Kinetic identification of protein ligands in a 51,200 small-molecule library using microarrays and a label-free ellipsometric scanner

James P. Landry, Andrew P. Proudian, Galina Malovichko, Xiangdong Zhu

Dept. of Physics, University of California at Davis, One Shields Ave, Davis, CA, USA 95616-5270

ABSTRACT

Drug discovery begins by identifying protein-small molecule binding pairs. Afterwards, binding kinetics and biofunctional assays are performed, to reduce candidates for further development. High-throughput screening, typically employing fluorescence, is widely used to find protein ligands in small-molecule libraries, but is rarely used for binding kinetics measurement because: (1) attaching fluorophores to proteins can alter kinetics and (2) most label-free technologies for kinetics measurement are inherently low-throughput and consume expensive sensing surfaces. We addressed this need with polarization-modulated ellipsometric scanning microscopes, called oblique-incidence reflectivity difference (OI-RD). Label-free ligand screening and kinetics measurement are performed simultaneously on small-molecule microarrays printed on relatively inexpensive isocyanate-functionalized glass slides. As a microarray is reacted, an OI-RD microscope tracks the change in surface-bound macromolecule density in real-time at every spot. We report progress applying OI-RD to screen purified proteins and virus particles against a 51,200-compound library from the National Cancer Institute. Four microarrays, each containing 12,800 library compounds, are installed in four flow cells in an automated OI-RD microscope. The slides are reacted serially, each giving 12,800 binding curves with ~30 sec time resolution. The entire library is kinetically screened against a single probe in ~14 hours and multiple probes can be reacted sequentially under automation. Real-time binding detection identifies both high-affinity and low-affinity (transient binding) interactions; fluorescence endpoint images miss the latter. OI-RD and microarrays together is a powerful high-throughput tool for early stage drug discovery and development. The platform also has great potential for downstream steps such as *in vitro* inhibition assays.

Keywords: label-free biosensing, biomolecular interaction analysis, small-molecule microarrays, chemical microarrays, high-throughput, drug screening, oblique incidence reflectivity difference (OI-RD), ellipsometry

1. INTRODUCTION

The objective of small-molecule compound screening is to find lead structures from which to develop therapeutically useful compounds. Drug discovery and development ultimately requires empirical assays of biomacromolecules (typically proteins) against libraries containing 10^4 to 10^6 compounds, from natural or synthetic sources. Traditional screening methods such as ELISA and yeast two-hybrid are mature and have been scaled to high-throughput formats, but require expensive robotics and can suffer from artifacts from the necessary artificial constructs and tags (e.g. fusion proteins and exogenous fluorophores). Over the last decade, microarrays (leveraging larger numbers of assays performed concurrently per unit area) and microfluidics (leveraging more assays performed per unit time) have emerged as major enabling platforms for high-throughput and cost-effective drug screening. After initial screening, it is understood that many "hit" compounds will be false-positives while others will fail at some step in a gauntlet of downstream biological and chemical selection tests. For example, when a fluorescently-labeled protein is observed to bind a compound in a small-molecule microarray, it is common to verify protein binding with a label-free real-time optical biosensor, such as surface plasmon resonance (SPR). Such an assay gives an independent validation of binding, removes the artificial fluorescent tag, and quantifies the reaction rate constants and binding affinity. The latter are important parameters for determining the best use of the lead structures. For example, high-affinity compounds might be effective on their own if they bind to the correct site while lower affinity compounds might reveal functional groups useful for fragment-based drug designs.

Although real-time label-free optical binding measurements give valuable information for the drug discovery pipeline, they are typically performed late in the process because traditional formats are low throughput and expensive to perform on a per reaction basis. There is thus a need to perform such measurements in a high-throughput and cost-effective format, earlier in the pipeline. To this end, we have developed an optical scanner for real-time and label-free detection of biomolecular binding in microarrays based on polarization-modulated oblique-incidence reflectivity difference

(OI-RD)[1-4]. The method measures changes in magnitude and phase of a laser beam when it reflects off a biomolecule covered solid surface, particularly in response to molecules binding to the surface. The method does not need specially structured substrates such as gold films, interference layers, gratings, or dielectric waveguide layers for detection and has a large "field of view" (presently ~60 cm^2). It is thus compatible with large microarrays printed on comparatively inexpensive chemically functionalized glass slides. In this paper, we report the successful scaling of the methodology to screen a library of 51,200 small-molecule compounds from the National Cancer Institute compound repository against a variety of purified protein probes and virus particles. The identified "hits" will be used for future *in vitro* inhibition assays using the same technology platform, followed-by independent verification assays.

2. METHODS AND MATERIALS

2.1 Oblique-Incidence Reflectivity Difference Scanning Microscope

The optical arrangement of the scanning OI-RD microscope is shown in Figure 1. A scan lens focuses a polarization-modulated laser beam (λ = 633 nm) to a 20-µm diameter spot on the back surface of a glass slide printed with a microarray. The incidence angle is θ = 36.6° inside the glass slide. The back surface is in contact with aqueous solution in a flow channel. Images of the microarray are obtained by scanning the focused beam across the back surface at a step size of 20 µm with a rotating mirror and scan lens along the *y*-axis and a translation stage that moves the microarray along the *x*-axis. Image "contrast" is based upon polarization changes of the laser beam upon reflection from the back surface. Specifically, let $r_p = \sqrt{R_p} \cdot \exp(i\phi_p)$ and $r_s = \sqrt{R_s} \cdot \exp(i\phi_s)$ be the *p*-polarized and *s*-polarized reflectivities when a layer of targets or probe-target complexes cover the surface. Further, let $r_{p0} = \sqrt{R_{p0}} \cdot \exp(i\phi_{p0})$ and $r_{s0} = \sqrt{R_{s0}} \cdot \exp(i\phi_{s0})$ be the reflectivities on the unprinted surface. The OI-RD scanner used in these studies directly measures the polarization changes $\Delta\delta = (\phi_p - \phi_{p0}) - (\phi_s - \phi_{s0})$ and $2 \cdot \Delta R = (R_p - R_{p0})/R_{p0} + (R_s - R_{s0})/R_{s0}$ in response to molecular binding on the surface where the light is reflected. For microarray detection, frequently the target, the probe, and the glass slide are transparent at the wavelength λ so probe binding primarily changes the phase signal $\Delta\delta$. When the thickness of the probe-target layer is small compared to the wavelength (valid for protein monolayers at visible wavelengths) $\Delta\delta$ varies linearly with the surface mass density Γ (mass per unit area) of the layer as[5-7]

