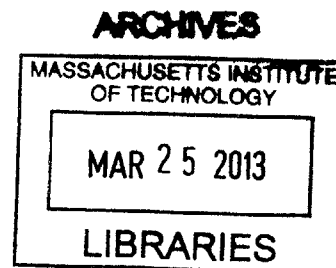


# Label-free Carbon Nanotube Sensors for Glycan and Protein Detection

By

Nigel F. Reuel

B.S. Chemical Engineering  
Brigham Young University (BYU), 2009



Submitted to the Department of Chemical Engineering  
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Chemical Engineering  
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2014

© 2013 Massachusetts Institute of Technology. All rights reserved.

Signature of Author:

~~Signature redacted~~

~~Department of Chemical Engineering~~  
Feb 6, 2014

Certified by:

~~Signature redacted~~

Michael S. Strano  
Professor of Chemical Engineering  
Thesis Supervisor

Accepted by:

~~Signature redacted~~

Patrick S. Doyle  
Professor of Chemical Engineering  
Chairman, Committee for Graduate Students

# Label-free Carbon Nanotube Sensors for Glycan and Protein Detection

By

Nigel F. Reuel

Submitted to the Department of Chemical Engineering  
on February 6, 2014 in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy in Chemical Engineering

## ABSTRACT

Nanoengineered glycan sensors may help realize the long-held goal of accurate and rapid glycoprotein profiling without labeling or glycan liberation steps. Current methods of profiling oligosaccharides displayed on protein surfaces, such as liquid chromatography, mass spectrometry, capillary electrophoresis, and microarray methods, are limited by sample pretreatment and quantitative accuracy. Microarrayed platforms can be improved with methods that better estimate kinetic parameters rather than simply reporting relative binding information. These quantitative glycan sensors are enabled by an emerging class of nanoengineered materials that differ in their mode of signal transduction from traditional methods. Platforms that respond to mass changes include a quartz crystal microbalance and cantilever sensors. Electronic response can be detected from electrochemical, field effect transistor, and pore impedance sensors. Optical methods include fluorescent frontal affinity chromatography, surface plasmon resonance methods, and fluorescent single walled carbon nanotubes (SWNT). Advantages of carbon nanotube sensors include their sensitivity and ability to multiplex. The focus of this work has been to develop carbon nanotube based sensors for glycans and proteins. Before detailing the development of these new sensors, the thesis will begin with a very brief primer on glycobiology, its connection to medicine, and the advantages and limitations of existing tools for glycan analysis. In the second chapter we model the use of quantitative nanosensors in a weak affinity dynamic microarray (WADM) to simulate practical uses of these sensors in bioprocessing and clinical diagnostics.

There is significant interest in developing new detection platforms for characterizing glycosylated proteins, despite the lack of easily synthesized model glycans or high affinity receptors for this analytical problem. In the third chapter we experimentally demonstrate ‘proof of concept’ of carbon nanotube-based glycan sensors. This is done with a sensor array employing recombinant lectins as glycan recognition sites tethered via Histidine tags to Ni<sup>2+</sup> complexes that act as fluorescent quenchers for SWNT embedded in a chitosan hydrogel spot to measure binding kinetics of model glycans. We examine as model glycans both free and streptavidin-tethered biotinylated monosaccharides. Two higher-affined glycan-lectin pairs are explored: fucose (Fuc) to PA-IIL and *N*-acetylglucosamine (GlcNAc) to GafD. The dissociation constants ( $K_D$ ) for these pairs as free glycans (106 and 19  $\mu$ M respectively) and streptavidin-tethered (142 and 50  $\mu$ M respectively) were found. The absolute detection limit for the first-generation platform was found to be 2  $\mu$ g of glycosylated protein or 100 ng of free glycan to 20  $\mu$ g of lectin. Glycan detection (GlcNAc-streptavidin at 10  $\mu$ M) is demonstrated at the single nanotube level as well by monitoring the fluorescence from individual SWNT sensors tethered to GafD lectin. Over a population of 1000 nanotubes, 289 of the SWNT sensors had signals strong enough to yield kinetic information ( $K_D$  of  $250 \pm 10$   $\mu$ M). We are also able to identify the locations of “strong-transducers” on the basis of dissociation constant (4 sensors with  $K_D < 10$   $\mu$ M) or overall signal modulation (8 sensors with > 5% quench response). We report the key finding that the brightest SWNT are not the best transducers of glycan binding. SWNT ranging in intensity between 50 and 75% of the maximum show the greatest response. The ability to pinpoint strong-binding, single sensors is promising to build a

nanoarray of glycan-lectin transducers as a high throughput method to profile glycans without protein labeling or glycan liberation pretreatment steps.

In the fourth chapter we move from detection of model glycoproteins (streptavidin with biotinylated glycans) to a more applied problem: detection of antibodies and their glycosylation. We do this with a second generation array of SWNT nanosensors in an array format. It is widely recognized that an array of addressable sensors can be multiplexed for the label-free detection of a library of analytes. However, such arrays have useful properties that emerge from the ensemble, even when monofunctionalized. As examples, we show that an array of nanosensors can estimate the mean and variance of the observed dissociation constant ( $K_D$ ), using three different examples of binding IgG with Protein-A as the recognition site, including polyclonal human IgG ( $K_D \mu = 19 \mu\text{M}$ ,  $\sigma^2 = 1000 \text{ mM}^2$ ), murine IgG ( $K_D \mu = 4.3 \text{ nM}$ ,  $\sigma^2 = 3 \mu\text{M}^2$ ), and human IgG from CHO cells ( $K_D \mu = 2.5 \text{ nM}$ ,  $\sigma^2 = 0.01 \mu\text{M}^2$ ). Second, we show that an array of nanosensors can uniquely monitor weakly-affined analyte interactions *via* the increased number of observed interactions. One application involves monitoring the metabolically-induced hypermannosylation of human IgG from CHO using PSA-lectin conjugated sensor arrays where temporal glycosylation patterns are measured and compared. Finally, the array of sensors can also spatially map the local production of an analyte from cellular biosynthesis. As an example we rank productivity of IgG-producing HEK colonies cultured directly on the array of nanosensors itself.

One great limitation to these practical applications, common to other new sensor developments, are the constraints of large, bulky, and capital-intensive excitation sources, optics, and detectors. In the fifth chapter we detail the design of a lightweight, field-portable detection platform for SWNT based sensors using stock parts with a total cost below \$3000. The portable detector is demonstrated with antibody detection in our lab and onsite at a commercial facility 3700 miles away with complex production samples.

Along the course of developing these sensors, there was a need to analyze noisy data sets from signal nanotubes (Chapter 3) to determine distinct binding states. NoRSE was developed to analyze high-frequency data sets collected from multi-state, dynamic experiments, such as molecular adsorption and desorption onto carbon nanotubes. As technology improves sampling frequency, these stochastic data sets become increasingly large with faster dynamic events. More efficient algorithms are needed to accurately locate the unique states in each time trace. NoRSE adapts and optimizes a previously published noise reduction algorithm (Chung *et al.*, 1991) and uses a custom peak flagging routine to rapidly identify unique event states. The algorithm is explained using experimental data from our lab and its fitting accuracy and efficiency are then shown with a generalized model of stochastic data sets. The algorithm is compared to another recently published state finding algorithm and is found to be 27 times faster and more accurate over 55% of the generalized experimental space. This work is detailed in Chapter 6.

