Fluorescence, XPS and ToF-SIMS Surface Chemical State Image Analysis of DNA Microarrays

Supporting Information

Chi-Ying Lee,[1,3] Gregory M. Harbers,[4 5] David W. Grainger,[4,5] Lara J. Gamble[1,2] and David G. Castner,[1,2,3,*]

National ESCA and Surface Analysis Center for Biomedical Problems[1]
Departments of Bioengineering[2] and Chemical Engineering[3]
Box 351750 University of Washington
Seattle, WA 98195-1750

Departments of Bioengineering[4] and Pharmaceutics and Pharmaceutical Chemistry[5]
University of Utah
Salt Lake City, UT 84112-5820

Correspondence to:

David G. Castner
National ESCA and Surface Analysis Center for Biomedical Problems
Departments of Bioengineering and Chemical Engineering
Box 351750, University of Washington, Seattle, WA 98195-1750

Email: castner@nb.engr.washington.edu
Telephone: 206-543-8094
Fax: 206-543-3778

Principal Component Analysis (PCA)

PCA was performed using a series of scripts written by NESAC/BIO for MATLAB (MathWorks, Inc., Natick, MA). For each set of data all significant mass peaks (i.e., peaks that were at least 3x background, including some peaks related to the fragmentation of the DNA molecules which have previously been reported) were incorporated into the peak list for analysis. A conversion routine transforms the ToF-SIMS data (saved as a binary file) into a matrix usable by the MATLAB software. The data are then "unfolded" so that an image that was originally 128 by 128 pixels with *m* spectral variables (mass fragments) is reshaped to form a 2D array that is 16384 (128 x 128) by *m*. All data were normalized to the total intensity of the selected peaks and ''autoscaled'' prior to PCA. In this procedure, the data are first mean-centered by subtracting the column mean from each column, thus forming a matrix where each column has a mean of zero. Each mean-centered variable is then divided by its standard deviation, which results in variables with unit variance. This procedure places all variables on an equal basis in the analysis. This allows the less intense higher mass peaks to receive the same weighting in the analysis as the intense, low mass peaks. PCA is then applied to the scaled matrix data set. In PCA, a principal axis rotation of the variance-covariance matrix of the unfolded matrix data set is performed, resulting in a diagonalized matrix. The direction cosines between the new and old axes are eigenvectors calculated by the singular value decomposition of the variance-covariance matrix. The resulting variables, or principal components (PCs), are orthogonal to each other, and therefore uncorrelated. The values for these new variables are called the PC scores, and the matrix that transforms the original unfolded data matrix into the PC scores matrix (i.e. the eigenvectors matrix) is called the PC loadings matrix. PCA was run on the normalized and autoscaled image matrix. PC score images were constructed by plotting the PC value, or score, at each pixel. PCs are "sorted" in order of decreasing total variance; the first PC (i.e. the one capturing the largest amount of the variance from the original data set) is called PC1. The chemical meaning of each PC is found by interpreting the loadings plot. For each PC, peaks with a high, positive loading contribute significantly to a positive score of that PC, while peaks with a large, negative loading contribute significantly to a negative score of that PC.