$$\Delta\delta \cong \frac{-4\pi\sqrt{\varepsilon_s}\cos\theta}{(\varepsilon_0 - \varepsilon_s)(\cot^2\theta - \varepsilon_s/\varepsilon_0)} \cdot \frac{(\varepsilon_d - \varepsilon_0)(\varepsilon_d - \varepsilon_s)}{\varepsilon_d}\left(\frac{\Gamma}{\rho\lambda}\right), \tag{1}$$

where ε_s, ε_0, and ε_d are the dielectric constants of the glass slide, the aqueous solution, and the probe-target layer, respectively, ρ = 1.35 g/cm^3 is the volume mass density of globular proteins, and $d = \Gamma / \rho$ is the ellipsometric thickness of the target-probe layer. For typical values of ε_s = 2.31, ε_0 = 1.77 and ε_d = 2.03, a protein layer with Γ = 1 ng/mm^2 yields $\Delta\delta$ = -2.3×10^{-3}. The OI-RD signal plotted in the figures is $-\Delta\delta$, except where noted.
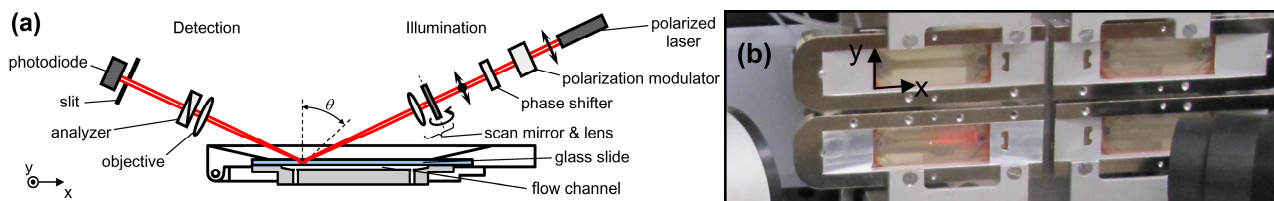

Figure 1. (a) Schematic representation of a scanning OI-RD microscope. Laser light from the OI-RD illumination arm is focused by a scan lens onto the bottom surface of a microarray-bearing glass slide. The laser light polarization is modulated by a photoelastic modulator and an adjustable phase shifter. Light reflected from the glass slide bottom surface is collected by an objective, passed through a polarization analyzer, separated from the top surface reflection by a slit, and converted to an electrical signal by a photodiode and lock-in amplifiers. The lock-in signal is zeroed on the unreacted microarray surface using the phase shifter so that as the beam is swept in space and time, resulting non-zero signal is proportional to the phase change in Eq. 1. (b) Four 2 cm × 5 cm flow cells, each containing a microarray of 13,960 spots, including 12,800 unique small-molecule targets. This arrangement allows a library of 51,200 compounds to be printed on four microarrays and measured in a single computer automated experiment over ~14 hours.

2.2 Small-Molecule Microarray Fabrication

We printed a library of 51,011 small organic molecules (molecular weights < ~800 Da) from the National Cancer Institute Developmental Therapeutic Program (NCI/DTP) including: (1) 60 compounds from the NCI-Challenge set, (2)

221 compounds from the NCI-Natural Products set, (3) 827 compounds from the NCI-Mechanistic Set, (4) 1,974 compounds from the NCI-Structural Diversity Set, (5) and 47,929 compounds from the NCI-Open Set. All of these compounds were provided at 1 mM concentration in DMSO in 96-well microplates (except for the Challenge, Natural Products, and Structural Diversity compounds, provided at 10 mM in DMSO). The library was printed on a set of 4 microarray slides, such as shown in Figure 2(a), using an OmniGrid 100 contact-printing robot (Digilab, Holliston, MA) and 100 μm diameter silicon printing pins (Parallel Synthesis, Santa Clara, CA). The microarrays consist of 74×185 = 13,690 spots over an area of 1.8 cm × 4.5 cm. The array was printed from the source microplates using a 2×5 arrangement of printing pins, resulting in the 9 mm square blocks seen in Figure 2. Within each pin block, borders along the upper-most row and right-most column were printed with 0.2 mg/mL bovine serum albumin (BSA) in 1× PBS, to serve as alignment guides. Additionally, 8 biotinylated-BSA spots (used to encode array identity) and 8 control peptides (used to encode block identity) were printed in the lower-left corner of each block. The remaining 1,280 spots in each block were printed with compounds from the library, yielding 12,800 compounds per slide and covering the entire ~51,200 compound library with four slides.

The compounds within the library come from a variety of sources and have diverse structures and functional groups. Therefore, the microarray slides had to be chemically functionalized to react with a wide range of groups to assure a reasonable percentage of the library remains covalently bound to the glass after washing. We adapted the methods of Bradner et al.[8-10], wherein the microarray glass is functionalized with isocyanate groups that can react covalently with many nucleophiles including primary and secondary amines, thiols, carboxylic acids, and alcohols[11]. To make isocyanate-functionalized glass slides in house, we used SuperAmine (ArrayIt, Sunnyvale, CA) functionalized slides by first coating the slides with short Fmoc-protected polyethylene glycol (PEG) spacers. After Fmoc removal using piperidine, 1,6-diisocyanatohexane was coupled to the surface by urea bond formation to deprotected PEG-amines, providing the isocyanate groups on the slides. Fresh slides were then printed with small molecules from the library microplates at concentrations of either 1 mM or 10 mM without further dilution and were subsequently exposed to pyridine vapor for several hours to catalyze slower isocyanate-nucleophile reactions. After catalysis, we stored the printed glass slides in a -20°C freezer. The unprinted areas of the isocyanate-functional slides were quenched and blocked with 1 mg/mL BSA in 1× PBS immediately before optical measurements, within the microscope flow cells.

(a) Dry Small-Molecule Microarray   **(b) Washed Small-Molecule Microarray**

OI-RD Signal x1000   4000 um   OI-RD Signal x1000   4000 um
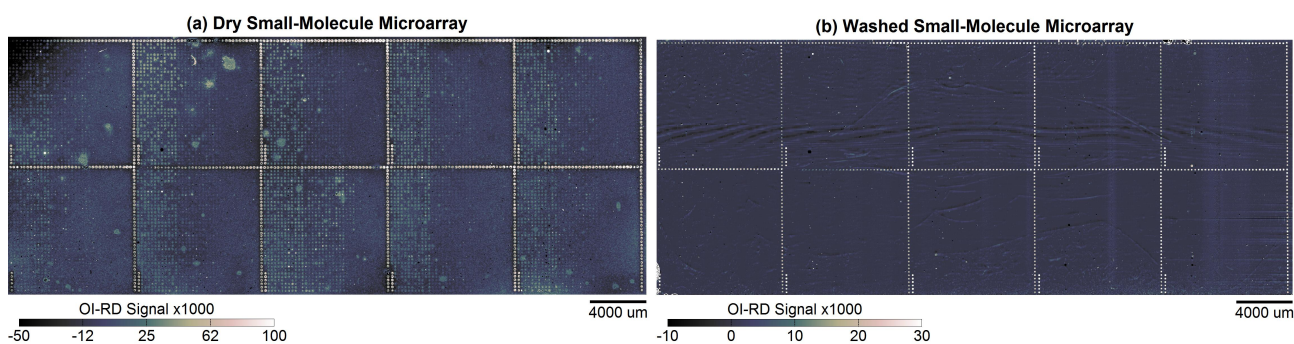-50 -12 25 62 100   -10 0 10 20 30

