to evaluate ToF-SIMS hyperspectral images of polymer microarrays. Hook et al.[24] investigated the surface chemistry of 70 different poly(meth)acrylate spots in a printed microarray, using ToF-SIMS imaging and multivariate curve resolution (MCR). Similar to the studies involving 1D spectra, MCR revealed correlations between polymers using their associated spectra. However, in this case spectra were extracted from individual pixels, such that spatial information was retained. MCR scores were then used to reconstruct images of the surface, providing visualization of the spectral similarities between different spots in the array. Thereby, correlations related to both the polymer backbones and the extending side groups were revealed.

Recently, we have demonstrated the power of machine learning for interpreting ToF-SIMS data, using a particular artificial neural network architecture known as Kohonen[32] self-organizing maps (SOMs). The design and function of the SOM in general is described elsewhere[32–34] and in our previous works. [13–17, 35] For brevity, a detailed mathematical description of the underlying algorithm of the SOM is omitted here. SOMs are designed to reduce the dimensionality of a high dimensional dataset, usually producing a 2D model of the data. A key feature of the SOM is the ability to preserve the *topology* of the data during the non-linear mapping to a lower dimensional space. SOMs are supposed to mimic the action of biological networks of neurons, which share information and thus evolve over time. Neurons are typically organized in a square 2D space and associated with a vector of weights. These weights are then updated iteratively on the basis of the input samples. Similar samples activate topologically close neurons in the network, and consequently SOMs are expected to learn and model patterns from the input data.

We employ a toroidal topology for the SOM, in which neurons located on the edges of the lattice are neighbors to those on the opposite edge, forming a continuous surface in all directions. This is topologically equivalent to the surface of a toroid. Critically, toroidal SOMs have been shown to be superior in performance to planar SOMs.[36] Specifically, planar SOMs suffer from boundary effects associated with neurons on the edges of the lattice, due primarily to these neurons having fewer neuron neighbors than those in the center of the lattice. [37, 38]

We have demonstrated previously that SOMs appear better suited to handling non-linear ToF-SIMS data than traditional linear approaches, such as PCA or MCR. [14–17] Recently, our group has presented a novel approach for applying toroidal SOMs in the analysis of ToF-SIMS hyperspectral images. [35] The output of the SOM is a 2D topology-preserving map. In order to visualize this topology intuitively as a reconstructed image, it is useful to overlay a coloring scheme onto the SOM. In this way, the neurons are colored according to their positions, such that adjacent neurons are assigned a similar color. The color-tagging is designed to match the toroidal topology, such that the color change is continuous across the boundaries of the SOM. Pixels in the original image can then be colored according to their winning neurons on the SOM, which enables the reconstruction of a single, color similarity map of the analysis area in which pixels are colored according to their spectral similarity. This process is described in more detail in the Experimental Methods section, and in our recent work. [35]

Here, we present the application of color-tagged SOMs for the analysis of a ToF-SIMS hyperspectral image of a polymer microarray. Specifically, we use the same data set used for MCR analysis by Hook et al..[24]Results from this work are discussed and evaluated with two primary outcomes. First, we focus on the technique itself, demonstrating the power and robustness of the SOM for visualizing complex ToF-SIMS hyperspectral imaging data. Second, we provide strong justification for the use of ToF-SIMS and color-tagged SOMs in the analysis of polymer microarrays, specifically. In this regard, we show how ToF-SIMS and SOMs together offer unique potential for the continued advancement of combinatorial materials discovery using polymer microarrays.

Experimental Methods

Microarray printing

Polymer microarray printing has been described elsewhere generally, and for this array specifically. [24] Briefly, epoxy-functionalized glass slides (Genetix) were dip coated in 4% (w/v) poly(hydroxy ethylmethacrylate) (pHEMA) (Sigma) in ethanol. Polymerization solution – composed of 75% (v/v) monomer (Sigma, dissolved in dimethylformamide (DMF)) and 1% (w/v) photoinitiator 2,2-dimethoxy-2-phenylacetophenone – was printed onto the pHEMA-coated slides, using an XYZ3200 dispensing workstation (Biodot). Slides were irradiated with a long wave UV source for 30 s after printing each material, then for a further 10 min following completion of the array. Finally, the slides were vacuum extracted at <50 mTorr for 7 days. The CAS number and chemical name for each of the 71 spots is included in the Supplementary Information (Table S-1).

Protein adsorption and fluorescence microscopy

Protein adsorption and fluorescence microscopy are described in Hook et al. [24] Briefly, polymer microarrays were immersed in a solution of 25 μg/ml tetramethylrhodamine isothiocyanate labeled albumin (Sigma) in phosphate buffered saline (Gibco; pH 7.4), then incubated for 1 h at 37 °C under stagnate conditions. Arrays were then washed for 1 min with ultrapure water, blotted and dried overnight. A Genepix fluorescence scanner was used for imaging, with a 532 nm laser and 5 μm pixel size. Images were captured before and after

Europe PMC Funders Author Manuscripts

protein adsorption, to account for background fluorescence. The same procedure was applied to a control array – without any added protein – to quantify fluorescence. A total of eight replicates were measured.

ToF-SIMS experimental

Hook et al. [24] describe the protocol used to acquire the ToF-SIMS data using an IONTOF (GmbH) TOF.SIMS 4 instrument. In brief, a 25 keV $Bi3^+$ primary ion source was scanned over a 9.2 × 9.2 mm analysis area, employing the macroraster scanning functionality. Negative ion mode was used, and a single scan performed using 15 pulses per pixel, with a pixel size of 10 × 10 μm. A low energy (~20 eV) electron flood gun was used for charge compensation.

Data preprocessing, export and analysis using SOMs

Peaks were selected automatically using the spatially-summed spectrum with the peak search function in the SurfaceLab6 software package. A minimum threshold of >100 counts was used over the entire mass range, producing a peak list comprising 717 mass peaks. Individual ion images were then exported as text files, compiled into a single data matrix and normalized to total ion intensity per pixel using in-house MATLAB scripts, as described previously.[35] SOM training was performed using the Kohonen and CP-ANN Toolbox[33, 34] in MATLAB. A squared, toroidal topology was used for all analyses, as well as the batch training approach, whereby all samples are introduced to the SOM simultaneously. The analysisspecific data scaling procedures such as Poisson scaling and/or individual peak normalization, as well as SOM parameters used, are described in more detail in the Results and Discussion section.

SOM visualization and analysis

The SOM output was visualized using the color-tagging approach described in our previous work. [35] All steps were performed using in-house MATLAB functions. Briefly, the neurons of the toroidal SOM were mapped to the surface of a 3D torus from their positions within the 2D plane. The equations describing this mapping are presented in Gardner et al. [35] Neurons were then colored using a red-green-blue (RGB) color scheme according to their *x*-, *y*- and *z*-coordinates on the torus, respectively. For image reconstruction, pixels were assigned the color of their winning neuron, producing a similarity map in which similar pixels (according to their mass spectra) appear as a similar color.

We used the Jaccard similarity index[39] to objectively evaluate SOM performance, with regards to the single-pixel discrimination of polymer spots. In general, for a pair of sample sets, A and B, the Jaccard similarity index, $J(A, B)$, is defined as

$$J(A, B) = \frac{A \cap B}{A \cup B}.$$

Hence, $J(A, B)$ provides a measure of similarity between two sample sets, with a value between 0 and 1; when there are no intersections, $J(A, B) = 0$, whereas complete intersection yields $J(A, B) = 1$. Here, we calculated the Jaccard similarity index between polymer pairs

by considering the winning neurons of their constituent pixels. This is discussed in more depth in the Results and Discussion section.

Results And Discussion

Background subtraction

SOM analysis of the ToF-SIMS image was split into two parts: background pixel subtraction and sample training. Background subtraction was implemented due to the large size of the data set (920 × 920 pixels, with 717 mass peaks), with the goal of reducing computation time by the removal of irrelevant data. The background subtraction workflow is illustrated by the flow chart in Fig 2 and visualized using the polymer microarray image in Fig 3. A 10 x 10 neuron rectangular SOM was initialized using the eigenvalue approach, whereby the neuron weights were assigned based on the eigenvectors associated with the first two principal components of the data set. [34] With this approach, data were initially clustered based on a linear PCA solution. Hence, we opted to first scale the ToF-SIMS data using Poisson scaling, which has been shown to improve performance of techniques such as PCA and MCR by accounting for Poisson noise. [40–42] This scaling attempts to correct heteroscedastic noise by transforming the data into an alternate space in which the uncertainty in the data is more uniform. [40, 41]

The SOM was trained for a nominal single epoch following initialization to produce the SOM output. Pixels representative of the background were then selected with regions of interest (ROIs), using the similarity map produced by the color-tagged SOM as a guide (Fig 3A). Pixels within the ROIs were surveyed for their winning neurons, producing a list of neurons corresponding to the substrate (represented in black on the SOM in Fig 3B). Finally, all pixels sharing these neurons were identified in the original image and removed from the data matrix. Matrix indices corresponding to the location of nonbackground pixels were stored for future image reconstruction. Fig 3C shows the reconstructed image after background subtraction, while Fig 3D highlights the origin of the neuron colors according to their position on a 3D toroid.