Future uses of these sensors include *in vivo* reporters of protein biomarkers. In Chapter 7, three-dimensional tracking of single walled carbon nanotubes (SWNT) with an orbital tracking microscope is demonstrated for this purpose. We determine the viscosity regime (above 250 cP) at which the rotational diffusion coefficient can be used for length estimation. We also demonstrate SWNT tracking within live HeLa cells and use these findings to spatially map corral volumes ( $0.27\text{-}1.32 \mu\text{m}^3$ ), determine an active transport velocity (455 nm/s), and calculate local viscosities (54-179 cP) within the cell. With respect to the future use of SWNTs as sensors in living cells, we conclude that the sensor must change the fluorescence signal by at least 4-13% to allow separation of the sensor signal from fluctuations due to rotation of the SWNT when measuring with a time resolution of 32 ms.

In the final chapter we draw conclusions from the development of this carbon nanotube-based sensor for glycan analysis and show the start of future work with arrays of SWNT sensors for glycoprofiling.

**Thesis Supervisor:** Michael S. Strano  
**Title:** Professor of Chemical Engineering

## Acknowledgments

First, I need to acknowledge my 'A' team at home, who has filled my PhD experience with joy; even when I had a tough day of experiments, I knew that they were at home waiting for me. Jenny, my wife and best friend of almost seven years now, has been a constant support and effective feedback loop to inform me when my focus on science may be overshadowing things of greater importance. She deserves many accolades and awards for coming with me on this journey and being willing to live in a shoebox-sized apartment with two kids. I look forward to our many, many adventures to come (hopefully with more living space and a backyard). Asher, my son, thanks for all the daddy de-stressing sessions of LEGO® building and playing outside. Liesel, my daughter, thanks for the huge smiles and always being the first to welcome me home after a day in the lab.

Next, I need to thank my advisor, Professor Michael S. Strano, who is the embodiment of enthused scientific curiosity. You are one of the busiest people I know, yet you take the time to listen to the smallest of ideas from the youngest of your students and give them sincere encouragement and support. Each idea has the potential to be something big, even 'revolutionary' in your eyes. I was one of these young students – full of many, many ideas. You never forced me down a prescribed path but gave me the perfect balance of autonomy and support. Many of my ideas did not work, but many others have blossomed into working, tangible creations. I know that the time I have spent working with you will forever inform the way, pace, and excitement at which I do science. Thank you for everything.

The Strano group is an incredible team. There is no semblance of hierarchy. Each member of this team has a unique skillset, and I have learned from so many. To do this justice, I will go in chronological order with many thanks: Daniel Heller from welcoming me into the group, answering my many questions, and getting me on my first Strano group paper; Jin Ho Ahn and Jong-Ho Kim for the year collaboration we did on protein sensing; Peter Bojo for the academic and start-up work we did together; Paul Barone, Joel Abrahamson, Jingqing Zhang, Andrew Hilmer, Geraldine Paulus, and Ardemis Boghossian for helping me a lot in learning to use almost every tool in the Strano lab; Chih-Jen Shih and Steven Shimizu for getting through core classes together and navigating the start of research; newer students Rishabh Jain, Lee Drahushuk, and Darin Bellisario for many insightful discussions; our recent cohort of post-docs Gili Raviv, Juan Pablo Giraldo, Markita Landry, Nicole Iverson, and Qing Hua Wang for their great advice and friendship. Especially I would like to thank Kevin Tvrdy (now Professor Tvrdy) for his wonderful enthusiasm and for introducing me to the wonderful world of optics; Sebastian Kruss for our many, many great discussions and for putting up with me as a desk-mate; and Justin Nelson without whom our Basel, Switzerland demonstration would have failed and my last few months in the lab would not have been nearly as fun. To all many thanks and I wish you the best in your future endeavors. Stay in touch!

The Strano group is one tiny island in the very collaborative sea we call MIT. There are many I need to thank for their help: to Mark Cutler (AeroAstro) for his sage wisdom in all things electronic and for letting me use the 'macro' tools in his lab (drillpress, CNC router, etc.), Cary Opel (ChemE) for teaching me the trade of mammalian cell culture, Timothy Politano and Adebola (Debo) Ogunniyi (ChemE) for their insight and additional cell lines, Sean Finney-Manchester for his friendship and -80°C freezer that saved us when ours failed, Professor Rajeev Ram and Professor J. Chris Love for their interest and insight in my work, and to all of the incoming class of 2009 for the many years of friendship and for surviving core classes long ago.

My research has also been influenced by a growing network of friends and colleagues outside of MIT: Prof. Lara K. Mahal (NYU) for helping us enter the field of glycobiology; Prof D.C. Lamb (LMU - Munich) and his researchers Aurélie Dupont and Olivier Thouvenin for letting me use their 3D tracking microscope for a summer and helping realize Chapter 7 of my thesis along with many memories of that

beautiful area of Europe; Drs. Bernhard Helk, Ramon Wahl, and Kamal Egodage of Novartis for their multi-year collaboration, monthly meetings, and lasting friendships; Daryl Fernandes (CEO of Ludger) for his friendship since we met at a conference in Vienna (2011) and his insight on the impact and use of my thesis work.

My work has also given me to opportunity to work with and mentor some great visiting students. Thanks to J. Zachary Mundy (NCU) for being a great friend and hard worker. Thanks to Brittany Grassbaugh (UC Fullerton) who was so responsible and autonomous. I could not have gotten as much work done that summer without you. Thanks especially for taking care of our cells when I was out with Liesel's birth. Thanks to Miguel de la Salle Rousseau Twahirwa (Morehouse College) for a great summer of research and putting up with my constant mispronunciation of his last name. All of them have gone on to graduate work of their own and I wish them all the best.

Many thanks go to my thesis committee members, Professor K. Dane Wittrup and Dr. Rahul Raman for finding time in their busy schedules to meet, for all their advice, and for their enthusiasm in this work.

Last I want to thank my extended family for all their love and support. I know that your thoughts and prayers have definitely helped me. I especially want to thank my parents, Lisa and Robert Reuel, for the countless hours they spent reading to me as a kid, remaining patient when I took apart household appliances, devoting many weekends to science competitions, and being a tireless cheering squad for all my scientific endeavors. I understand better now all of the unnoticed work that goes behind the scenes of being a parent. Thank you for all you have done and may you share in this achievement as I would not be here without you.