Figure 2. (a) OI-RD image of an unwashed ("dry") isocyanate-functionalized slide after printing with the NCI-DTP small molecule library at 1 mM concentration in DMSO. The microarray contains 13,690 spots, 12,800 of them distinct library compounds, over an area of 1.8 cm × 4.5 cm. The bright features along the top and right edges of each printing pin block are BSA borders, used for gridding and alignment. The compound spots vary in diameter from 60-150 μm and have varying contrast due to the differing surface tensions, refractive indices, and printing conditions (e.g. environmental humidity) for each compound. (b) OI-RD image of the same microarray after washing with 1× PBS buffer in the microscope flow cell. The microarray remains immersed in buffer for the remainder of the experiment. After washing, most printed compounds give little to no contrast due to the thinness of the immobilized target layer and/or refractive index matching to the glass or buffer. The BSA monolayers in the borders, however, maintain high contrast and therefore are used to construct a grid for locating all of the compound spots in the microarray.

2.3 Microarray Reactions

For end-point image analysis and real-time readout, a grid is extrapolated from the BSA borders (see supplemental methods in reference [3]). For real-time binding curve measurement, we select one pixel from each target spot and two pixels from the neighboring unprinted regions above and below each spot as references. We repeatedly scan and record the signals from this subset of pixels. The average signal from the two neighboring reference pixels is subtracted from the corresponding spot pixel, reducing background from instrumental drift, ambient refractive index changes, and flow-

induced signal transients. The background-corrected spot signal during a probe binding reaction constitutes a "binding curve". In the present experiments, all 27,565 pixels (13,690 target pixels + 13,875 reference pixels) are readout every 35 seconds.

After washing and blocking the microarray with a 1 mg/mL solution of BSA in 1× PBS for 30 minutes, we acquire an image (in buffer) before probe reaction, then flow probe solution to react the microarray (while recording all of the binding curves), and finally acquire an image after the reactions are complete. In the present experiments, the microarrays were reacted at ambient temperature (25 °C) with purified protein in buffer solution. For each reaction, we first acquire a baseline for 30 min (buffer flowing at 0.01 mL/min), quickly replace the buffer with probe solution and monitor the association reaction for 30-240 min (probe solution flowing at 0.01 mL/min for 30 min reactions), and lastly quickly replace the probe solution with buffer and monitor the dissociation reaction for 60 min (buffer flowing at 0.01 mL/min). Under computer control, multiple reactions can be performed serially on each of the four reaction chambers in the microscope. BSA blocking takes 30 min, each image scan takes 30 min, and each reaction is typically monitored for 120 min, for a total of 3.5 hours per chamber, or 14 hours total to screen the entire library of ~51,000 compounds against a single probe.

2.4 First-Pass Data Analysis and Hit Identification

The goal of the experiments was to identify "hit" compounds within the library which bind a purified protein or virion probe of interest. Hits on the microarrays are identified and filtered in a series steps. Currently, early steps are performed by computer algorithms on a subset of the data immediately after data acquisition, while later steps consider all available data (e.g. multiple data types, related reaction sequences, replicate experiments) and require human input. With OI-RD scanners, two types of complementary data are available from each reaction: (1) binding endpoint images and (2) real-time binding curves. We will briefly discuss how each data type is processed and how the data are successively filtered to identify hits.

First-pass endpoint analysis. Immediately before and after each reaction, an image of each microarray is acquired similar to Figure 2(b). By subtracting the initial image from the final image, we acquire an "endpoint image" showing only the net effect of probe binding to the surface at the end of the reaction, such as Figure 3(a) and 3(c). Ideally, all background features, such as stress birefringence in the microarray glass, subtract out of the endpoint image. In practice, some of these background features drift during the reaction and the residual must therefore be removed from the endpoint image using methods described in reference [3]. In this paper, residual background in endpoint images was isolated by filtering out the reacted microarray spots using a 5×19 pixel median filter. The residual background was then subtracted from the original endpoint image to give the final "flattened" endpoint images shown in Figure 3(a) and 3(c). In these images residual strain can still be seen because the spatial variation of the background in some regions is comparable in length to the spot-to-spot separation and therefore the median filter fails to discriminate between spot and background. This problem is specific to the current flow cell design, not the OI-RD method or data processing; with an improved design, we can reduce the stress applied to the glass and thus reduce the magnitude and increase the length scale of the stress birefringence.

After background subtraction, the endpoint image is algorithmically analyzed to identify spots that reacted. For this, the image is partitioned into small rectangular regions around each nominal spot location, determined from the grid extrapolated from the BSA borders. "Reacted spot" pixels are segmented from "background pixels" by computing the median (M) and median absolute deviation (MAD) of each of these regions. The regions are sized so that for typical spot diameters (~100 μm), the spot pixels occupy roughly 20% of the area and therefore are reliably segmented from background pixels by setting thresholds at $M \pm 3 \cdot MAD$. The resulting binary "blobs" are morphologically filtered and microarray regions containing blobs above a threshold size and threshold signal-to-background ratio are flagged as "endpoint hits" in this analysis round. For example, all of the regions highlighted with light green and red boxes in Figure 3(a) and 3(c) were flagged by this process. The thresholds were set to accept greater numbers of false positives in trade for fewer false negatives. False positives occur in regions of un-subtracted stress birefringence, the edges of air bubbles (from leaks in the flow cells), and occasionally along BSA borders (due to conformational changes or desorption from the thick BSA borders, uncorrelated to the probe reaction as revealed by real-time binding curves). False negatives often occur for very large spots that occupy a large fraction of the partition area (>50%), wherein the intensity based segmentation fails. We visually scan the images for "obvious" false negatives and manually flag them before further analysis.