This workflow has the same analytical goals as the manual selection of ROIs for background removal, however is more objective, automated, data-oriented and not restricted to shapes that can be drawn manually. For example, background pixels were identified dispersed throughout regions in which spot material had diffused into the surroundings, such as at spot 56. These pixels would be impossible to locate and remove manually. The identification of background pixels is determined by their spectral similarity, and is based on the complete mass spectrum at each pixel rather than the intensity of an individually selected peak, set of peaks or total ion intensity. Here, we opted to use only the eigenvalue initialization of the SOM for background identification, as the polymer materials were clearly linearly separable from the substrate and this was the least time-consuming approach. However, it would be equally feasible to train the SOM for the required number of epochs to reach convergence, which would be necessary when there is a high level of spectral similaritybetween the sample and substrate. This would still reduce computation time, as a much smaller SOM can be used to complete a broad classification of background and sample pixels. As we demonstrate, a larger SOM can then be trained using the collapsed data matrix. It should be

Europe PMC Funders Author Manuscripts

noted that caution is required during background subtraction to avoid removing useful data by the misclassification of pixels. Hence, it is recommended that this approach should only be used when the approximate layout of the image is known. In this case the dimensions and layout of the microarray were clearly defined.

SOM visualization and analysis

Following background subtraction, the remaining pixels were used to train a new 40 × 40 neuron SOM, in this instance for the 200 training epochs required to reach convergence. Using the stored background pixel indices, it was then possible to reconstruct the original image using the SOM output. Specifically, non-background pixels were assigned the RGB colors of their winning neurons, then reinserted into their original positions in the image according to their stored pixel indices. Background pixels were colored white, producing a color similarity map of non-background pixels only. This image is presented in Fig 4A. Note that Poisson scaling was not used for this SOM. As discussed earlier, Poisson scaling transforms the data such that the uncertainty is more uniform. This is critical for techniques such as PCA, which rely on identifying directions of maximum variance. However, this process also inevitably modifies the underlying topological arrangement of the data, and hence is less applicable to techniques such as SOMs. We therefore opted to use the same preprocessing steps that we have reported previously, [13–17], [35] which have consistently delivered positive analytical outcomes from the resulting SOM. Nevertheless, while outside the scope of this work, a direct investigation of the effects of Poission scaling on SOM performance is an important future direction.

Qualitatively, it is clear from Fig 4A that the SOM successfully discriminated many of the polymers based on their mass spectra, as evidenced by their respective coloring. Furthermore, most of the polymers exhibited high color uniformity, suggesting insensitivity to spectral noise and chemical homogeneity within each spot. Polymer spots showing variation in assigned color were those in which there appeared to be real heterogeneities across the spot (for example, spots 38, 48 and 53), likely associated with drying effects, incomplete/inconsistent polymerization, or some other phenomenon. The SOM also revealed diffusion of polymer material, especially from polymer 56. Note that the DMF spot (71) was almost entirely absent after background subtraction, suggesting minimal difference in composition relative to the substrate, at least according to the PCA-based separation. Hence this spot was excluded from further chemical analyses.

Fig 4B shows the locations of the most abundant winning neurons from each of the 70 polymers, plotted onthe color-tagged SOM. Briefly, each polymer was selected individually using the polygon ROI select tool in MATLAB (polygon selections are displayed in Fig S-1 in Supporting Information (SI)). For each polymer, a histogram of winning neuron frequency was then calculated, and the most abundant neuron was selected and plotted on the color-tagged SOM. Each data point was then colored by the polymer backbone categories, namely acrylates, diacrylates, triacrylates, methacrylates and dimethacrylates. These categories are shown in Fig 4B, as well as in Fig 1A.

Fig 4B clearly illustrates the capabilities of the SOM in distinguishing each of the 70 polymers, at least with regards to their most abundant neurons. Only the top neurons are

Europe PMC Funders Author Manuscripts

shown so that global SOM performance can be visualized in a single image. However, the individual winning neuron compositions of all the polymers are also presented separately in Fig S-2 in the SI.

Fig 4C presents the winning neurons overlaid on the U-matrix for the SOM. The U-matrix provides a measure of the local distance information between adjacent neurons on the SOM, and hence can be used to infer the presence of clusters in the data.[43] Specifically, the greyscale value assigned to each neuron represents the mean distance between that neuron's weights and the weights of the adjacent neurons. Fig 4C shows more clearly the boundaries between many of the polymers on the SOM, as well as those polymers (for example, those in the corners of the SOM) that are more closely clustered. This is consistent with the individual winning neurons distributions presented in Fig S-2. However, it is important to note that the U-matrix is an approximation of the clusters in the SOM, and only provides entirely local and directionless distance information.

It is important to reiterate that these results were achieved using an entirely unsupervised approach, including the peak selection process. The only information that was required was the approximate layout of the array, for background subtraction. Hence, Figs 4–5 demonstrate the separation of 70 unique polymer structures without any prior knowledge of their surface chemistry. This exemplifies the power of the SOM in interpreting complex ToF-SIMS image data with minimal user input or intervention. According to Figs 4B–C, polymers in the array were generally well clustered with regard to their backbone structure. In particular, the acrylates, methacrylates and dimethacrylates appeared separately clustered to well defined regions on the SOM, whereas the diacrylates and triacrylates were clustered together. There also appeared to be ordering on the SOM reflecting the side group moieties present within each polymer. For example, spots 21, 27, 34, 37, 42, 58 and 69 all contained aromatic functional groups (specifically phenol or benzyl) and are all adjacent on the SOM in Figs 4B–C. The only other polymer in the array in this category is spot 63. This is strong evidence that the SOM recognized the chemical commonality with the side groups and weighted the neurons accordingly. Looking more closely at the chemical structure of each polymer and their corresponding weights reveals this explicitly.

Fig S-4 in the SI shows the mean weights (normalized to total peak weighting) associated with each of these 8 polymer spots. The weights can be used to infer which mass peaks were most important in distinguishing each polymer, according to the SOM algorithm. Note that the corresponding normalized mean spectra from each polymer are included in Fig S-5 in the SI. The $C_6H_6O^-$ ion, relating to the phenol group, was the top weighted fragment in spots 27, 34 and 37, and ranked 8th in 58 (in which the top-ranking weight is the similar phenolate ion, $C_6H_5O^-$). These polymers all contain phenol moieties situated at the end of their side groups, comprising a phenyl group with attached oxygen atom. Hence, the release of either of these ions is expected given this structure.[12] Similarly, the top weighted peak for both 21 and 69 was $C_7H_8O^-$, which is the fragmented benzyloxy group characteristic of the side groups in these polymers. The benzoate ion, $C_7H_5O_2^-$, was the 2nd highest ranked weight for spot 63, which again is in direct agreement with the structure of this polymer. Finally, the phenyl group in spot 42 is part of a larger nonylphenyl functional group, containing a characteristic 9-carbon tail. The corresponding nonylphenol fragment ion, $C_{15}H_{24}O^-$, was

ranked as the 15th highest weight for this polymer. This lower ranking is consistent with the increased size of this moiety, which is expected to fragment more readily into its constituents.

It is remarkable that, despite the different aromatic fragment ions produced by these eight polymers, they were still mostly grouped together on the SOM. This suggests that there were other, more subtle similarities in their spectra that correlated with aromatic groups. Interestingly, this does not seem to be the phenyl anion itself, $C_6H_5^-$, which was ranked at 6th, 203rd, 129th, 103rd, 174th, 149th, 3rd and 7th for spots 21, 27, 34, 37, 42, 58, 63 and 69, respectively. It is more likely that the similarities detected by the SOM are distributed across a range of mass peaks – including many that do not exclusively originate from the aromatic groups -which would otherwise be a significant challenge or practically impossible to identify manually. As a reference, Table S-2 in the SI shows the weights of the top 50 weighted mass fragments corresponding to each spot.

While these results are compelling, Fig 5 provides a more objective evaluation of the overall performance of the SOM. We used the Jaccard similarity index[39] to measure how well each pair of polymers was discriminated. Specifically, we used the Jaccard index to compare the winning neuron compositions of each polymer spot. The index represents, in this case, the number of unique winning neurons shared by two polymers (the inter-section) divided by the total number of unique winning neurons assigned to the pair (the union). Fig 5A presents this information as a greyscale matrix, in which the brightness of each square is proportional to the Jaccard index for the corresponding polymer pair. As a reference, Fig 5B shows a similar matrix, in which brightness is inversely proportional to the Euclidean distance between the mean spectra extracted from each polymer pair, and hence shows spectral similarity. Note that the histogram of the distances shown in Fig 5B is provided in Fig S-3 in the SI.

There are similarities between Fig 5A and 5B, suggesting that, qualitatively, the SOM is providing an accurate model of the underlying spectral similarities between polymers. More notably, 90% of polymer pairs yielded a Jaccard index of 0, meaning there was no intersection of winning neuron assignments to their pixels. This is remarkable, especially considering that a single common pixel assignment would produce a nonzero Jaccard index. This is further represented in the pie chart in Fig 5C, which shows the distribution of Jaccard similarity indices for the polymer pairs. Only 4% of polymer pairs yielded a Jaccard index greater than 0.1, and only 1% above 0.5. This is a clear, objective demonstration of the degree of singlepixel molecular discrimination achieved by the SOM.