## Table of Contents

1. Introduction .....	11
1.1 Motivation and Overview.....	11
1.1.2 A brief history of glycobiology and glycoprofiling.....	12
1.1.3 Brief glycobiology primer .....	13
1.1.4 Medicine and glycoprofiling .....	15
1.1.5 Current methods of glycoprofiling .....	16
1.2 Mechanical platforms .....	19
1.2.1 Principles of QCM and measuring strategies .....	19
1.2.2 Chemistry of SAM and comparisons .....	20
1.2.3 Applications of using QCM on SAMs .....	23
1.2.4 Cantilever array sensors detecting carbohydrate-protein interactions.....	35
1.3. Electrical Platforms .....	38
1.3.1 Electrochemical sensors .....	38
1.3.2 Nanotube field effect transistor (NTFET) sensors .....	42
1.3.3 Nanopore sensors .....	44
1.4. Optical platforms.....	46
1.4.1 Fluorescent frontal affinity chromatography for kinetic parameters.....	46
1.4.2 Surface plasmon resonance for kinetic parameters .....	50
1.4.3 Optical carbon nanotube based sensor for measuring glycan-lectin kinetics .....	59
1.5 Summary .....	60
1.6 References .....	62
2. Modeling Native Glycoprofiling with Glycan-CRD Sensors.....	73
2.1 Overview .....	73
2.2 Introduction .....	73
2.3 Experimental Model.....	76

2.4 Results and Discussion.....	79
2.4.1 Screening of Protein Therapeutics – Galactose- $\alpha$ 1,3-Galactose .....	79
2.4.2 Screening Biomarkers – Arthritic Disease and Glycosylation of IgG.....	82
2.4.3 Complete Profiling of Unknown Glycoprotein .....	83
2.4.4 Adding Accessibility to Profiling Algorithm .....	84
2.5 Conclusion.....	85
2.6 References .....	89
Appendix 2.1 .....	91
Appendix 2.2 .....	106
Appendix 2.3 .....	108
Appendix 2.4 .....	109
Appendix 2.5 .....	110
3. Proof of Concept - Nanotube Sensors for Proteins and Glycans.....	111
3.1. Introduction .....	111
3.2 Methods .....	113
3.2.1 Lectin Expression .....	113
3.2.2 Glycan and Model Glycoprotein Probes .....	113
3.2.3 Native PAGE Binding Analysis .....	114
3.2.4 Construction of Chitosan-SWNT Sensor Chips .....	114
3.2.5 Ensemble Measurements of Sensors .....	116
3.2.6 Single SWNT Sensor Measurements .....	118
3.3. Results and Discussion.....	118
3.4. Conclusions .....	129
3.5 References .....	130
4. SWNT Glycan Sensors Applied to Biomanufacturing Antibodies .....	135
4.1 Background and Motivation.....	135
4.2 Methods .....	136



4.2.1 SWNT Sensor and Gel Platform Fabrication .....	136
4.2.2 Poroelastic Relaxation Indentation and Dextran Release Curves.....	137
4.2.3 Data collection on nIR Inverted Microscope.....	139
4.2.4 HEK Cell Line Generation and CHO Origin.....	139
4.2.5 Cell Passaging and Culture.....	140
4.2.6 Hypermannosylation CHO Culture Experiment.....	140
4.3 Results and Discussion.....	140
4.3.1 Sensor Fabrication, Detection Method, and Mechanism.....	140
4.3.2 Modeling a Nanosensor Array to Measure $K_D$ Distributions .....	147
4.3.3 $K_D$ Distributions and Calibration Curves from Experimental Data.....	153
4.4.4 Hypermannosylation Detection – Weakly affined interactions on nanosensor arrays .....	156
4.4.5 Nanosensor Arrays for Monitoring Local Cell Colony Production.....	160
4.4 Conclusion.....	161
4.5 References .....	164
Appendix 4.1 .....	167
Appendix 4.2 .....	170
Appendix 4.3 .....	176
5. Portable Detection of Carbon Nanotube Optical Sensors .....	181
5.1 Background and Motivation.....	181
5.2 Methods.....	181
5.2.1 Fabrication of Gel Modules for Portable Sensing of SWNT Sensors .....	181
5.2.2 Functionalization of Chitosan Wrapped SWNT for IgG Sensing .....	182
5.2.3 Preparation of (GT)15-SWCNT for dopamine detection .....	182
5.3 Results .....	182
5.4 Conclusions .....	194
5.5 References .....	194
6. Sensor Signal State Evaluator Program (NoRSE).....	197

6.1 Summary .....	197
6.2 Motivation .....	197
6.3 Algorithm Explanation .....	197
6.4 Gauging Performance of Algorithm with a Generalized Model .....	200
6.4.1 Scaling comparison of NoRSE and SFA Algorithms .....	203
6.5 References .....	204
Appendix 6.1 .....	204
Appendix 6.2 .....	215
7. 3D Tracking of Carbon Nanotubes within Living Cells.....	237
7.1 Background and Motivation .....	237
7.2 Methods .....	238
7.2.1 Preparation of Chitosan Wrapped SWNT Suspension .....	238
7.2.2 Tracking of SWNT Particles .....	238
7.2.3 Tracking of SWNT Particles in Live HeLa Cells.....	239
7.2.4 AFM Imaging of SWNT .....	240
7.3 Results and Discussion .....	240
7.4 Conclusion.....	253
7.5 References .....	254
Appendix 7.1 .....	256
Appendix 7.2 .....	260
Appendix 7.3 .....	266
Appendix 7.4 .....	269
8. Conclusion and Future Outlook.....	273

## 1. Introduction

Some of the work, text and figures presented in this chapter are reprinted or adapted from reference<sup>1</sup> (reproduced under thesis use allowance from The Royal Society of Chemistry)

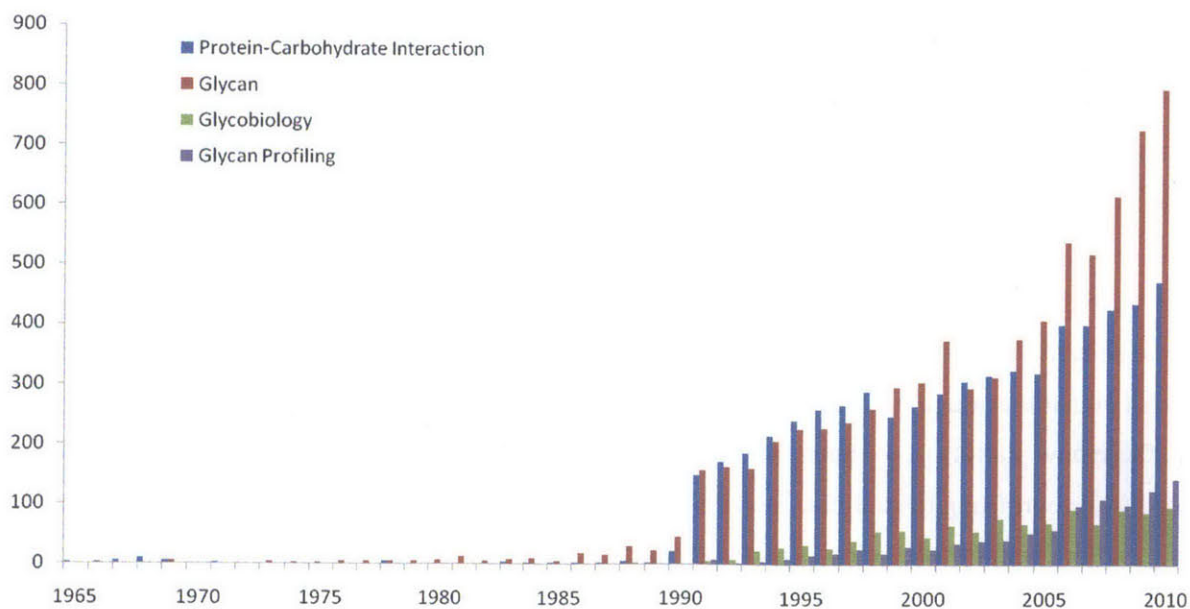
### 1.1 Motivation and Overview

Surface sugars that cover the majority of our proteins and cells, called glycans, are key players in human health and disease.<sup>2</sup> Characterizing their complex structure is a key piece of understanding their structure-function relationship. Current methods of glycan analysis are complex. It was the motivation of this work to develop a simple, label free method of glycan characterization.