First-pass real-time analysis. During the probe reaction, binding curves such as shown in the insets of Figure 3(b) and 3(d) are acquired for *every* spot in the microarray. The vertical dashed lines in the real-time traces separate the initial baseline (nominally flat) from the association reaction (where probe is introduced to the microarray) and the final dissociation reaction (where probe solution is removed from the flow cell). Since we are interested in compounds *with high or low affinity*, in the first-pass analysis we calculate the "association signal-to-noise ratio" for each spot, defined as the median signal of the binding curve for 10 data points before the transition to the dissociation phase (second vertical dashed line in insets) divided by the median absolute deviation of the baseline signal. We create maps of the association SNR to visualize the real-time data globally. All spots yielding association SNR > 4 are flagged as "real-time hits" (light green and red boxes in Figure 3(b) and 3(d)). As seen in these figures, the real-time data is particularly susceptible to false positives in the high-stress regions and BSA borders, but is less susceptible to false positives from air bubbles than endpoints.

2.5 Hit Refinement

In the experiments reported here, ~50-1,000 spots per microarray were algorithmically flagged as endpoint or real-time hits in the first-pass analysis, dependent upon the actual reactivity of the probe, the concentration and molecular weight of the probe, and the random background and noise. The corresponding endpoint images and real-time binding curves are imported into a database along with associated metadata (target/compound identity, probe identity and concentration, microarray ID, experiment ID, etc). Within the database a variety of "data features" are calculated from the imported data. For endpoint images, features include: spot diameter, spot signal, spot SNR, spot centering, spot principal moment of inertia ratio, and partition standard deviation. For real-time binding curves, features include: baseline noise, association SNR, dissociation SNR, total curve standard deviation, reference correlation, and Langmuir binding model least-squares fit association rate, dissociation rate, and equilibrium dissociation constant (binding affinity). The database format allows more flexibility for calculating and considering many classification features compared to the software mechanisms used in the first-pass analysis. At this stage, false positives due to air bubbles are easily identified by thresholding the total binding curve or image partition standard deviations. Also, false positives along the BSA borders are easily eliminated using the metadata (i.e. spots labeled "Border" are ignored). The remaining "hit candidates" and their associated features are good inputs for a multitude of machine learning algorithms[12]. Supervised algorithms are most appropriate for this application, but they require "training sets" of data wherein each hit candidate is manually flagged as being either a "true hit" or a "miss" (or other classifications, as appropriate). Therefore, for now we employ human inspection to further refine hits and misses. The present data sets will be used as training sets for hit refinement algorithms at a later time.

We manually refined the first pass hits as follows: For a given probe reaction, all real-time and endpoint hit candidates from the small-molecule library were imported into a database file. We then examined the subset of data which were flagged in the first round by *both* endpoint and real-time modes. *Coincidence* of detection in both modalities correlates strongly with true hits. However, true hits can be missed in one channel but not the other for a variety of reasons. For example, a small-diameter true hit can be missed in the real-time mode because the extrapolated readout pixel missed the spot, but the endpoint analysis still correctly identifies the spot within the grid partition. As another example, a piece of dirt might settle on the surface and be flagged by endpoint analysis, but the binding curve will reveal the event does not correlate with the introduction and removal of probe solution to the microarray. Therefore, the remainder of the hits scoring in only one mode (real-time or endpoint) are also inspected, but prioritized by data features that correlate strongly with true hits (e.g. spot SNR, spot moment of inertia ratio, association SNR, reference correlation). At the end of this process a list of "semi-final hits" are identified for each reaction. For each semi-final hit, all available data (real-time and endpoint) from all available reactions (related reaction sequences or replicate experiments) are imported into another database file. We then make a final judgment about the proper classification of the probe-target interactions based upon all available data.
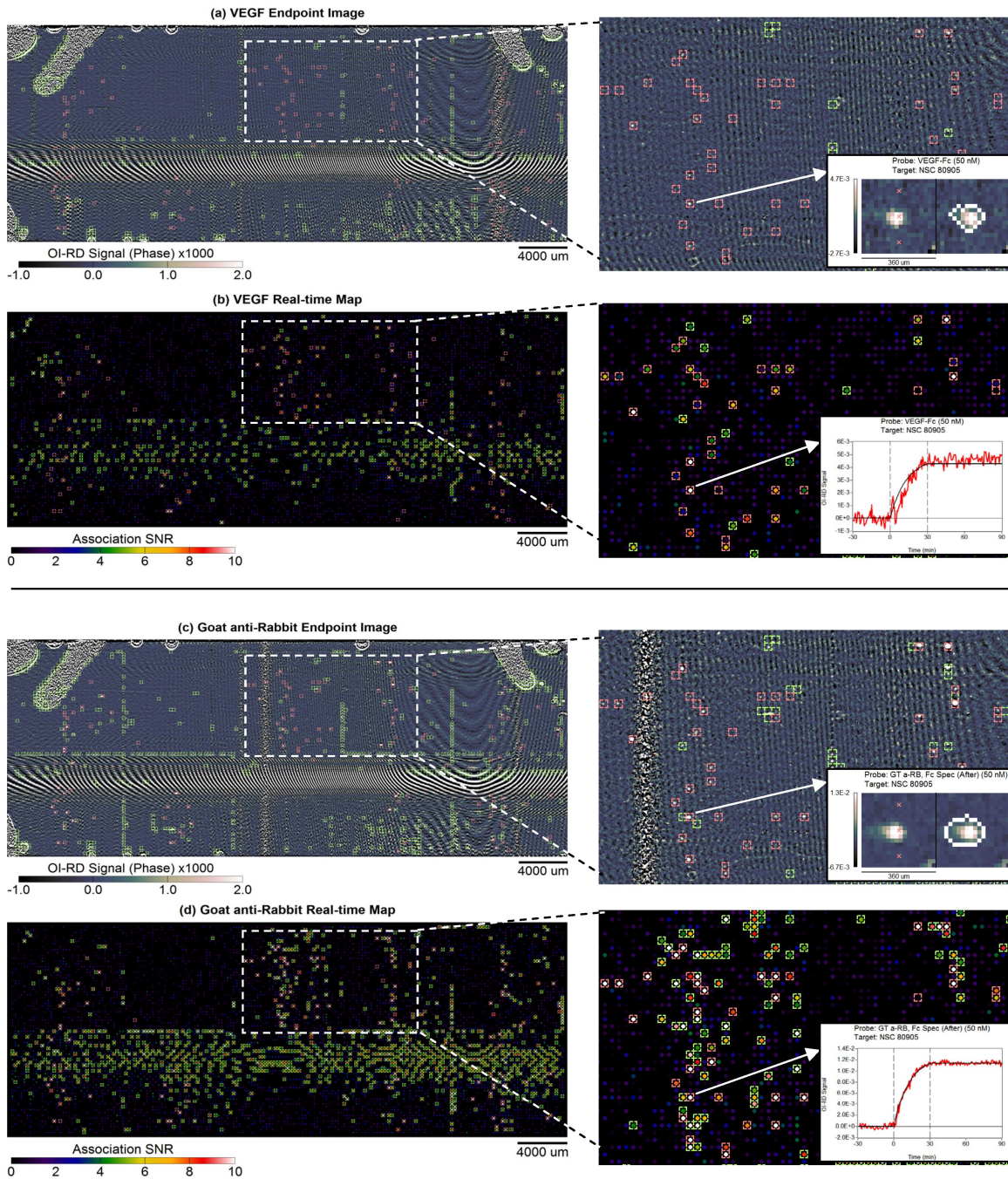
Figure 3. (a) and (b) show global views of a reaction of one microarray slide (1/4 of the library) with a VEGF-Rabbit IgG Fc fusion protein at 50 nM. (c) and (d) show global views of a secondary reaction of the *same* slide with polyclonal Goat anti-Rabbit Fc antibody at 50 nM. (a) and (c) show *endpoint images* of the reactions, while (b) and (d) show *real-time maps* of the reactions, derived from the $-\Delta\delta$ signal. The small boxes show spots that were flagged as "hits" in first-pass analysis. The light red boxes indicate spots that remained flagged as hits in later analysis, while the light green boxes indicate spots later determined to be false positives. The side panels to the right show a magnified view of a region with many hits, while the insets show examples of endpoint and real-time analyzed data. In the endpoint insets, the left image shows the image region analyzed and the light red crosses indicate the pixels used for target and reference real-time readouts, while the right image shows the same region with a highlight around the algorithmically detected spot pixels. In the real-time insets, the binding curve (target signal compensated by reference signals) is displayed and the dashed vertical lines separate the baseline, association phase (probe introduced to flow cell), and dissociation phase (probe removed from flow cell).