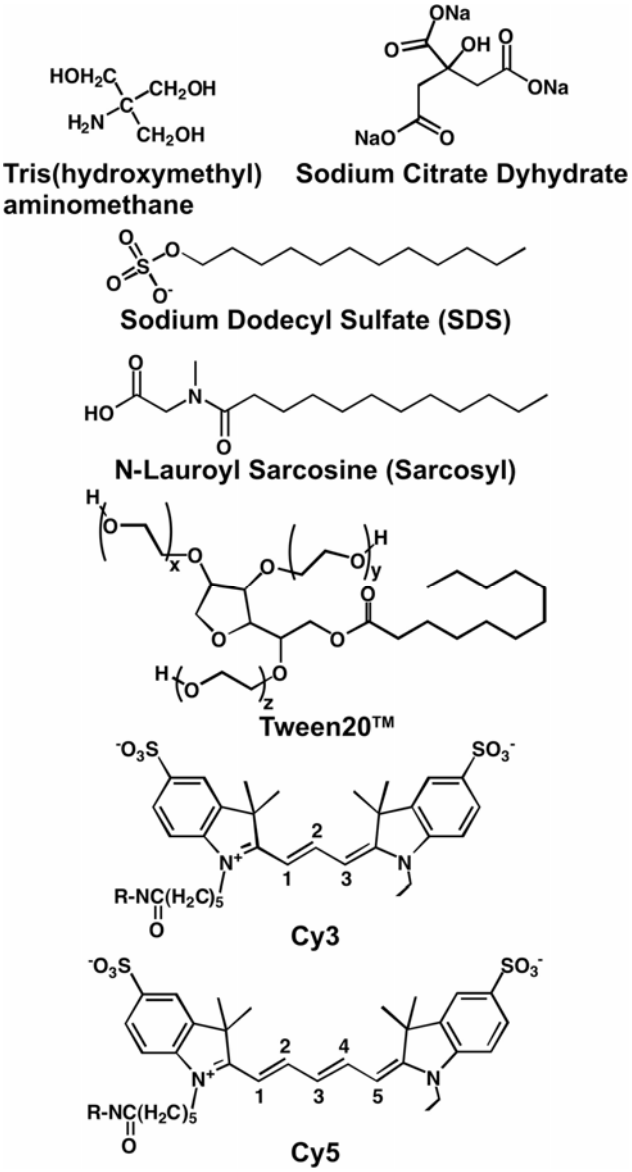

Figure S1. Chemical structures for the additives and components of the print and hybridization buffers as described in the main text as well as the structures for the Cyanine Dyes used as fluorescent tags on the DNA Probe Oligo (Cy3) and Target Oligo (Cy5). The structures have been provided to allow better assessment of the potential source of the molecular fragments identified in the ToF-SIMS data. For Tween20, x+y+z=20 ethylene oxide units; for Cy3 and Cy5, 1-5 reference respective backbone carbons that result in the specified Cy**3** and Cy**5** nomenclature.

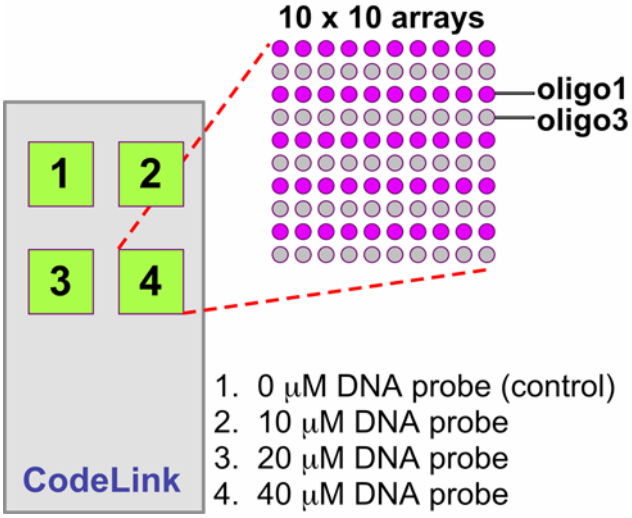

Figure S2. Layout of printed microarray regions on CodeLink slides.

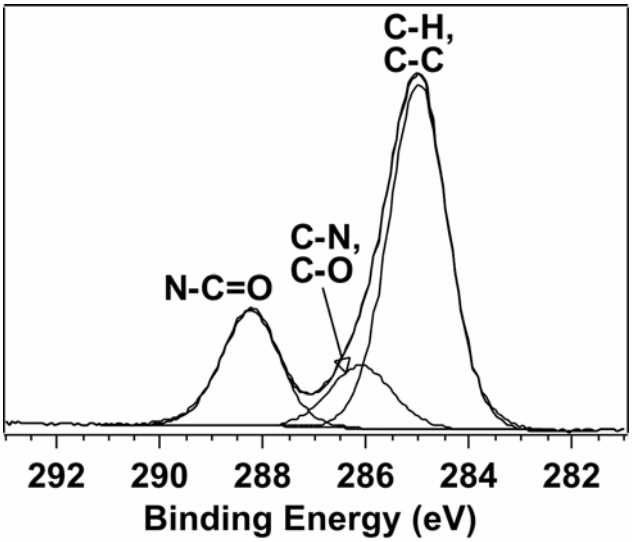

Figure S3. High-resolution XPS C1s spectrum and fits used for area measurements from a fresh, unmodified CodeLink microarray slide.