This first chapter does not attempt to summarize the vast field of glycobiology nor the current, well-established methods of glycoprofiling, but rather critically reviews a few emerging, nanoengineered, carbohydrate sensors to frame the motivation and need for the subsequent experimental work presented in this thesis. These technologies can form the basis of new analytical tools for rapid glycoprofiling without the traditional steps of glycan liberation or labeling. These sensors can reduce our current macroscopic approaches to mimic the intricate and precise processes of cellular machinery<sup>3</sup> just as nanotechnology has done in other areas like synthetic pores which regulate molecular transport<sup>4</sup>, specific nano-catalysts that mimic selective enzymes<sup>5,6</sup>, and self-assembled materials that resemble cellular vesicles or lipid walls.<sup>7-11</sup> Although routine for nascent proteins in the cell's endoplasmic reticulum<sup>12-14</sup>, the cellular quality-control mechanism of accurate glycan profiling on the surface of proteins<sup>15</sup> has been very difficult to engineer in man-made sensors.<sup>16</sup> Current approaches for high-throughput glycoprofiling require significant pretreatment of the sample, namely chemical/enzymatic liberation of sugar groups from the protein and labeling steps, for accurate detection.<sup>17</sup> However, an emerging class of nanoengineered glycan sensors promise to sense the glycoprotein as the cell would, in its native state. Each of these emerging nanosensors employ some synthetic or naturally-occurring glycan (carbohydrate) recognition domain (CRD) to selectively bind the glycan but they can be classified in three categories according to their method of signal transduction: 1) mechanical, 2) electrical, and 3) optical. However, before we critically review these new platforms, we must briefly frame the history and nomenclature of Glycobiology, the relation of glycoprofiling to medicine, and the current state of glycoprofiling techniques that require labeling and/or chemical liberation steps for accurate results.

### 1.1.2 A brief history of glycobiology and glycoprofiling

The first seeds of discovery in the burgeoning field of Glycobiology were sown over 100 years ago. These pioneering works included the structural analysis of monosaccharides by Hermann Emil Fischer in 1891<sup>18</sup> and the discovery of human blood groups by Karl Landsteiner in 1900<sup>19</sup> (for a more complete list of early Glycobiology pioneers we recommend Table 1.1 of reference 18<sup>20</sup>). But the work of these early scientists lay largely dormant during the explosive growth phase of genetic and proteomic discovery. It was not until the late 1980s that general interest returned to physiologically relevant glycans and the term “Glycobiology” was coined.<sup>21</sup> In the past two decades we have witnessed an exponential growth of discovery and interest in the field of Glycobiology, evidenced by increased publications (Fig. 1.1), the founding of a journal solely dedicated to the field in 1990<sup>22</sup>, and major general scientific awards given to Glycobiologists, like the MIT-Lemelson prize to Prof. Carolyn Bertozzi in 2010<sup>23</sup>. Growing commercial interest in glycoprotein profiling<sup>15</sup> is also evidenced by the increasing number of academic spin-off companies in the past few years (Table 1.1) whose main products are glycan analysis services or tools. The current market emphasis is on services, or contracted profiling of glycosylated proteins, as reliable accuracy and expense have impeded the adoption of off-the-shelf glycoprofiling assays such as those offered by GP Biosciences Ltd or Procognia.



**Fig. 1.1** Growing number of publications in the field of glycobiology. Searches for publications made on Web of Science<sup>SM</sup> for the following subjects: protein-carbohydrate interactions, glycan, glycobiology, and glycan profiling.

**Table 1.1** Growing interest in glycoprofiling evidenced by new companies in the past decade with profiling as their primary product (not an exhaustive list).

Name	Service	Academic Source	Year Founded
Glycan Connections, LLC	On-site carbohydrate analysis for pharmaceuticals, antibodies, and vaccines	Univ NH	2009
ProGlycAn	Service arm of university group offering glycan analysis services	BOKU Vienna	---
GP Biosciences Ltd.	Tools for glycan structure analysis: lectin chip array, evanescent field reader	Japan-AIST	2009
Glycome Technologies	On-site MS analysis and available software packages for MS data interpretation	Univ NH	2009
Ezose	On-site MALDI-MS analysis based on GlycanMap™ technology	Hokkaido Univ	2009
Ludger	On-site glycan analysis for pharmaceuticals and available kits/standards for purchase	Oxford Univ	1999
Procognia	High-throuput lectin array for glycan analysis	Tel Aviv Univ	2000

### 1.1.3 Brief glycobiology primer

The Glycome,<sup>24-27</sup> sometimes referred to as the “third alphabet” in biology<sup>28</sup> (following the language of genetics and proteomics), is incredibly diverse and complex. The library of basic building blocks (analogous to base pairs for genetics or amino acids for proteomics) consist of 9 monosaccharides for mammalian cells (until they are further derivatized by sulfation and other chemical modifications)<sup>20,29</sup> (Fig. 1.2) and many more for bacterial<sup>30</sup>. However the number of branching points available on each carbohydrate, compounded by relevant stereochemistry leads to a huge number of possible linked combinations. For example, a simple glycan consisting of six monosaccharide units from the library of 9 mammalian sugars can theoretically exist in  $\sim 10^{12}$  different branched and linear isomeric forms<sup>29</sup>. Thankfully, nature has settled on a more finite set of structures that are recognized by the glycan-binding proteins and mediate biological functions (referred to as glycan determinants); the current estimate in the number of human glycan determinants is  $\sim 7000$ .<sup>31</sup> Glycans can also differ in how they are bound to proteins: via glycosidic linkages to asparagine residues (N-linked glycans) or to serine and threonine residues (O-linked glycans).<sup>20</sup> To catalogue the vast number of glycan structures, a systematic International Union of Pure and Applied Chemistry (IUPAC) nomenclature for carbohydrates was developed (also available online here <http://www.chem.qmul.ac.uk/iupac/2carb/>) in which numbers are used to note which saccharide carbons are linked and the symbols  $\alpha$  and  $\beta$  are used to specify the linkage

stereochemistry.<sup>32</sup> A widely-accepted figure representation (Fig. 1.3) was initially proposed<sup>33,34</sup> and later standardized<sup>35</sup> to visually depict these long IUPAC named species. Colored or shaded blocks represent the monosaccharides and the drawn orientations of the blocks specify the types of linkages between the sugars.