3. RESULTS

3.1 VEGF Protein Ligands

Our first probe of interest was vascular endothelial growth factor[13-16] (VEGF), a protein ligand that binds to corresponding cell surface receptors (VEGFR) in the biochemical pathway for the growth of blood vessels (angiogenesis), including to cancer tumors. Much effort has gone into the development of small molecule and antibody inhibitors of this process for cancer therapy, targeting either VEGF or its receptors[17-19]. We previously used OI-RD to screen a subset of 8,000 compounds from the NCI library and identified high-affinity VEGF ligands which were then further screened in real-time OI-RD-based VEGF-VEGFR2 binding inhibition experiments, resulting in identification of 12 novel small molecule inhibitors (manuscript in review). In the present experiments, we seek new VEGF ligands within the larger 51,200 compound library. Our VEGF probe is a fusion protein containing VEGF-A coupled to Rabbit IgG Fc domain and has a 48 kDa molecular weight (Epitomics, Burlingame, CA). The Fc domain was used for expressed protein purification with Protein A/G and for sandwich assays with a polyclonal Goat anti-Rabbit Fc antibody (Jackson ImmunoResearch, West Grove, PA). If the fusion protein is observed to bind to compounds in the library, it is natural to ask if it bound to the VEGF domain (desired) or to the Rabbit Fc domain (undesired). To discriminate between these possibilities we performed three reactions sequences on the library (separate slides for each sequence) as follows: [Sequence 1] (1) 50 nM Goat anti-Rabbit Fc, (2) 50 nM VEGF-Fc, (3) 50 nM Goat anti-Rabbit Fc; [Sequence 2] (1) 50 nM Goat IgG, (2) 50 nM VEGF-Fc, (3) 50 nM Goat anti-Rabbit Fc; [Sequence 3] (1) 50 nM monoclonal Rabbit IgG, (2) 50 nM VEGF-Fc, (3) 50 nM Goat anti-Rabbit Fc. The primary goal was to observe direct binding of VEGF to compounds; however, the secondary Goat anti-Rabbit Fc reactions greatly amplified the primary VEGF binding signal because the polyclonal antibodies are larger in mass and bind many epitopes, allowing VEGF signals at or just below the direct-detection limit to be easily observed. If binding is observed in the primary VEGF-Fc and secondary Goat anti-Rabbit Fc reactions, then it is clear that the target compound is a ligand to the VEGF-Fc fusion protein. If negligible binding response is also observed for Rabbit or Goat IgG, this suggests the compounds bind directly to the VEGF domain of the probe and are therefore of interest for follow-up inhibition assays.

From the library of 51,011 compounds, 336 semi-final hits were identified from analysis of Reaction Sequence 1. Data from these compounds was then compiled from all three Reaction Sequences and used to identify the most promising VEGF ligands for future experiments. 39 compounds gave good quality direct VEGF-Fc and secondary anti-Rabbit Fc reactions *and* showed no evidence of direct Rabbit IgG binding, such as shown in Figure 4(a) and 4(b). Such compounds are prime candidates for follow-up inhibition experiments. Among these candidates was one compound that we previously found to inhibit VEGF-VEGFR2 binding. Additionally, 44 compounds gave good primary and secondary VEGF reactions *but* showed evidence of direct Rabbit Fc binding in Sequence 3, such as shown in Figure 4(c), and therefore are less likely to be of interest in further experiments. Another 63 compounds appear to be false positives and are not worth further consideration, while the remaining 190 compounds gave lower quality or less reproducible binding signals (among the three reaction sequences) and are only worth further consideration if the 39 best candidates prove uninteresting. A complete listing of these compounds will be published in a forth coming publication.

3.2 Pyl1 and Pyl2 Protein Ligands

Our second probes of interest were Pyl1 (M_W = 25 kDa) and Pyl2 (M_W = 20 kDa) proteins from the model plant *Arabidopsis Thaliana*. These are members of a recently discovered class of abscisic acid (ABA)-binding signal transduction proteins, involved in plant resistance to environmental stress and plant development[20, 21]. We screened the library serially with 200 nM Pyl1 and then 200 nM Pyl2 (on a single set of microarray slides). In this single screen we identified 35 Pyl1 and 5 Pyl2 "semi-final hits", such as shown in Figure 5.

3.3 RNPC1a Protein Ligands

Our third probes of interest were RNPC1a (M_W = 25 kDa) and a mutant version, DRNPC2. RNPC1a is an RNA-binding protein and a target of the p53 family of proteins[22-24], and is thus an important component in the regulatory network responsible for cancer development. DRNPC2 is a mutant version of RNPC1a with a critical portion of the RNA-binding domain sequence knocked-out. We only had enough probe to perform experiments on 25% of the library (one slide), and thus focused on the slide that had most of the special compounds (Natural Products, Mechanistic, etc) as well as a large number Open Set compounds. We screened the library subset twice with RNPC1a at 100 nM and twice with DRNPC2 at 100 nM. Out of the 12,800 screened compounds, we identified one compound that reproducibly bound both

proteins with an affinity ~100 pM (Figure 6). Interestingly, the observed on-rate of the mutant DRNPC2 is roughly three times higher than RNPC1a.