The ability of the SOM to detect subtle difference between individual pixels is exemplified in Fig 6. Here, it is clear that the SOM identified two distinct regions within spot 38, as highlighted in Fig 6A. This image was used as a guide to select a set of representative pixels from each region (Fig 6B). After ROI selection, pixels were surveyed for their winning neurons, similar to the process described earlier used to produce Fig 4. However, rather than only considering the most abundant neuron in each ROI, the neurons accounting for > 95% of the pixel count were identified and included in the analysis. This approach was used to avoid including any outlier pixels in the selection, which could have skewed the mean

weights. Fig 6C shows the distribution of pixels with these same winning neurons across the entire analysis area, demonstrating the precision of the SOM clustering with almost all of these pixels localized within polymer 38. This is further emphasized in the close-up of the spot, which reveals that the SOM clearly distinguished ROI 1 from ROI 2, with minimal shared neurons. This is explored further in Fig 6D, in which the SOM has been shaded in greyscale according to neuron abundance in each ROI, based on the corresponding pixel histograms.

Visualizing the distribution on the SOM is useful to gauge the relatedness of the two ROIs. Here it is obvious that, while clearly distinct, ROI 1 and ROI 2 were highly similar with slight overlaps. This is also visibly evident in the weights extracted from each ROI (Fig 6E and Table 1). Note that the Euclidean distance between the mean spectra extracted from each region was 0.0269. For context, the median spectral distance between all of the polymer pairs was 0.085 (the complete distribution of distance is provided in Fig S-3 in the SI), suggesting a high level of similarity between the two regions.

Table 1 shows the top 16 weights extracted from each ROI – according to their normalized weightings – and their respective peak assignments. Also shown are the weights in the adjacent region, as well as the ratio between the two regions. The top weights provide insight into the fragment ions that best represent the underlying chemical composition of each region. In ROI 1, all of the weights shown correspond well with the chemical structure of the polymer, isodecyl acrylate. Further, the ratios of these weights between ROI 1 and ROI 2 are relatively consistent, varying between 0.72 and 1.36. This is also true for 13 of the top 16 weighted fragment ions in ROI 2. However, the remaining three fragment ions, $C_4H_5O_2^-$, $C_2H_5O_2^-$ and $C_2H_3O_2^-$ (highlighted in Table 1), were assigned significantly higher weighting in ROI 2 than in ROI 1. Further, these fragment ions appear highly related and may be indicative of the underlying pHEMA coating. For example, the $C_2H_5O_2^-$ ion – which is characteristic of the pHEMA structure[44] – was weighted 9.24 times higher in ROI 2 than in ROI 1. This also explains why the $C3H3O2^-$, $C_5H_5O^-$ and $C_9H_{15}O_2^-$ ions, which likely originated from the isodecyl acrylate polymer, were more highly weighted in ROI 1 than in ROI 2. Note that the differences in these peak weightings are minor, suggesting that the polymer was still a major component in ROI 2.

The pHEMA slide coating is intended to resist cell and protein attachment but is also known to intermix with the polymer spots. In some cases, for example when the spot material has high affinity for pHEMA or is particular thin, pHEMA has been detected at the surface of the spots. [44] For this spot, these findings suggest that pHEMA is present at the surface in ROI 2, indicative of intermixing and/or a thinner deposition of the polymer.

By visualizing these differences using the color-tagged SOM, the two regions were clearly distinguishable by their unique color compositions. This exemplifies the ability of this approach to reveal minor chemical difference between highly similar regions. More importantly, this is achieved using an entirely unsupervised workflow, requiring no prior knowledge of the sample composition.

Previous work has demonstrated how polymer microarrays can be used for high-throughput materials discovery, by correlating multivariate ToF-SIMS data with univariate performance metrics, such as protein adsorption, surface wettability or cell proliferation rates. [20, 23, 24] These studies have been successful in identifying individual mass fragments that either positively or negatively correlated with a given indicator, using partial least squares (PLS) regression. However, the relationship between ToF-SIMS data *topology* and performance has not been thoroughly explored. This was the motivation for Fig 7, which presents protein adsorption levels for each of the 70 polymers in the microarray in the context of their topological arrangement. Fig 7 was constructed in a similar fashion to Fig 4A, by considering the most abundant neuron associated with each polymer. However, rather than coloring the plot using the polymer categories, a heat map was used to display protein fluorescence and hence adsorption levels.

Exploring the relationship between data topology and performance in this way is useful, as it reveals global trends in the influence of surface chemistry. This contrasts with techniques such as PLS, which provides information regarding which *specific* variables correlate with performance. To demonstrate this, consider the bottom and top left of Fig 7. This region appears to be heavily populated with polymers exhibiting high levels of protein adsorption, which suggests common underlying mechanisms or properties that aid in protein binding. It is also useful to investigate polymers that appear to be highly similar according to their mass spectra, yet exhibit vastly different protein binding. For example, polymers 10, 68 and 62 are adjacent on the SOM, yet were correlated with high, moderate and low protein binding, respectively.

The advantages of the SOM are complementary to those of PLS, and hence combining these two approaches in the analysis is likely to provide unique insights into structure-function correlations. For example, applying PLS to this same dataset, Hook et al. [24] showed that $C8H3O4^-$ and $C6H3O^-$ ions correlated with high protein adsorption. It was then argued that these ions originated from the acrylate backbone and therefore that the backbone itself could promote protein adsorption. With regards to the SOM analysis, for spots 3, 5, 7, 10, 11, and 65 – which all exhibited moderate to high protein adsorption and are all clustered on the SOM – $C_8H_3O_4^-$ was ranked as the 2nd, 3rd, 5th, 2nd, 17th and 2nd highest weighted fragment, respectively. These results are consistent with the PLS results.

It is intriguing that the same fragment ion is also ranked as the 5th highest weight in the adjacent spot 62, which is correlated with poor protein adsorption. The weights associated with spot 62, comprising ethylene (glycol) diacrylate, are indeed very similar to those of the adjacent high-adsorbing spots. Using PLS analysis, Hook et al. presented the top fragment ions that correlated – either positively or negatively – with protein adsorption. We considered how highly these fragments were weighted in spot 62, in comparison to the high-adsorbing spots. Interestingly, none of these peaks distinguished spot 62 from the other spots, either alone or in combination. This suggests that spot 62 did not fit well with the PLS results, which is not surprising considering its location on the SOM.

We also considered the weights of spot 27, which exhibited moderate protein adsorption yet was distinctly separated from the other high-adsorbing spots on the SOM (Fig 7). Contrary

to the other spots, $C_8H_3O_4^-$ was ranked as the 207th highest weight for spot 27. However, C6H3O$^-$ was ranked as 11th, whereas this fragment ranks between 17th – 34th for the other spots. Together, these results suggest that, while polymers 3, 5, 7, 10, 11, 27 and 65 all exhibited moderate to high protein adsorption, the binding mechanism for polymer 27 may be different than that of the other six polymers.

It is noteworthy that polymer 27 contains a phenol group, which distinguishes it from the other high-adsorbing polymers. Furthermore, there appears to be a cluster of moderate protein binding associated with most of the other polymers containing aromatic functional groups (21, 34, 37, 42 and 58, as discussed earlier). Hence, this could suggest a weak correlation between the presence of an aromatic functional group and protein adsorption, although further investigations are necessary to support or refute this hypothesis.

It is interesting that Hook et al.[24] did not identify any correlations between aromatic-containing polymers and protein adsorption. There are two important considerations here. First, PLS is a linear method, so any non-linear relationships are less likely to be identifiable. This contrasts with the non-linear SOM. Second, it is useful to consider our discussion regarding Fig 4B and the weights associated with these polymers (Fig S-3). Specifically, in this discussion we argued that there weren't obvious single variables that connected each of these polymers according to their weights (for example, we highlighted how the weighting of the phenyl anion for each polymer varied considerably). Instead, we suggested that the SOM based the topological arrangement of these polymers on a collection of mass fragments, which were not obviously associated with the aromatic groups. In this way, the SOM uses considerable latent information from the spectra to cluster data. Without any obvious single variables connecting these polymers, it is expected that PLS would not be ideal for correlating aromatics with adsorption. In this regard, it is clear that the SOM provides a unique perspective for investigating correlations between groups of polymers and their performance that is not possible with PLS.

As a final point of discussion, it is important to include a caveat regarding the use of the proposed methodology. While we described a completely unsupervised workflow using the color-tagged SOM, a potential limitation is the reliance on the automated peak selection tool when making inferences about surface chemistry. In this work, we had access to information about the chemical composition of the polymers in the array, which was used to verify results from the analysis and confirm peak assignments. However, for many MSI experiments this information is not available, and hence it is important to manually check key peaks to avoid any misiterpretations due to, for example, peak shoulders or overlapping peaks. We have shown previously that the SOM still performs well using mass segmented data, where no peak selection is required and the data is represented as mass bins with uniform width.[14–16]This suggests that the peak selection process is not critical for the discrimination of samples. Nevertheless, manually confirming peak assignments is still important when drawing conclusions about the underlying surface chemistry.