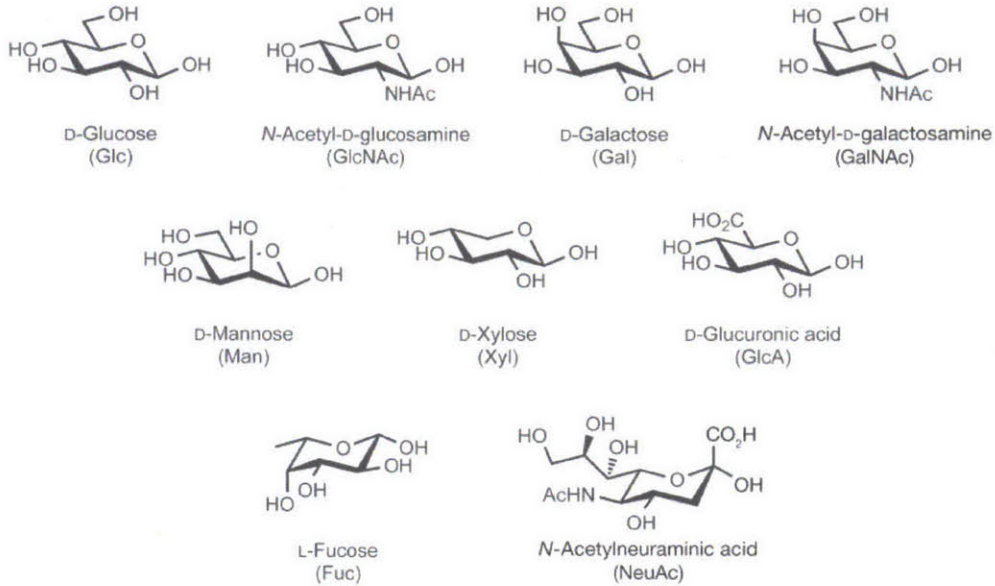


Fig. 1.2 Common monosaccharides found in vertebrates (Reproduced with permission from *Essentials*<sup>20</sup>).

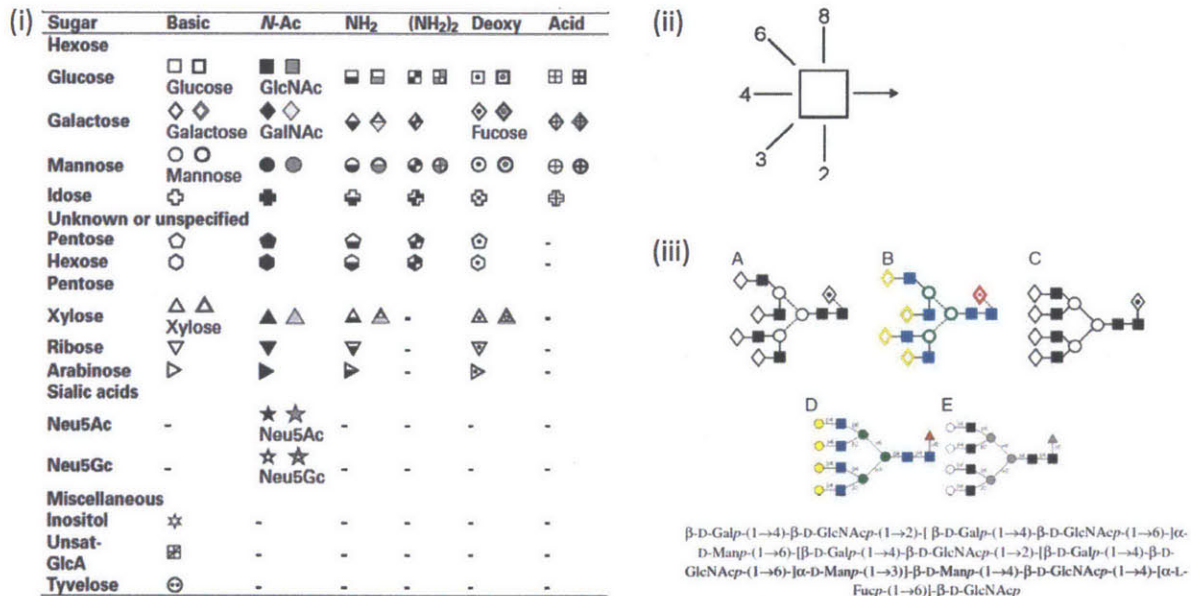


Fig. 1.3 Figure representation of glycan structures. (i) A colored or shaded key is used for the most common monosaccharide units, (ii) carbon linkage positions are denoted by the bond position drawn on each

monosachharide, (iii) examples of colored and shaded glycan figures with corresponding IUPAC nomenclature. Reproduced with permission.<sup>34</sup>

### 1.1.4 Medicine and glycoprofiling

The careful bookkeeping of all bond linkages and stereochemistries is more than an academic injunction; it is vital to track the physiological impact of glycans. Subtle changes in a glycan's structure can have tremendous impact, and thus similar glycans must be carefully distinguished. To quote a recent case,<sup>36</sup> researchers found that when the glycosylated, murine-expressed, monoclonal antibody drug called Cetuximab (used to treat colorectal cancer and squamouscell carcinoma) bears the wrong disaccharide, Galactose- $\alpha$ 1,3-Galactose, rather than one of the other 21 similar forms,<sup>37</sup> the drug becomes immunogenic to one third of its intended population. Thus production of this drug requires stringent glycoprofiling to eliminate batches that contain this deleterious disaccharide.

Glycoprofiling is a crucial step in all glycosylated protein production, especially when the expressed product is intended as a drug<sup>38</sup>. Such proteins are efficiently expressed in non-native cell lines that lack the inherent glycosylation regulation machinery. The expressed proteins emerge as a heterogeneous mixture of glycoforms.<sup>39</sup> In order for a glycosylated protein drug to gain clinical approval, the range of produced glycoforms has to be carefully established, tested, and approved.<sup>40</sup> If changes in the production of the drug alter the glycoform population in any way, for example through environmental changes or in process scale-up, the drug is considered as a new entity and must be reapproved.<sup>38</sup> During ongoing production, each batch of proteins must be carefully profiled to ensure that the approved glycosylation pattern is present. There is a huge research interest in engineering cell lines to produce specific homogenous glycoproteins<sup>41-43</sup>; the impact of such work is witnessed by the competitive acquisition of GlycoFi, which successfully produced homogenous glycoproteins from yeast cells, by Merck for \$400 million in 2006.<sup>44</sup> Yet, the growing market of glycosylated therapeutics, now at over two thirds of the biologics market, is still dominated by traditional expression mechanisms.<sup>40</sup> Thus high-throughput platforms that can easily integrate directly into the production line and screen native proteins would be of great academic and industrial worth.

Another direct medicinal application of glycoprofiling is in the screening of known biomarkers and the discovery of new ones. Arthritis is an example of a human disease with known glycosylated biomarkers. Arthritis has been linked with changes to the single glycosylation site in the constant domain of the IgG Fc region.<sup>45</sup> It is thought that by accurate screening of the glycan group on these proteins, clinicians would be able to tell what type of arthritis an individual has. Healthy individuals have a large glycan group in this region, yet as the arthritis severity increases, this glycan group loses many of its

terminal sugars.<sup>45</sup> When an increased count of the IgG proteins contain only the base glycan structure ((GlcNAc2Man6)-Man4GlcNAc-4GlcNAc also known as IgG-G0), the individual is likely suffering from Rheumatoid Arthritis (RA).<sup>46</sup> In a clinical setting it would be advantageous to create a rapid screening tool for this prescient glycan group from all the other possible IgG variants. There are many other glycosylated protein biomarkers known for cancers and carcinomas.<sup>47-50</sup> An accurate glycoprofiling tool could also be used for discovery of new disease biomarkers<sup>51</sup> by screening serum samples from infected populations.