3.4 Influenza A/PR8/34 (H1N1) Virion Ligands

Our fourth and final probe of interest was influenza A/PR/8/34 (H1N1) virus particles[25, 26]. The influenza virus particles are large macromolecular assemblies (~100 nm diameter) consisting of a lipid membrane embedded with a number of glycoproteins and surrounding single-stranded RNA. Thus, binding reactions on the chemical microarrays are biochemically less constrained than with purified protein probes, but more so than with whole cells (with heterogeneous surface compositions) or cellular lysate probes. Optically, because the particles are so large, the polarization response is no longer well described by the thin film result in Equation 1. In fact, for the influenza A particles the amplitude signal ΔR is most important rather than the amplitude signal $\Delta\delta$ dominant for protein monolayers presented before. We screened the entire 51,200 compound library and identified 18 compounds that bound the virus particles (Figure 7). The ability to measure both the optical phase $\Delta\delta$ and amplitude ΔR signals enables OI-RD to track binding of a wide variety of large biological particles on microarrays in both endpoint and real-time modes.
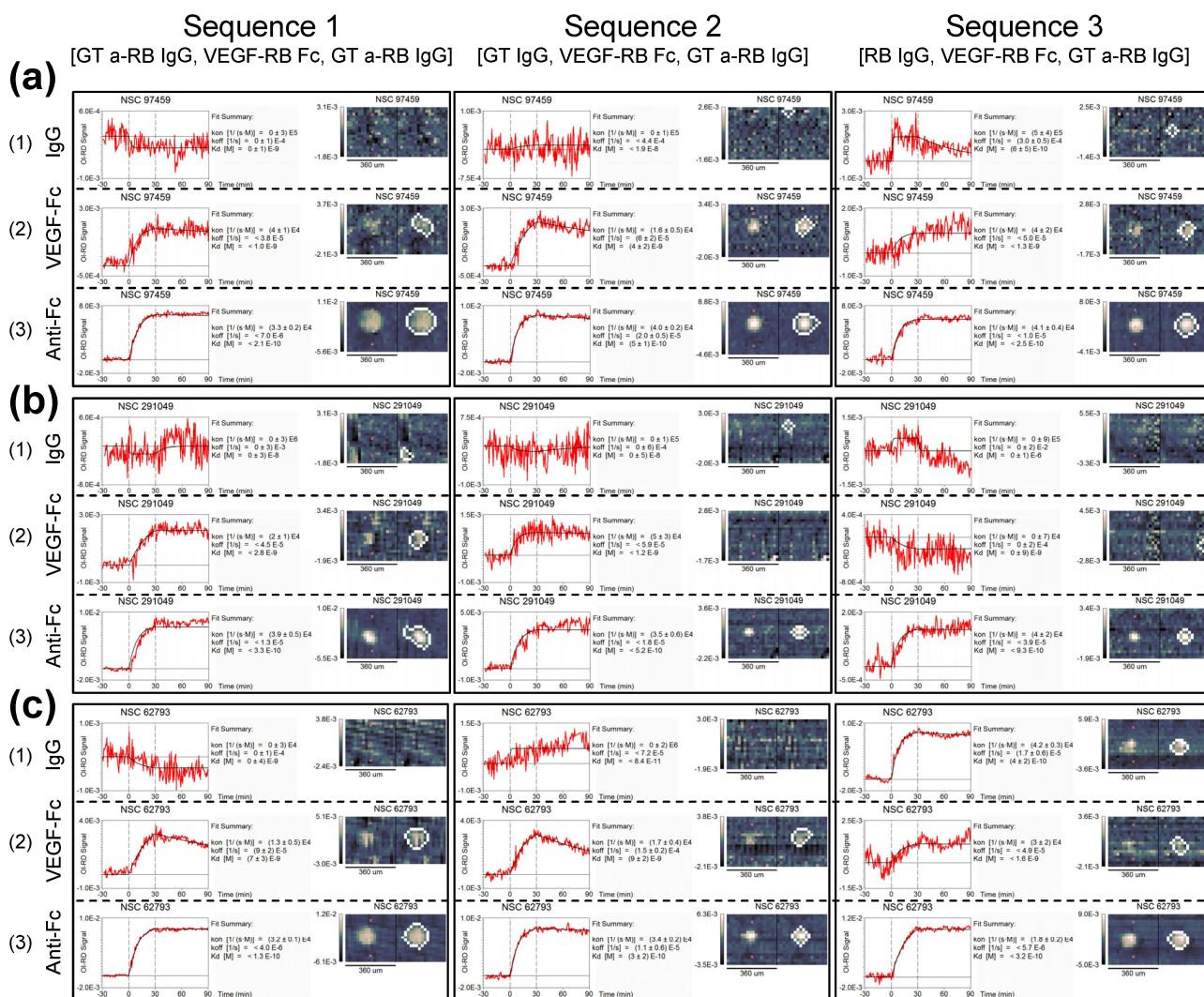

Figure 4. Rows (a) and (b) show complete screening $-\Delta\delta$ signal data for two library compounds that bind to the VEGF domain of the VEGF-Fc fusion protein probe with an equilibrium dissociation constant ~1 nM (determined from direct VEGF-Fc binding curves, sub-row 2). These compounds are considered "final hits" because direct VEGF-Fc binding (sub-row 2) and secondary Goat anti-Rabbit Fc binding (sub-row 3) were observed in multiple modes (real-time and endpoint), in multiple reactions (Sequence columns 1-3), *and* there is little evidence that Rabbit IgG (in reaction Sequence column 3, sub-

row 1) binds to the compounds by itself. Such compounds are prime candidates for further VEGF-receptor inhibition studies. Row (c) shows complete screening data for a compound that appears to capture the VEGF-Fc probe through the Rabbit Fc domain (as indicated by the direct binding of Rabbit IgG in Sequence 3, sub-row 1).
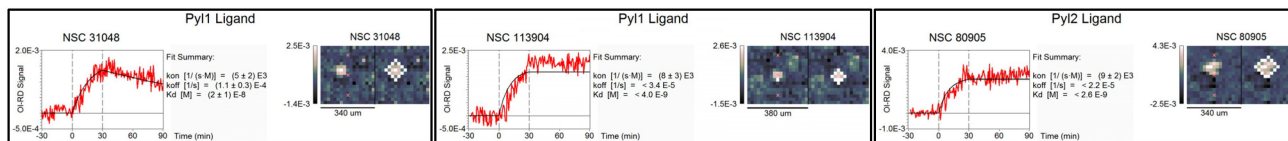