Conclusion

The study reported here builds on our previously published work, demonstrating the ability of color-tagged SOMs to interpret complex ToF-SIMS hyperspectral images in an unsupervised manner. The advantages in using a polymer microarray for this purpose are twofold: first, this type of system offers a unique opportunity for investigating the capabilities of the SOM. A broad range of surface chemistries can be included on a single chip, and hence captured within a single ToF-SIMS image. This removes uncertainties associated with preparing multiple samples, as well as when acquiring data under slightly different instrument conditions. Further, the underlying composition of each spot is known, which enables direct comparison between known chemical information and the output from the SOM. As we have demonstrated, this is useful to clarify whether the SOM is revealing accurate information about the polymer surface chemistries, and hence demonstrate the suitability of this approach for more complex samples with an unknown chemical distribution.

Color-tagged SOMs were demonstrably able to produce unique analytical outcomes, with respect to other multivariate approaches.[24] The methodology described offers an intuitive approach for visualizing complex relationships between the spectra of individual pixels. We have demonstrated the accuracy of the output by considering not only how the polymers were clustered with regards to their polymer backbone, but also with respect to their individual side groups. Specifically, each of the 70 polymers were well discriminated which, considering the high similarity in the spectra produced by many of the polymers, is a nontrivial task. Further, polymers were clearly clustered on a global scale on the basis of their backbone structures, while side groups were influencing their local arrangements.

One of the most important applications of polymer microarrays is the identification of correlations between surface chemistry and polymer performance. This has been demonstrated using PLS paired with ToF-SIMS. [12, 22, 23, 25] We have shown how SOMs can be used to visualize the performance of individual polymers within the context of the topological relatedness of their surface chemistries. This provides a unique perspective that is complementary to the information provided by PLS and similar techniques -rather than focusing on specific fragment ions, the SOM can be used to identify entire subspaces in high dimensional space that correlate with high or low performance. Further, polymers with surface chemistries that are adjacent topologically, yet exhibit entirely different properties, can be identified and studied. Coupled with PLS, this can provide the necessary information to unravel the complicated relationship between surface chemistry and polymer performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

This work was performed in part using resources funded by the Australian National Fabrication Facility (ANFF), a company established under the National Collaborative Research Infrastructure Strategy, through the La Trobe

University Centre for Materials and Surface Science. The authors thank Robert Madiona, La Trobe University, for underpinning contributions in the use of self-organizing maps in the interpretation of ToF-SIMS data. The authors acknowledge the Milano Chemometrics and QSAR Research Group for the development of the Kohonen and CP-ANN Toolbox for MATLAB[33, 34]. Andrew Hook thanks the University of Nottingham for provision of his Nottingham Research Fellowship. Funding from the Wellcome Trust (103882 and 103884) is also kindly acknowledged.

References

1. Colliver TL, Brummel CL, Pacholski ML, Swanek FD, Ewing AG, Winograd N. Atomic and Molecular Imaging at the Single-Cell Level with TOF-SIMS. Anal Chem. 1997; 69(13):2225–2231. [PubMed: 9212701]
2. Pacholski ML, Cannon DM Jr, Ewing AG, Winograd N. Static time-of-flight secondary ion mass spectrometry imaging of freeze-fractured, frozen-hydrated biological membranes. Rapid Commun Mass Spectrom. 1998; 12(18):1232–1235. [PubMed: 9772765]
3. Brulet M, Seyer A, Edelman A, Brunelle A, Fritsch J, Ollero M, Laprevote O. Lipid mapping of colonic mucosa by cluster TOF-SIMS imaging and multivariate analysis in cftr knockout mice. J Lipid Res. 2010; 51(10):3034–45. [PubMed: 20616379]
4. Passarelli MK, Winograd N. Lipid imaging with time-of-flight secondary ion mass spectrometry (ToF-SIMS). Biochim Biophys Acta. 2011; 1811(11):976–990. [PubMed: 21664291]
5. Kokesch-Himmelreich J, Woltmann B, Torger B, Rohnke M, Arnhold S, Hempel U, Muller M, Janek J. Detection of organic nanoparticles in human bone marrow-derived stromal cells using ToF-SIMS and PCA. Anal BioAnal Chem. 2015; 407(16):4555–65. [PubMed: 25869483]
6. Passarelli MK, Newman CF, Marshall PS, West A, Gilmore IS, Bunch J, Alexander MR, Dollery CT. Single-Cell Analysis: Visualizing Pharmaceutical and Metabolite Uptake in Cells with Label-Free 3D Mass Spectrometry Imaging. Anal Chem. 2015; 87(13):6696–6702. [PubMed: 26023862]
7. Castellanos A, Ramirez CE, Michalkova V, Nouzova M, Noriega FG, Fernndez-Lima F. Three dimensional secondary ion mass spectrometry imaging (3D-SIMS) of Aedes aegypti ovarian follicles. J Anal At Spectrom. 2019; 34(5):874–883. [PubMed: 31680712]
8. Dowlatshahi Pour M, Ren L, Jennische E, Lange S, Ewing AG, Malmberg P. Mass spectrometry imaging as a novel approach to measure hippocampal zinc. J Anal At Spectrom. 2019; 34(8):1581–1587.
9. Henss A, Otto S-K, Schaepe K, Pauksch L, Lips KS, Rohnke M. High resolution imaging and 3D analysis of Ag nanoparticles in cells with ToF-SIMS and delayed extraction. Biointerphases. 2018; 13(3) 03B410
10. Sierros KA, Morris NJ, Ramji K, Cairns DR. Stress corrosion cracking of indium tin oxide coated polyethylene terephthalate for flexible optoelectronic devices. Thin Solid Films. 2009; 517(8):2590–2595.
11. Zhou LF, Liu ZY, Wu W, Li XG, Du CW, Jiang B. Stress corrosion cracking behavior of ZK60 magnesium alloy under different conditions. Int J Hydrogen Energy. 2017; 42(41):26162–26174.
12. Hook AL, Scurr DJ. ToF-SIMS analysis of a polymer microarray composed of poly(meth)acrylates with C6 derivative pendant groups. Surf Interface Anal. 2016; 48(4):226–236. [PubMed: 27134321]
13. Madiona RMT, Winkler DA, Muir BW, Pigram PJ. Optimal machine learning models for robust materials classification using ToF-SIMS data. Appl Surf Sci. 2019; 487:773–783.
14. Madiona RMT, Bamford SE, Winkler DA, Muir BW, Pigram PJ. Distinguishing Chemically Similar Polyamide Materials with ToF-SIMS Using Self-Organizing Maps and a Universal Data Matrix. Anal Chem. 2018; 90(21):12475–12484. [PubMed: 30260219]
15. Madiona RMT, Welch NG, Russell SB, Winkler DA, Scoble JA, Muir BW, Pigram PJ. Multivariate analysis of ToF-SIMS data using mass segmented peak lists. Surf Interface Anal. 2018; 50(7):713–728.
16. Madiona RMT, Winkler DA, Muir BW, Pigram PJ. Effect of mass segment size on polymer ToF-SIMS multivariate analysis using a universal data matrix. Appl Surf Sci. 2019; 478:465–477.