Lastly, there are many other fundamental areas of clinical science that glycoprofiling can help address, as glycans compose the fourth most abundant ingredient of all cellular structures (after proteins, DNA, and lipids)<sup>52</sup> and glycosylated proteins compose ~50% and >90% of the cell and secreted proteome, respectively<sup>53,54</sup>. Glycans help dictate protein folding, function, and transport.<sup>55-57</sup> As such they are integral in homeostatic processes, and likewise play a role in many human diseases. These are too numerous to address in this short review article, but we refer the reader to a number of more exhaustive reviews for glycan connections to cancer<sup>58-60</sup>, chronic inflammation<sup>61</sup>, immunology<sup>62</sup>, immunoglobulin function<sup>50</sup>, and other human diseases.<sup>63,64</sup> Profiling tools aid basic research in determining the structure of physiologically relevant glycans. Additionally, as the structural roles of glycans are further understood, better engineering of emerging glyconanomaterials for medicinal applications will be possible.<sup>65</sup>

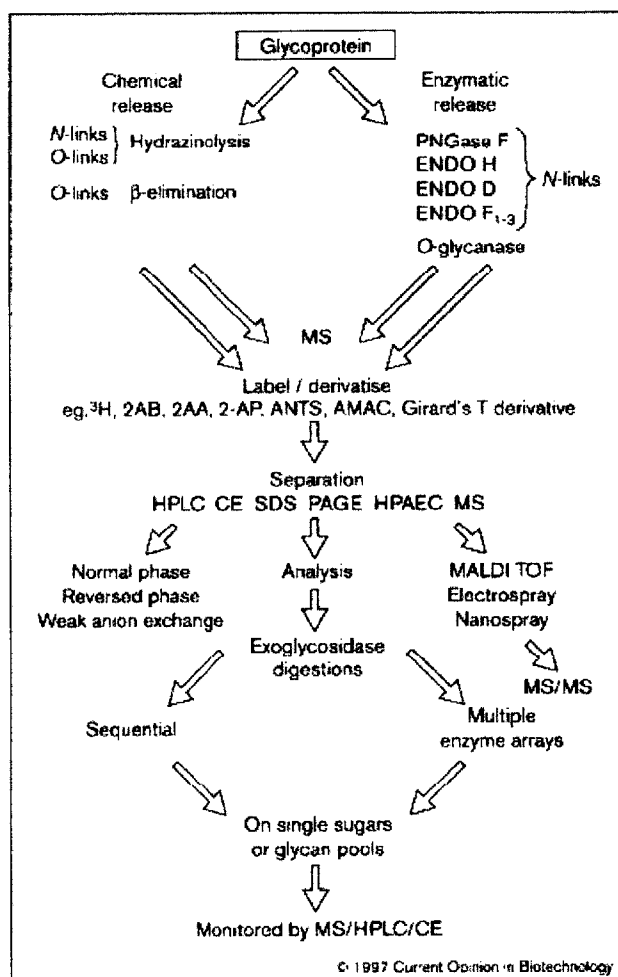
### 1.1.5 Current methods of glycoprofiling

For a more detailed account of current methods of glycoprofiling, including the established “acronym tools” (mass spectrometry (MS), capillary electrophoresis (CE), liquid chromatography (LC)) and more recent microarray methods, we recommend these recent reviews.<sup>17,66-69</sup> The common acronym tools, listed above, have spawned an even greater list of specialized tools also known by acronyms; see table 47.1 of the *Essentials of Glycobiology* text<sup>20</sup> for a concise guide. Herein we wish to briefly review these current methods to illustrate their capacities and limitations to frame the need for emerging nanoengineered sensors for medical applications. A good summary of their advantages and limitations is provided in the Vanderschaeghe *et al.* review<sup>17</sup>.

The Rudd *et al.* review<sup>67</sup> provides a clear roadmap of how the “acronym” technologies are often linked together for glycan analysis (reproduced in Fig. 1.4). The glycoprotein solution, which consists of a heterogeneous mixture of glycoforms<sup>39</sup>, is pretreated to release the glycans from the surface of the protein referred as the liberation step. Liberation is done by either chemical (hydrazinolysis) or enzymatic release (general peptide N-glycanases for N-glycans and more specialized enzymes for O-glycans)<sup>39</sup>. This liberation step must be done carefully to ensure that the glycans are neither destroyed nor altered so



the subsequent analysis is an accurate representation of the protein's glycan profile. Liberation of the glycans is necessary for the subsequent profiling tools as the protein's presence will confound the separation technique or influence the spectrometric analysis.



**Fig. 1.4** Common coupling strategies of macroscopic profiling tools. MS is the work horse of the systems, while HPLC, CE, and SDS PAGE provide atomistic details. Reproduced with permission.<sup>67</sup>

The choice of tool after liberation largely depends on the quantity of the glycan sample and the prior knowledge of the glycoform population. If a large amount of sample is available (>5mg) and standards are available for known species in the population, than simple liquid chromatographic techniques<sup>70,71</sup> can be used. The column will separate the glycans, and the elution times are compared to standards. However, identical elution times do not guarantee the glycan fraction is identical to the standard, so LC methods are often coupled with atomistic analyses like mass spectrometry.<sup>72,73</sup> LC methods do not necessarily require labeling of the glycan species, but this is often done to enhance separation.

Mass spectrometry (MS) is the old work horse of these coupled systems, giving a clear, atomistic structure of the glycan.<sup>74</sup> It was first used in 1978 to analyze the surface carbohydrates of Antarctic fish “antifreeze” proteins<sup>75</sup> and has subsequently been used for many structural determinations.<sup>76</sup> For a clearer glycan structural analysis, the sample must be relatively pure, thus it is coupled with separation tools, like LC mentioned above. Electrophoretic separation, such as capillary electrophoresis (CE), can also be coupled with MS for glycan analysis<sup>77,78</sup> but is more difficult to interface than LC methods and is less frequently used. Like LC tools, CE can be used without MS coupling for glycan analysis, as in this study of screening milk oligosaccharides<sup>79</sup>, but it requires labeling<sup>67,80</sup> and yields indirect structural information (comparison to standards). Mass spectrometry is enhanced by molecule ionization techniques, the most common being matrix-assisted laser desorption ionization (MALDI) yielding “mass-maps” of the glycan constituents. The use of MALDI-MS systems to profile glycans has been rigorously chronicled in the reviews<sup>81-89</sup> of D.J. Harvey (the developer of the widely-used figure nomenclature of glycans). Recent improvements in MS techniques for glycan analysis include better data analysis methods<sup>90</sup> and reduction of sample size with microfluidic platforms<sup>91</sup>. NMR can also probe the atomistic structure of glycans, but it requires specific expertise to analyze the complex signal. Thus, these analyses are often reserved for determining the finer details of glycan structures like the anomericity of the glycosidic bonds<sup>20</sup> but can also be used in profiling, as the A. E. Manzi *et al.* study does with the glycans of genetically modified mice.<sup>92</sup> The interplay of all these systems is demonstrated well in the D. Muller *et al.* work,<sup>93</sup> wherein they use MALDI-MS, HPLC, and NMR to glycoprofile a population of hybrid plasminogen activators.

These “acronym” tools have been strongly established over the past 30 years, but they still remain expensive and time-consuming.<sup>17</sup> They give clear, detailed images of the glycan, including composition and linkage structure, but at the expense of destroying the protein and glycan. For certain clinical applications, such as serum screening, the liberation requirement may be suitable, but the cost may be prohibitive. For large scale, continuous glycoprotein profiling, such as in screening the expressed production of a protein therapeutic, the liberation steps and time-consuming analysis would be unsuitable. Thus, there has been an interest in platforms that are more rapid and can do profiling without labor intensive glycan liberation.