Figure 5. Representative $-\Delta\delta$ signal data hits for Pyl1 and Pyl2 proteins screened against the library serially (Pyl1 followed by Pyl2) at 200 nM.
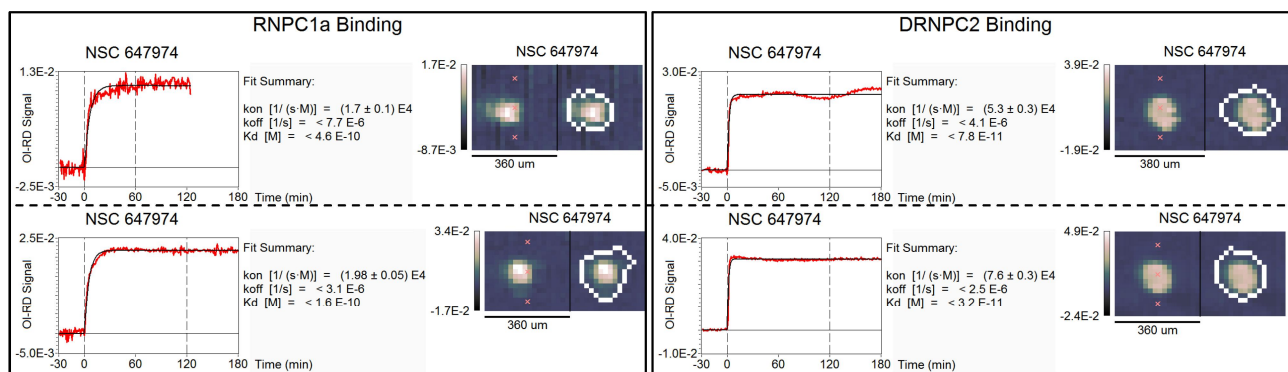

Figure 6. Complete $-\Delta\delta$ signal data for a hit compound identified in a screen against RNPC1a and DRNPC2 proteins. Each reaction was performed on a separate slide with 100 nM protein concentration.
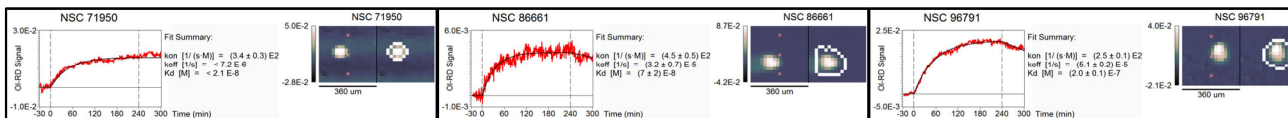

Figure 7. Representative $-\Delta R$ signal data hits for influenza A/PR/8/34 (H1N1) virus particles screened against the library at a concentration of 1,000 PFU/mL.

4. DISCUSSION

In the present experiments, most of the identified hits were for interactions having binding affinities on the order of 1 nM. Such high-affinity interactions are often of primary interest, but weaker interactions on the order 1 μM or even up to 1 mM can also be valuable for fragment-based drug development[27, 28]. In such interactions, the half-life of the bound probes can be on the order of a few minutes. Therefore, during post-reaction washing applied in conventional *ex situ* fluorescence-based endpoint detection of small-molecule microarrays, probe will be lost from low-affinity spots and will not be detected in the fluorescence endpoint scan. However, with our OI-RD-based method, because the microarray is processed *in situ* and detected in real-time *we will detect low-affinity interactions using the real-time mode*. Even if the half-life is comparable to our scanning time resolution (~2.6 ms × # spots in microarray), we will still observe step-like responses at the introduction and removal of probe as long as the association binding signal is above the limit of detection. Naturally, the maximum observable binding affinity depends upon the concentration of the applied probe. We estimate a cut-off affinity as follows: Suppose we *don't* observe binding for a particular spot, but we assume that binding did in fact occur (probe and target were functional and present), but the bound surface mass density was below the limit of detection (LOD). To give an upper-bound, assume the binding reaches equilibrium during the association phase so that $\Gamma_{LOD} > \Gamma_{mono} \cdot C / (C + K_d)$, where Γ_{LOD} is the system limit of detection, Γ_{mono} is the surface mass density of a monolayer of bound probe, C is the probe concentration in solution, and K_d is the equilibrium binding constant of the interaction. Therefore, the maximum directly observable binding affinity is $K_{d,obs} = C \cdot (\Gamma_{mono} / \Gamma_{LOD} - 1)$; if no reaction is observed, we expect that the true $K_d > K_{d,obs}$ because, otherwise, the transient binding response during the association reaction would have been observable above the LOD. In the present experiments $\Gamma_{mono} \sim 1$ ng/mm^2 and $\Gamma_{LOD} \sim 0.1$ ng/mm^2. Therefore, in the present experiments with concentrations ~50-100 nM, we could have detected K_d up to 500-

1000 nM, if such interactions were present. Higher applied concentrations set the cutoff proportionally higher. On the other end of the spectrum, very low values of K_d (high-affinity) are difficult to resolve over short (~1 hour) dissociation observations. We have previously reported a method for estimating an upper-bound for the off-rates in such cases[3], which applied to the present data gives a minimum directly observable K_d ~10-100 pM.

5. CONCLUSION

The results presented here demonstrate the versatility of the combined OI-RD scanner and chemical-microarray platform for high-throughput screening of small-molecule libraries against a variety of protein probes and virus particles with different biological functions and molecular weights. The platform enables *label-free*, *high-throughput*, *real-time*, *automated*, and *cost-effective* screening. We identified specific and reproducible interactions of protein probes with ~1-40 compounds in the 51,200 compound library, demonstrating that the microarray fabrication and processing maintains the selectivity and functionality of the printed compounds. Therefore, the platform is well suited to identify small-molecule ligands in libraries ranging from 10,000 to at least 200,000 compounds, dependent on the desired turn-around time and number of such instruments available for parallel operation. At present, the limiting step is the hit analysis. This step will be sped up by (1) implementing machine-learning algorithms to speed-up and objectify hit determination and (2) improving the flow-cell design to reduce artifacts (such as stress birefringence and air leaks) and therefore reduce the number of false-positives flagged in the first-pass analysis. In addition to qualitatively identifying binding, the real-time data gives a first-pass quantitative measurement of on-rates, off-rates, and equilibrium dissociation constants. These values are useful for designing follow-up experiments such as determining IC50 values for protein-protein inhibition measurements. In fact, for small libraries or a subset of a library of interest, all four flow cells can be loaded with identical microarrays and reacted with the same probe, but each chamber at *different probe concentrations* to obtain high-quality binding affinity measurements for up to ~13,000 interactions in ~14 hours[3]. Furthermore, with appropriate modifications to the scanning systems, OI-RD can also be applied to detect interactions in small microarrays printed in a glass-bottomed microplate format. This will allow us to perform high-throughput follow-up experiments with small-molecule library hits, such as dose-response IC50 measurements.