17. Welch NG, Madiona RMT, Payten TB, Jones RT, Brack N, Muir BW, Pigram PJ. Surface Adsorbed Antibody Characterization Using ToF-SIMS with Principal Component Analysis and Artificial Neural Networks. Langmuir. 2016; 32(34):8717–8728. [PubMed: 27494212]
18. Madiona RMT, Alexander DLJ, Winkler DA, Muir BW, Pigram PJ. Information content of ToF-SIMS data: Effect of spectral binning. Appl Surf Sci. 2019; 493:1067–1074.
19. Leo BF, Fearn S, Gonzalez-Cater D, Theodorou I, Ruenraroengsak P, Goode AE, McPhail D, Dexter DT, Shaffer M, Chung KF, Porter AE, et al. Label-Free Time-of-Flight Secondary Ion Mass Spectrometry Imaging of Sulfur-Producing Enzymes inside Microglia Cells following Exposure to Silver Nanowires. Anal Chem. 2019; 91(17):11098–11107. [PubMed: 31310103]
20. Celiz AD, Hook AL, Scurr DJ, Anderson DG, Langer R, Davies MC, Alexander MR. ToF-SIMS imaging of a polymer microarray prepared using ink-jet printing of acrylate monomers. Surf Interface Anal. 2013; 45(1):202–205.
21. Scoutaris N, Hook AL, Gellert PR, Roberts CJ, Alexander MR, Scurr DJ. ToF-SIMS analysis of chemical heterogenities in inkjet micro-array printed drug/polymer formulations. J Mater Sci Mater Med. 2012; 23(2):385–391. [PubMed: 22083856]
22. Urquhart AJ, Taylor M, Anderson DG, Langer R, Davies MC, Alexander MR. TOF-SIMS Analysis of a 576 Micropatterned Copolymer Array To Reveal Surface Moieties That Control Wettability. Anal Chem. 2008; 80(1):135–142. [PubMed: 18044847]
23. Yang J, Mei Y, Hook AL, Taylor M, Urquhart AJ, Bogatyrev SR, Langer R, Anderson DG, Davies MC, Alexander MR. Polymer surface functionalities that control human embryoid body cell adhesion revealed by high throughput surface characterization of combinatorial material microarrays. Biomaterials. 2010; 31(34):8827–8838. [PubMed: 20832108]
24. Hook AL, Williams PM, Alexander MR, Scurr DJ. Multivariate ToF-SIMS image analysis of polymer microarrays and protein adsorption. Biointerphases. 2015; 10(1) 019005 [PubMed: 25708635]
25. Epa VC, Hook AL, Chang C, Yang J, Langer R, Anderson DG, Williams P, Davies MC, Alexander MR, Winkler DA. Modelling and Prediction of Bacterial Attachment to Polymers. Adv Funct Mater. 2014; 24(14):2085–2093.
26. Hook AL, Anderson DG, Langer R, Williams P, Davies MC, Alexander MR. High throughput methods applied in biomaterial development and discovery. Biomaterials. 2010; 31(2):187–198. [PubMed: 19815273]
27. Celiz AD, Smith JG, Patel AK, Hook AL, Rajamohan D, George VT, Flatt L, Patel MJ, Epa VC, Singh T, Langer R. Discovery of a Novel Polymer for Human Pluripotent Stem Cell Expansion and Multilineage Differentiation. Advanced materials (Deerfield Beach, Fla). 2015; 27(27):4006–12.
28. Mei Y, Saha K, Bogatyrev SR, Yang J, Hook AL, Kalcioglu ZI, Cho SW, Mitalipova M, Pyzocha N, Rojas F, Van Vliet KJ, et al. Combinatorial development of biomaterials for clonal growth of human pluripotent stem cells. Nat Mater. 2010; 9(9):768–78. [PubMed: 20729850]
29. Zhang R, Mjoseng HK, Hoeve MA, Bauer NG, Pells S, Besseling R, Velugotla S, Tourniaire G, Kishen REB, Tsenkina Y, Armit C. A thermoresponsive and chemically defined hydrogel for long-term culture of human embryonic stem cells. Nat Commun. 2013; 4(1) 1335 [PubMed: 23299885]
30. Hook AL, Chang CY, Yang J, Atkinson S, Langer R, Anderson DG, Davies MC, Williams P, Alexander MR. Discovery of novel materials with broad resistance to bacterial attachment using combinatorial polymer microarrays. Advanced materials (DeerfieldBeach, Fla). 2013; 25(18):2542–7.
31. Hook AL, Chang CY, Yang J, Luckett J, Cockayne A, Atkinson S, Mei Y, Bayston R, Irvine DJ, Langer R, Anderson DG, et al. Combinatorial discovery of polymers resistant to bacterial attachment. Nat Biotechnol. 2012; 30(9):868–875. [PubMed: 22885723]
32. Kohonen T. Self-organized formation of topologically correct feature maps. Biol Cybern. 1982; 43(1):59–69.
33. Ballabio D, Consonni V, Todeschini R. The Kohonen and CP-ANN toolbox: A collection of MATLAB modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks. Chemometr Intell Lab. 2009; 98(2):115–122.

34. Ballabio D, Vasighi M. A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies. Chemometr Intell Lab. 2012; 118:24–32.
35. Gardner W, Cutts SM, Muir BW, Jones RT, Pigram PJ. Visualizing ToF-SIMS Hyperspectral Imaging Data Using Color-Tagged Toroidal Self-Organizing Maps. Anal Chem. 2019; 91(21):13855–13865. [PubMed: 31549810]
36. Mount N, Weaver D. Self-organizing maps and boundary effects: quantifying the benefits of torus wrapping for mapping SOM trajectories. Pattern Anal Appl. 2011; 14(2):139–148.
37. Andreu, G; Crespo, A; Valiente, JM. Selecting the toroidal self-organizing feature maps (TSOFM) best organized to object recognition; Proceedings of International Conference on Neural Networks (ICNN'97), 12-12 June 1997; 1997. 1341–1346.
38. Schmidt C, Rey S, Skupin A. Effects of Irregular Topology in Spherical Self-Organizing Maps. Int Reg Sci Rev. 2011; 34:215–229.
39. Jaccard P. Etude de la distribution florale dans une portion des Alpes et du Jura. Bull de la Soc Vaud des Sci Nat. 1901; 37:547–579.
40. Keenan MR, Kotula PG. Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images. Surf Interface Anal. 2004; 36(3):203–212.
41. Keenan MR, Kotula PG. Optimal scaling of TOF-SIMS spectrum-images prior to multivariate statistical analysis. Appl Surf Sci. 2004; 231-232:240–244.
42. Lee JLS, Gilmore IS, Seah MP. Quantification and methodology issues in multivariate analysis of ToF-SIMS data for mixed organic systems. Surf Interface Anal. 2008; 40(1):1–14.
43. Ultsch, A. Data Mining and Knowledge Discovery with Emergent Self-Organizing Feature Maps for Multivariate Time SeriesKohonen Maps. Oja, E, Kaski, S, editors. Elsevier Science B.V; Amsterdam: 1999. 33–45.
44. Celiz AD, Smith JGW, Patel AK, Langer R, Anderson DG, Barrett DA, Young LE, Davies MC, Denning C, Alexander MR. Chemically diverse polymer microarrays and high throughput surface characterisation: a method for discovery of materials for stem cell culture. Biomater Sci. 2014; 2(11):1604–1611. [PubMed: 25328672]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

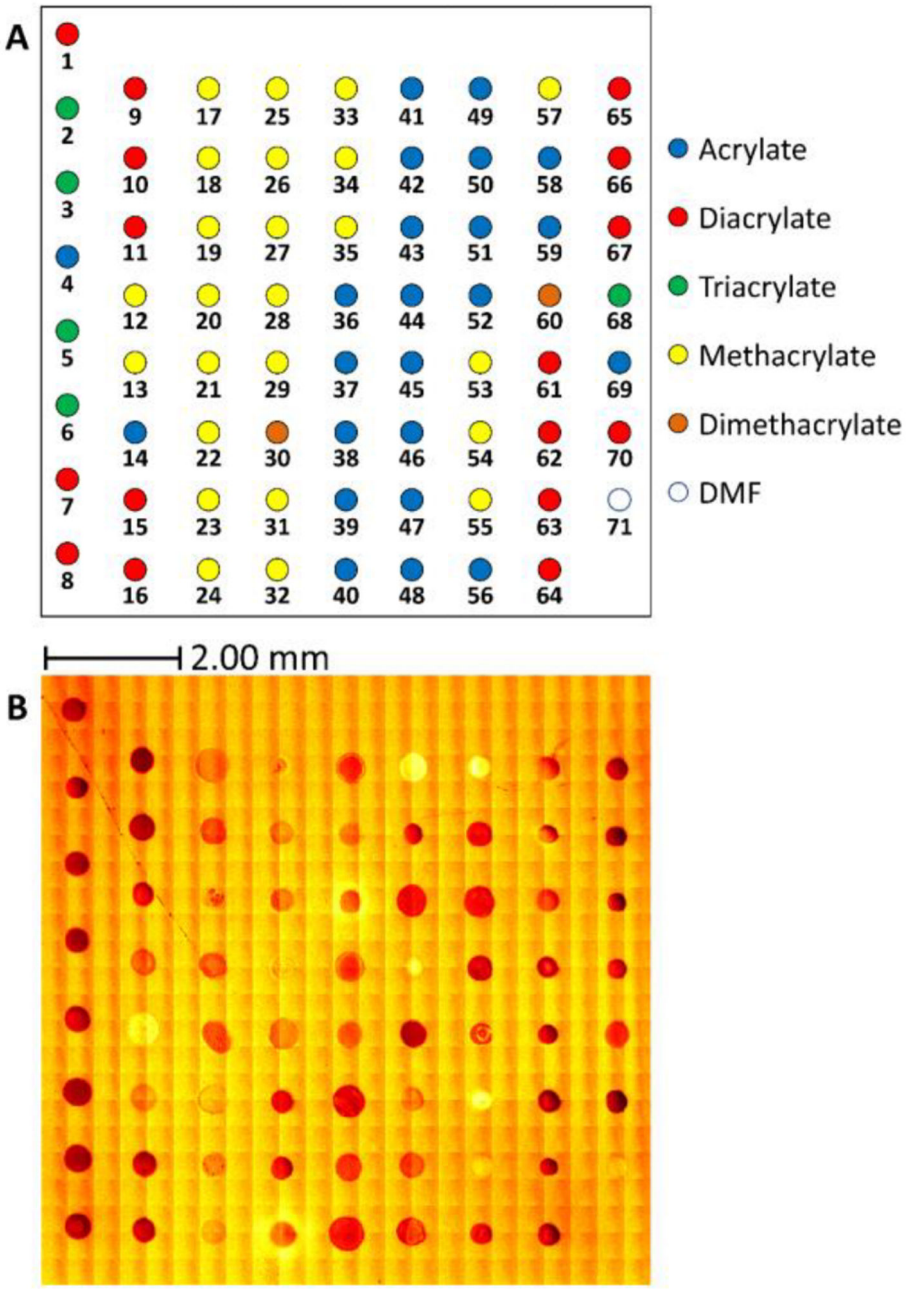


Figure 1.