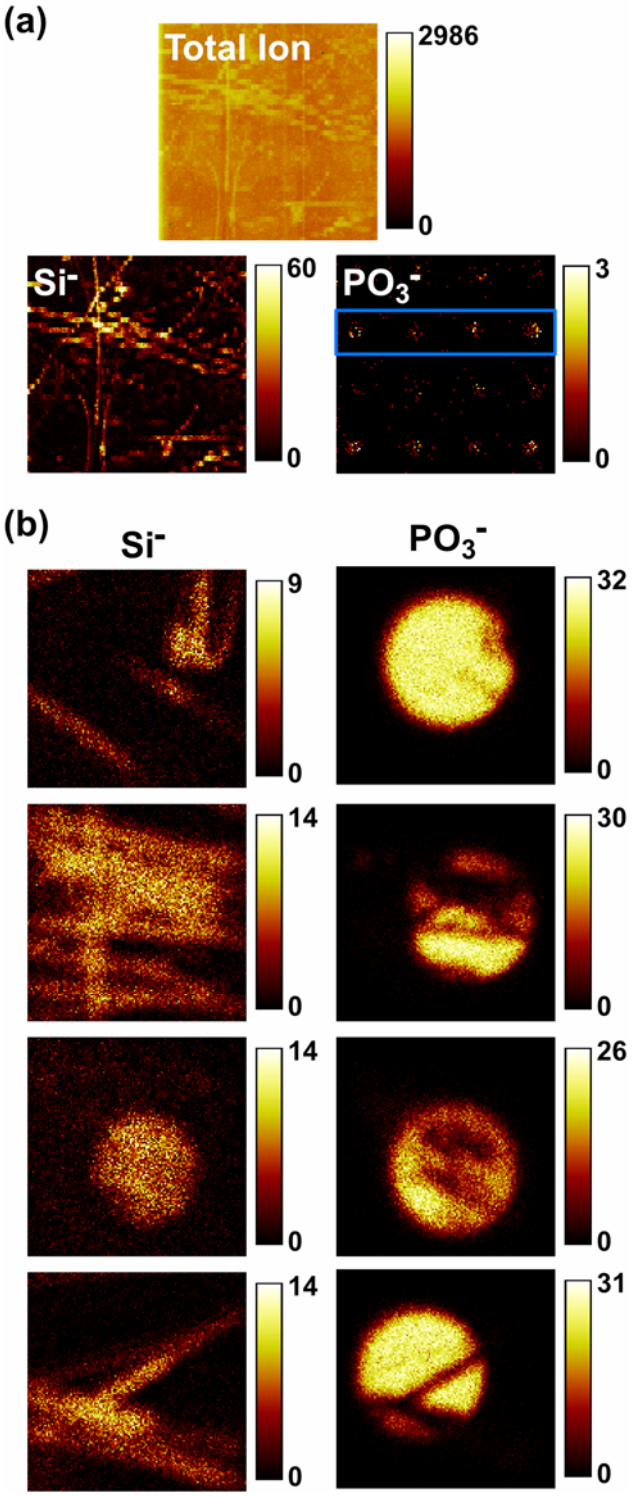

Figure S4. Representative negative ion ToF-SIMS images showing surface damage over a 1500 µm x 1500 µm microarray region (a). 200 µm x 200 µm images characteristic of the printed DNA (PO_3^-) and the substrate (Si^-), showing surface damage in the individual micro-spots within the microarray region (b).