ACKNOWLEDGMENTS

REFERENCES

[1] X. D. Zhu, J. P. Landry, Y. S. Sun *et al.*, "Oblique-incidence reflectivity difference microscope for label-free high-throughput detection of biochemical reactions in a microarray format," Applied Optics, 46(10), 1890-1895 (2007).

[2] Y. Fei, J. P. Landry, Y. Sun *et al.*, "Screening small-molecule compound microarrays for protein ligands without fluorescence labeling with a high-throughput scanning microscope," Journal of Biomedical Optics, 15(1), 016018 (2010).

[3] J. P. Landry, Y. Fei, and X. Zhu, "Simultaneous Measurement of 10,000 Protein-Ligand Affinity Constants Using Microarray-Based Kinetic Constant Assays," ASSAY and Drug Development Technologies, 10, 250-259 (2012).

[4] J. P. Landry, Y. Y. Fei, and X. D. Zhu, "High Throughput, Label-free Screening Small Molecule Compound Libraries for Protein-Ligands using Combination of Small Molecule Microarrays and a Special Ellipsometry-based Optical Scanner," International Drug Discovery, 6, 8-13 (2012).

[5] J. P. Landry, X. D. Zhu, and J. P. Gregg, "Label-free detection of microarrays of biomolecules by oblique-incidence reflectivity difference microscopy," Optics Letters, 29(6), 581-3 (2004).

[6] A. Wong, and X. D. Zhu, "An optical differential reflectance study of adsorption and desorption of xenon and deuterium on Ni(111)," Applied Physics A (Materials Science Processing), 63(1), 1-8 (1996).

[7] J. P. Landry, Y. S. Sun, X. W. Guo *et al.*, "Protein reactions with surface-bound molecular targets detected by oblique-incidence reflectivity difference microscopes," Appl. Opt., 47(18), 3275-3288 (2008).

[8] J. E. Bradner, O. M. McPherson, and A. N. Koehler, "A method for the covalent capture and screening of diverse small molecules in a microarray format," Nature Protocols, 1(5), 2344-2352 (2006).

[9] J. E. Bradner, O. M. McPherson, R. Mazitschek *et al.*, "A Robust Small-Molecule Microarray Platform for Screening Cell Lysates," Chemistry & Biology, 13(5), 493-504 (2006).

[10] K. Schmitz, S. J. Haggarty, O. M. McPherson *et al.*, "Detecting Binding Interactions Using Microarrays of Natural Product Extracts," Journal of the American Chemical Society, 129(37), 11346-11347 (2007).

[11] O. Vandenabeele-Trambouze, L. Mion, L. Garrelly *et al.*, "Reactivity of organic isocyanates with nucleophilic compounds: amines; alcohols; thiols; oximes; and phenols in dilute organic solutions," Advances in Environmental Research, 6(1), 45-55 (2001).

[12] A. J. Izenman, [Modern multivariate statistical techniques : regression, classification, and manifold learning] Springer, (2008).

[13] S. A. Cunningham, T. M. Tran, M. P. Arrate *et al.*, "Characterization of Vascular Endothelial Cell Growth Factor Interactions with the Kinase Insert Domain-containing Receptor Tyrosine Kinase," Journal of Biological Chemistry, 274(26), 18421-18427 (1999).

[14] N. Ferrara, "Vascular Endothelial Growth Factor: Basic Science and Clinical Progress," Endocr Rev, 25(4), 581-611 (2004).

[15] A. Hoeben, B. Landuyt, M. S. Highley *et al.*, "Vascular Endothelial Growth Factor and Angiogenesis," Pharmacological Reviews, 56(4), 549-580 (2004).

[16] A. K. Olsson, A. Dimberg, J. Kreuger *et al.*, "VEGF receptor signalling -- in control of vascular function," Nature Reviews, 7(5), 359-371 (2006).

[17] I. Sousa Moreira, P. Alexandrino Fernandes, and M. Joao Ramos, "Vascular Endothelial Growth Factor (VEGF) Inhibition - A Critical Review," Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry - Anti-Cancer Agents), 7(2), 223-245 (2007).

[18] R. A. Brekken, J. P. Overholser, V. A. Stastny *et al.*, "Selective Inhibition of Vascular Endothelial Growth Factor (VEGF) Receptor 2 (KDR/Flk-1) Activity by a Monoclonal Anti-VEGF Antibody Blocks Tumor Growth in Mice," Cancer Research, 60(18), 5117-5124 (2000).

[19] Y. A. Muller, Y. Chen, H. W. Christinger *et al.*, "VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 å resolution and mutational analysis of the interface," Structure, 6(9), 1153-1167 (1998).

[20] N. Nishimura, A. Sarkeshik, K. Nito *et al.*, "PYR/PYL/RCAR family members are major in-vivo ABI1 protein phosphatase 2C-interacting proteins in Arabidopsis," The Plant Journal, 61(2), 290-299 (2010).

[21] S.-Y. Park, P. Fung, N. Nishimura *et al.*, "Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins," Science, 324(5930), 1068-1071 (2009).

[22] S. J. Cho, J. Zhang, and X. Chen, "RNPC1 modulates the RNA-binding activity of, and cooperates with, HuR to regulate p21 mRNA stability," Nucleic Acids Research, 38(7), 2256-2267 (2010).

[23] J. Zhang, S.-J. Cho, L. Shu *et al.*, "Translational repression of p53 by RNPC1, a p53 target overexpressed in lymphomas," Genes & Development, 25(14), 1528-1543 (2011).

[24] L. Shu, W. Yan, and X. Chen, "RNPC1, an RNA-binding protein and a target of the p53 family, is required for maintaining the stability of the basal and stress-induced p21 transcript," Genes & Development, 20(21), 2961-2972 (2006).

[25] K. M. Coombs, A. Berard, W. Xu *et al.*, "Quantitative Proteomic Analyses of Influenza Virus-Infected Cultured Human Lung Cells," Journal of Virology, 84(20), 10888-10906 (2010).

[26] A. L. Kroeker, P. Ezzati, A. J. Halayko *et al.*, "Response of Primary Human Airway Epithelial Cells to Influenza Infection: A Quantitative Proteomic Study," Journal of Proteome Research, 11(8), 4132-4146 (2012).

[27] D. E. Scott, A. G. Coyne, S. A. Hudson *et al.*, "Fragment-Based Approaches in Drug Discovery and Chemical Biology," Biochemistry, 51(25), 4990-5003 (2012).

[28] D. A. Erlanson, [Introduction to Fragment-Based Drug Discovery] Springer Berlin Heidelberg, 180 (2012).