(A) Schematic showing the layout of the 70 polymers in the microarray, in which each polymer is colored according to the broadly defined polymer groupings provided on the right. Spot 71 contained only the solvent, DMF. (B) Total ion image of the 9.2 × 9.2 mm analysis area obtained using negative ion mode. Figure adapted, with permission, from Hook et al. [24]

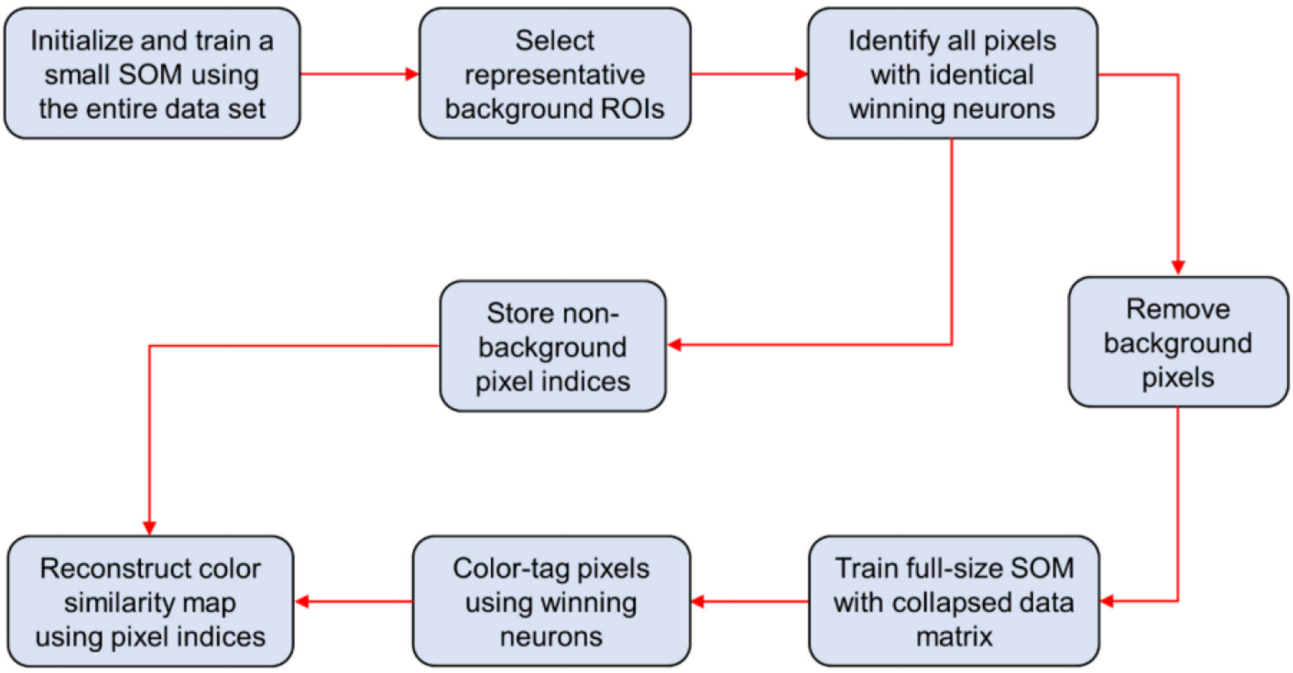


Figure 2.

Flowchart outlining the data processing and analysis workflow. Note that in this work eigenvalue initialization of the small SOM was used in the first step. This was sufficient to distinguish background pixels.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

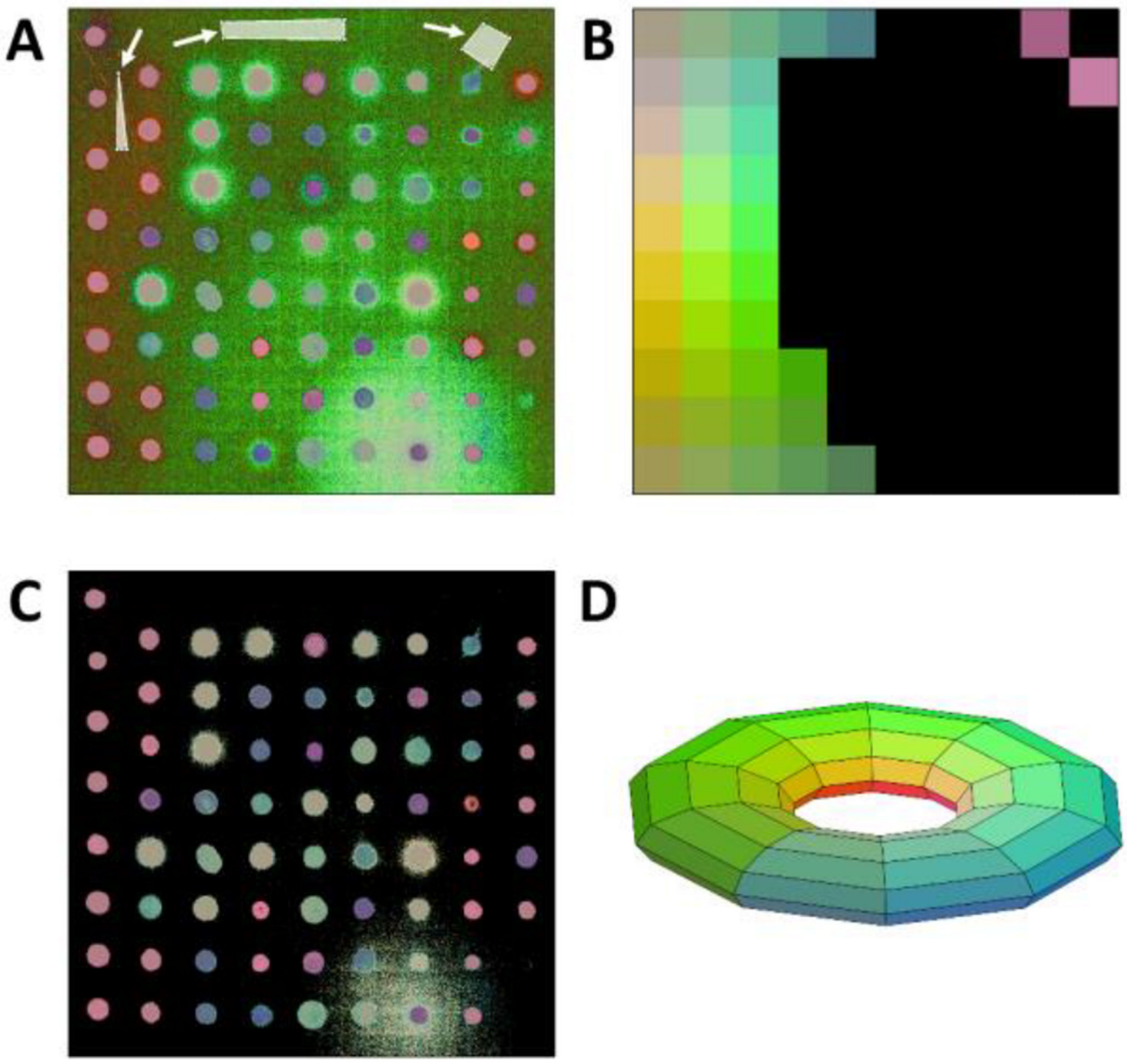


Figure 3.

Applying the data processing workflow presented in Figure 2 to the ToF-SIMS image of the polymer microarray. (A) Selection of three polygon ROIs representative of the entire set of background pixels, indicated by the white arrows. (B) The localization of the winning neurons associated with the pixels selected in (A) on the SOM, shown in black. This represents the cluster of neurons that are modelling the set of background pixel spectra. (C) Subtraction of background pixels sharing the winning neurons displayed in (B). (D) The origin of the neuron colors, which are based on their positions in the toroidal SOM. RGB colors are proportional to the *x*-, *y*- and *z*-coordinates of the neurons on the 3D toroidal surface, respectively.