In the past decade microarray platforms for glycoprofiling have emerged to fulfill the need for rapid glycoprofiling.<sup>66</sup> They do not require glycan liberation from the protein surface, but they do require labeling steps for detection.<sup>94</sup> The premise of these platforms is to create an array of carbohydrate recognition domains (CRDs) that interact with varying degrees of affinity to glycans. If enough CRDs are present, a fingerprint for each glycan can be determined; this effectively mimics nature’s use of glycans as a “sugar code.”<sup>28,95-97</sup> There is a distinct emphasis in the field on using naturally occurring

proteins with CRDs, known as lectins,<sup>98</sup> as the arrayed sensors. These lectin-based platforms have been used for glycoprotein profiling<sup>99-101</sup>, biomarker detection<sup>102</sup>, and analysis of cell-surface carbohydrates.<sup>103,104</sup> In these platforms the labeling can occur as a pretreatment step (labeling the glycoprotein) or as a sandwich assay<sup>105</sup> (adding a fluorescent molecule that absorbs selectively to the glycoproteins bound to CRDs on the surface). CRDs have also been formed from a new class of carbohydrate binding peptides,<sup>106</sup> random peptide sequences<sup>107</sup>, and other carbohydrate antibodies.<sup>108</sup> These array platforms are currently limited to rough, high throughput analyses of glycoproteins. Their level of detail is limited to the number of CRDs present on the array and their ability to accurately define the relative binding affinities of the glycoproteins to the CRDs.<sup>17,95,98</sup> Advances have been made in the basic platform to boost the fluorescent signal of the bound species, such as using evanescent field illumination<sup>109</sup> or optimized labeling and sensor positioning<sup>110</sup>, but there is still a need for better resolution. The ideal platform would be able to clearly distinguish between the binding strengths of glycans to CRDs without the need of labeling the glycoprotein, just as the ER is able to screen nascent proteins without any pretreatment steps.

## 1.2 Mechanical platforms

Mechanical transduction of single molecules has been done with a myriad of micro and nano-platforms. Extension of these platforms to the detection of glycan-CRD interactions has occurred in the past two decades, primarily using a quartz crystal microbalance (QCM) platform and more recently with fine-tuned cantilever arm detectors. The self-assembled monolayer (SAM) has recently appeared to be the most promising approach to create a scaffold for either lectins or sugars, in conjunction with the use of classical QCM. In this portion of the review, we will focus our discussion on the application of QCM and cantilever sensors to studying the binding between lectins and sugars, as well as the more recent approaches of creating SAM layers.

### 1.2.1 Principles of QCM and measuring strategies

The QCM technique relies on the change in the resonant frequency of a quartz crystal associated with a mass change of the crystal, following the equation given by Sauerbrey in 1959<sup>111</sup>.

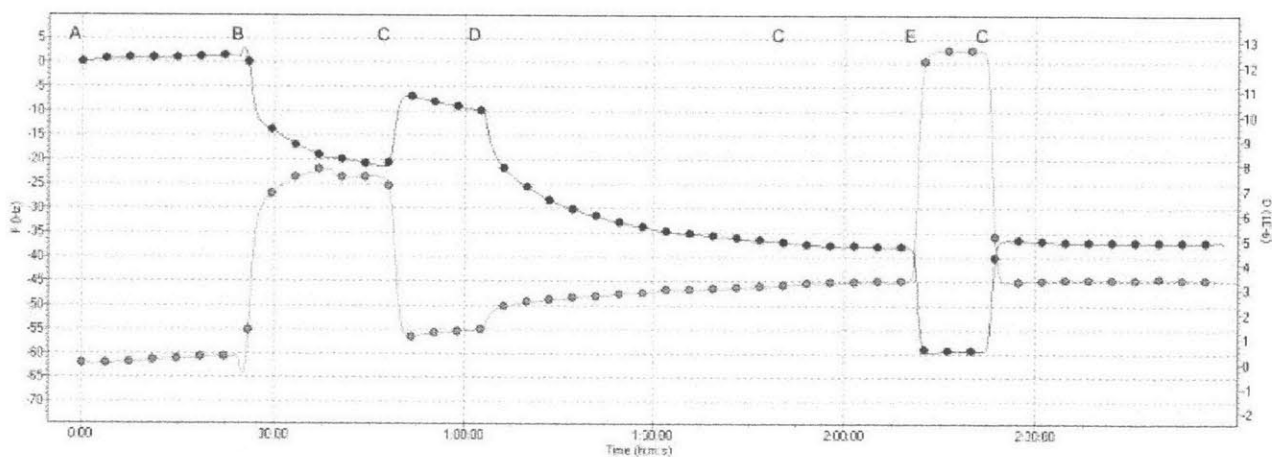
$$\Delta f = -2\Delta m n f_0^2 / [A(\mu_q \rho_q)^{1/2}] \tag{Eq. 1.1}$$

where  $n$  is the overtone number,  $\mu_q$  is the shear modulus of the quartz, and  $\rho_q$  is the density of the quartz. The equation is obtained by solving a one-dimensional equation of motion, with an important assumption:

the frequency change of a resonator is due to the deposition of thin, rigid, and uniform films, where the foreign mass is strongly coupled to the resonator.<sup>112</sup>

QCM can be used to track both the thermodynamics and the kinetics of the binding between two molecules. Both damping resistance and frequency modulation can be monitored (a typical graph inserted below, Fig. 1.5). Small damping resistance changes indicate rigid and strong attachment, whereas large changes indicate viscoelastic behavior of the adsorption. Specific interaction between the molecule and the surface will cause a mass increase of the crystal, resulting in a frequency decrease.

Typical QCM operates in such a way that the QCM quartz surface is first modified to strongly adsorb or covalently attach a receptor molecule; frequency modulation and dissipation functions are then monitored in real-time as the ligand molecule flows over the quartz surface. The example measurements (Fig. 1.5) reveal typical output of a QCM experiment where adsorption of the molecule results in a decrease in frequency and an increase in dissipation.