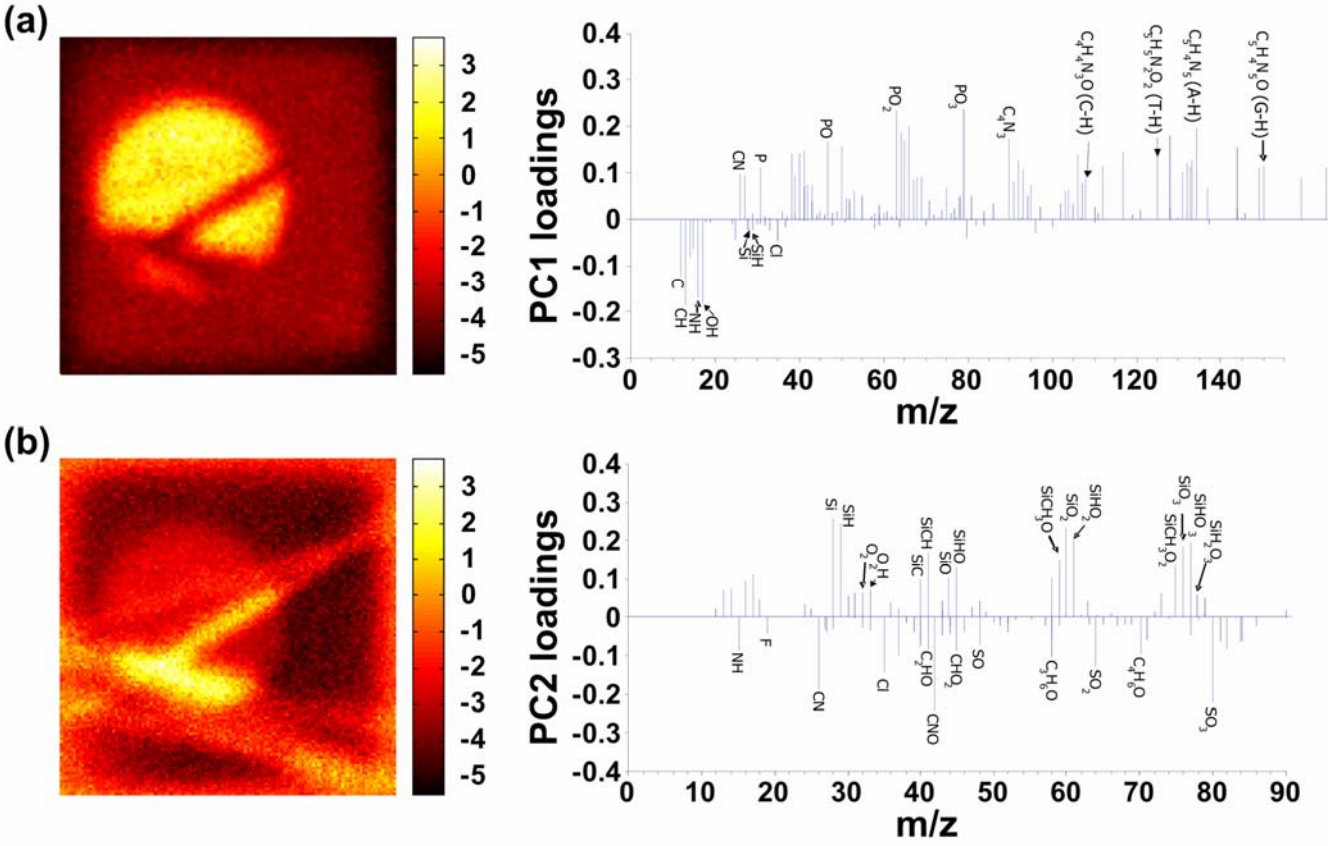

Figure S5. Image scores and loadings for PC 1 (a) and PC 2 (b) for a damaged micro-spot (negative ion image). The bright regions in image (b) correspond to the damaged areas. PC2 loadings from the negative ToF-SIMS image data matrix indicate that damaged areas are characterized by Si-containing fragments from the polymer-coated glass substrate, similar to those found associating with the "halo" feature around unhybridized probe micro-spots. Images are 200 µm x 200 µm.