Europe PMC Funders Author Manuscripts

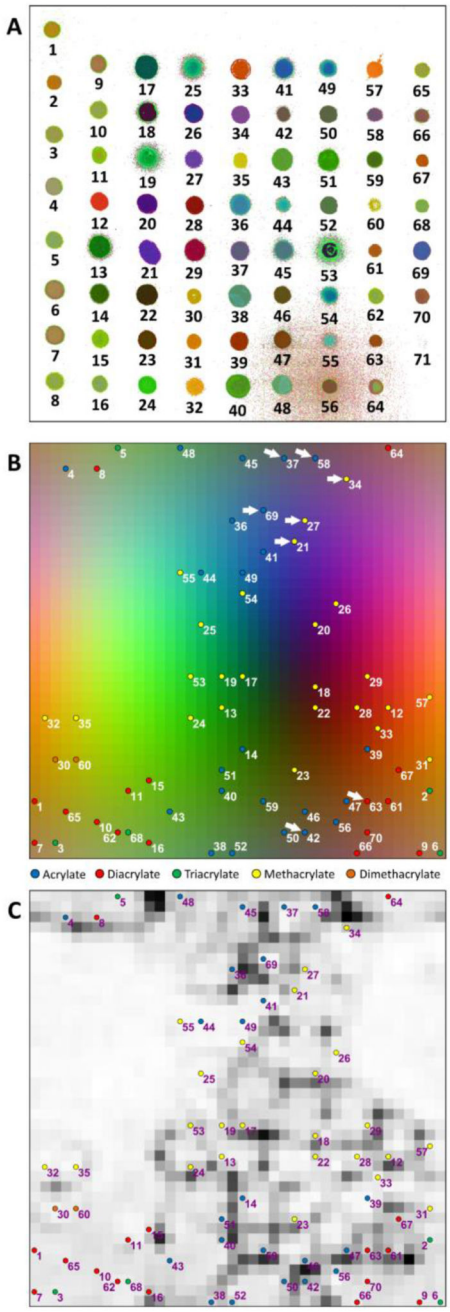


Figure 4.

(A) Reconstructed RGB similarity map after background subtraction and subsequent training with a 40 × 40 SOM. Pixels are colored by their winning neuron colors, such that spectral similarity between pixels is captured by their color similarity. White pixels represent background-subtracted pixels. (B) Localization of the most abundant winning neuron from each of the 70 polymers on the color-tagged SOM. Numbers correspond to the polymer numbers shown in (A), and the datapoint colors represent to the broad polymer grouping scheme shown below. Arrows point to the exclusive set of aromatic-containing polymers.

(C) Most abundant winning neurons, presented overlaid on the U-matrix for the SOM. The neurons in the U-matrix were assigned a greyscale value between 0-1, which is inversely proportional the mean distance to adjacent neurons, according to their weights.

Europe PMC Funders Author Manuscripts

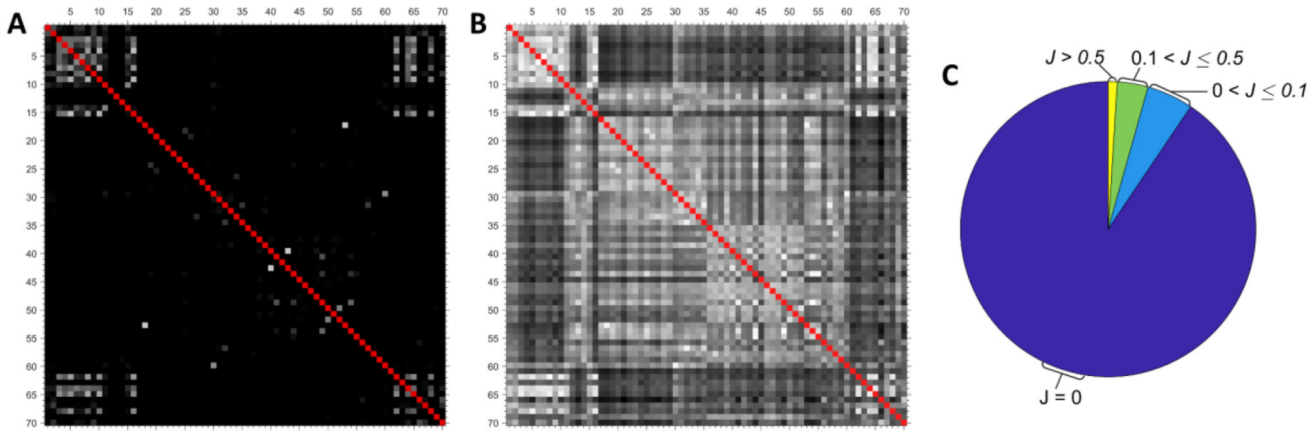


Figure 5.

(A) Greyscale matrix showing the Jaccard index for polymer pairs, where the brightness of each square is proportional to the index value. (B) Greyscale matrix showing the Euclidean distance between polymers, where the brightness is inversely proportional to distance. (C) Pie chart showing the distribution of Jaccard indices for the polymer pairs.

Europe PMC Funders Author Manuscripts

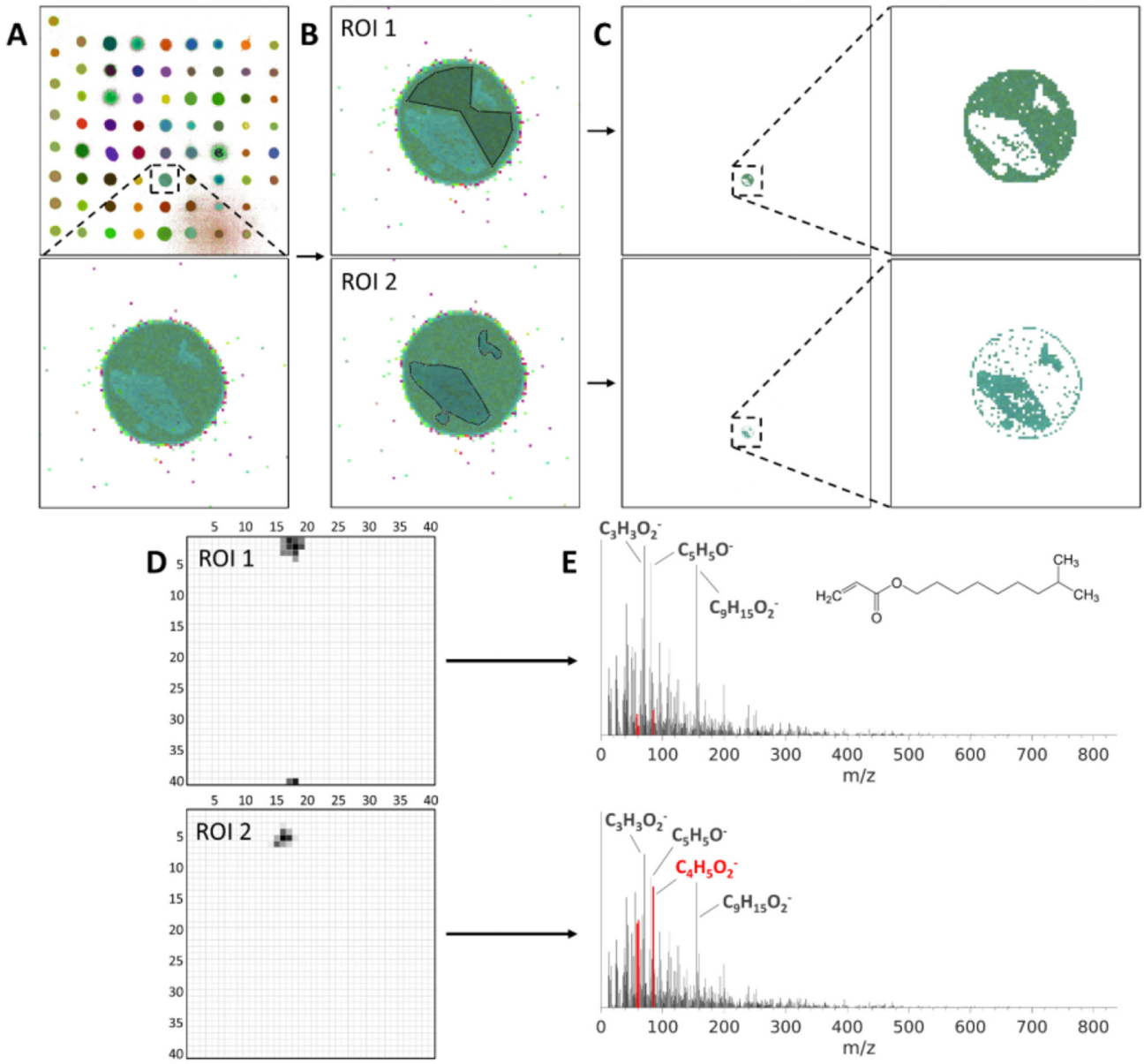


Figure 6.

Identification of intra-spot surface chemistry heterogeneity using the color-tagged SOM. (A) Reconstructed RGB similarity map of the polymer microarray, with a close-up of polymer spot 38 (isodecyl acrylate) illustrating the presence of two distinct regions. (B) Polygon selection of ROI 1 and ROI 2. (C) Distribution of the winning neurons in ROI 1 and ROI 2 across the entire analysis area, indicating localization within spot 38. Also shown is a close-up of spot 38, highlighting the separation of ROI 1 and ROI 2. (D) SOM localization for winning neurons located with ROI 1 and ROI 2, colored using a grey scale that represents relative neuron abundance within the ROI (with white equal to zero). (E) Weights extracted from each ROI using the winning neurons presented in (D), presented as a weighted mean (based on neuron abundance) and normalized to total weighting for each ROI. The labels

show some of the prominently weighted fragments originating from the polymer structure, and the red bars refer to three of the major differences in peak weightings between the two ROI.

Europe PMC Funders Author Manuscripts

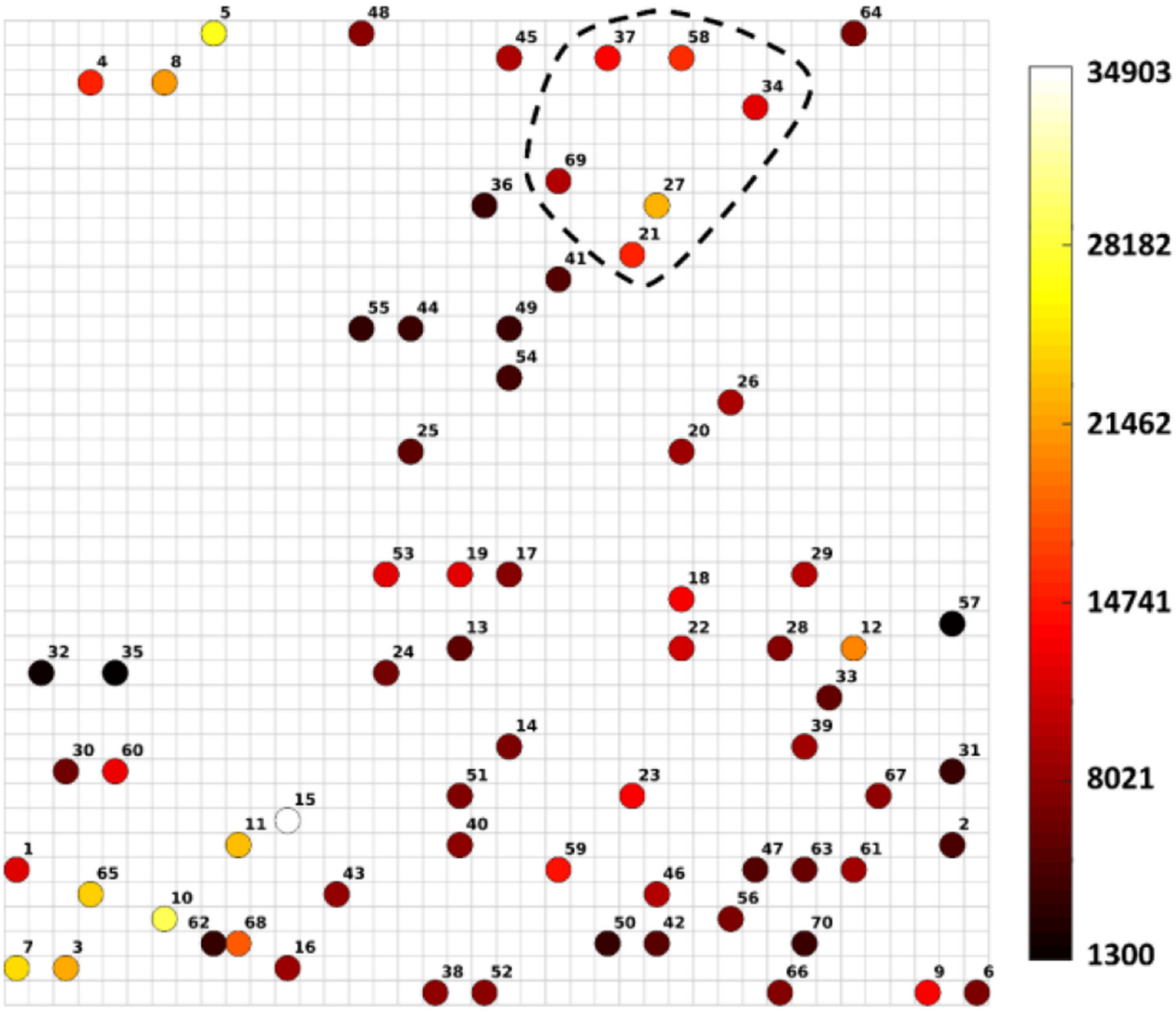
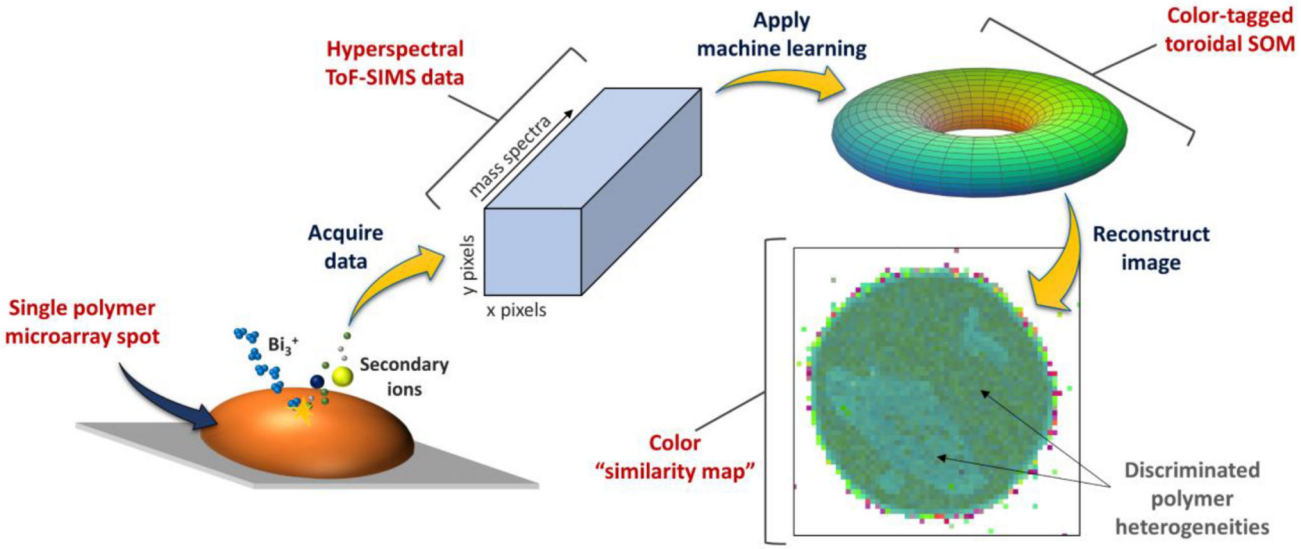


Figure 7.

Localization of the most abundant winning neuron from each of the 70 polymers on the color-tagged SOM, colored with a heat map that represents the mean protein binding (n = 8) for each polymer. The color bar on the right shows the arbitrary fluorescence scale. The inset outlines a cluster of aromatic-containing polymers exhibiting moderate-high protein adsorption.



Europe PMC Funders Author Manuscripts

Table 1

Top weighted peaks from each ROI in spot 38 (isodecyl acrylate) with peak assignments and weighting ratios.

Region	Peak m/z	Assignment	Weight in ROI 1 (%)	Weight in ROI 2 (%)	Ratio	Region	Peak m/z	Assignment	Weight in ROI 2 (%)	Weight in ROI 1 (%)	Ratio
ROI 1	71.01	$C_3H_3O_2^-$	2.91	2.35	1.24	ROI 2	71.01	$C_3H_3O_2^-$	2.35	2.91	0.81
	81.04	$C_5H_5O^-$	2.65	2.00	1.32		81.04	$C_5H_5O^-$	2.00	2.65	0.76
	155.09	$C_9H_{15}O_2^-$	2.62	1.92	1.36		155.09	$C_9H_{15}O_2^-$	1.92	2.62	0.73
	41.01	C_2HO^-	2.02	1.69	1.19		85.04	**$C_4H_5O_2^-$**	1.86	0.38	**4.92**
	67.02	$C_4H_3O^-$	1.8	1.42	1.27		55.02	$C_3H_3O^-$	1.77	1.28	1.38
	95.05	C_6H_7O-	1.42	1.16	1.22		41.01	C_2HO^-	1.69	2.02	0.84
	43.02	$C_2H_3O^-$	1.39	1.27	1.1		67.02	$C_4H_3O^-$	1.42	1.80	0.79
	51.02	$C_4H_3^-$	1.36	1.16	1.17		61.04	**$C_2H_5O_2^-$**	1.34	0.15	**9.24**
	111.05	$C_6H_7O_2$-	1.33	1.17	1.14		59.02	**$C_2H_3O_2^-$**	1.29	0.32	**4.00**
	69.04	$C_4H_5O^-$	1.32	1.05	1.25		43.02	$C_2H_3O^-$	1.27	1.39	0.91
	55.02	$C_3H_3O^-$	1.28	1.77	0.72		111.05	$C_6H_7O_2$-	1.17	1.33	0.88
	49.01	C_4H^-	1.19	1.13	1.06		95.05	$C_6H_7O^-$	1.16	1.42	0.82
	38.02	$C_3H_2^-$	1.06	1.00	1.06		49.01	C_4H^-	1.13	1.19	0.95
	79.02	$C_5H_3O^-$	1.04	0.91	1.14		69.04	$C_4H_5O^-$	1.05	1.32	0.8
	13.01	CH-	1.03	0.89	1.16		25.01	C_2H^-	1.04	1.22	0.85
	53.04	$C_4H_5^-$	1.00	0.79	1.27		45.00	CHO_2^-	1.00	0.83	1.2