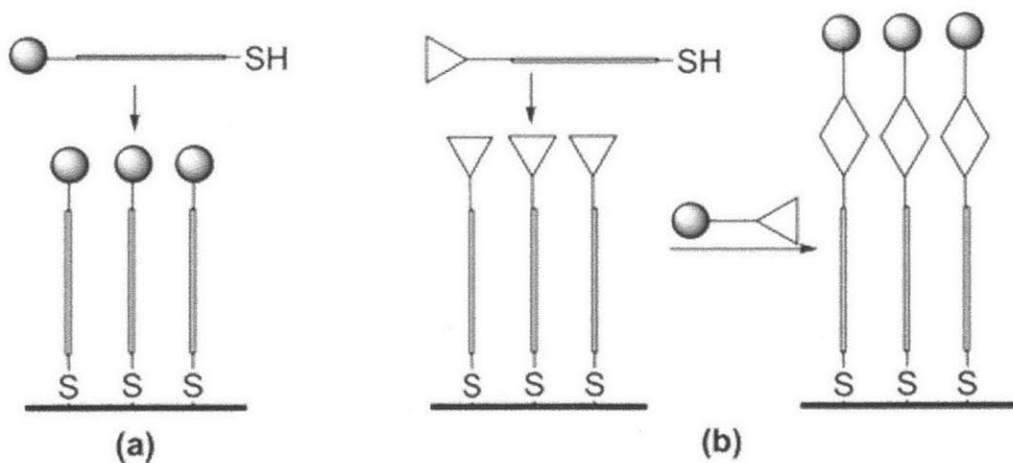
**Fig. 1.5** The time-dependent frequency (upper curve) and dissipation shift (lower curve) recorded during the immobilization of Concanavalin A lectin on the surface of quartz crystal electrode: (A) thiol-modified electrode; (B) activation of the carboxyl groups with NHS:EDC mixture; (C) rinse with the running buffer; (D) immobilization of the lectin ( $1\text{ mg mL}^{-1}$ , 60 min); (E) blocking of unspecific sites with ethanolamine-HCl (pH 8.5). PBS (10mM, pH 7.4) containing 1mM metal ions ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  and  $\text{Mn}^{2+}$ ) was pumped through the system at a flow rate  $50\ \mu\text{l min}^{-1}$ . Reproduced with permission.<sup>113</sup>

## 1.2.2 Chemistry of SAM and comparisons

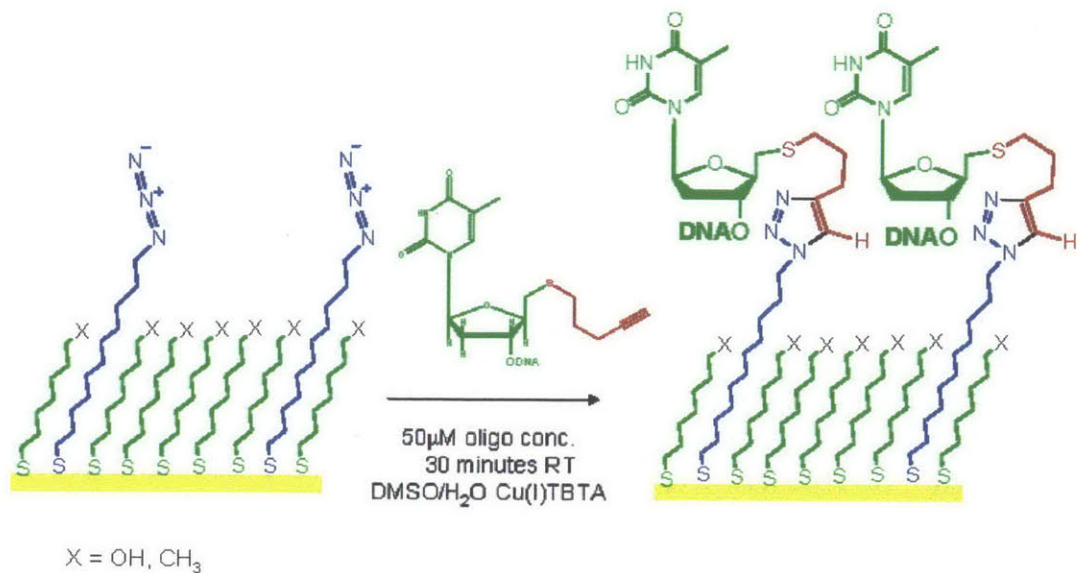
SAMs are used to immobilize lectins, sugars, or glycoproteins onto the surface of the QCM gold substrate. A properly made SAM should have two important aspects: 1) the first layer of molecules (either lectins or sugars) should bind to the surface strongly (mostly through covalent interaction) such that the introduction of the counter-molecules will not easily remove them; 2) non-specific binding should largely be eliminated. Among the many types of SAMs, alkanethiolates on gold currently constitute the best model system.<sup>114,115</sup> Modification of the gold surface with sugars or lectins is easily performed with thiol-gold chemistry.

### 1.2.2.1 Construction of the sugar-SAM

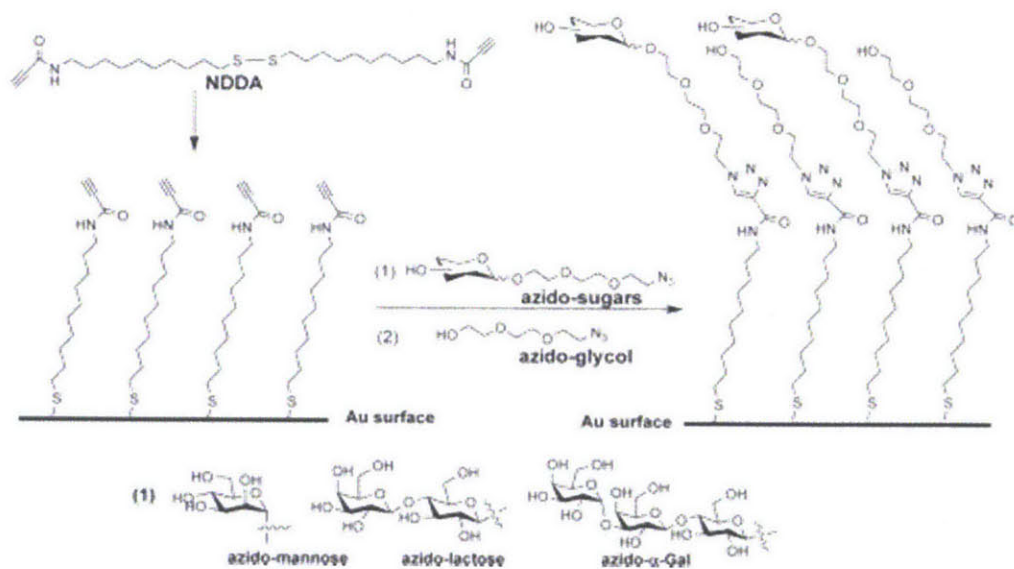
In order to create sugar SAM layers, two general strategies have been used, as shown in Figure 1.6. One approach is to synthesize sugars with an alkanethiol group attached (Fig. 1.6a), although the synthetic effort of this approach is not trivial. The second approach is to first modify the gold surface by introducing a SAM layer with specific functional groups, and then introduce the sugars through reactions with these functional groups (Fig. 1.6b); ‘click chemistries’ play a major role in this latter scheme. Collman and co-workers made azide-functionalized SAMs and then attached the alkyne-modified ligodeoxyribonucleotide through ‘click chemistry;’ the resultant layer is well-defined (Fig. 1.7). Following this strategy, Wang *et al.*<sup>116</sup> created a sugar SAM through introduction of an alkyne-terminated alkanethiol linker to the gold surface on which the azido-sugars could be ‘clicked’ (Fig. 1.8).



**Fig. 1.6** Carbohydrate SAM fabrication strategies on the gold substrate: (a) direct coupling through thiol anchor; (b) indirect coupling through click reactions on preformed template. Reproduced with permission.<sup>116</sup>



**Fig. 1.7** Surface modification of a mixed SAM on gold by chemoselective coupling of acetylene-bearing oligonucleotides.<sup>117</sup>



**Fig. 1.8** Carbohydrate (i.e., mannose, lactose, or R-Gal) SAM fabricating strategies using a Huisgen 1,3-dipolar cycloaddition “click” chemistry. Reproduced with permission.<sup>116</sup>

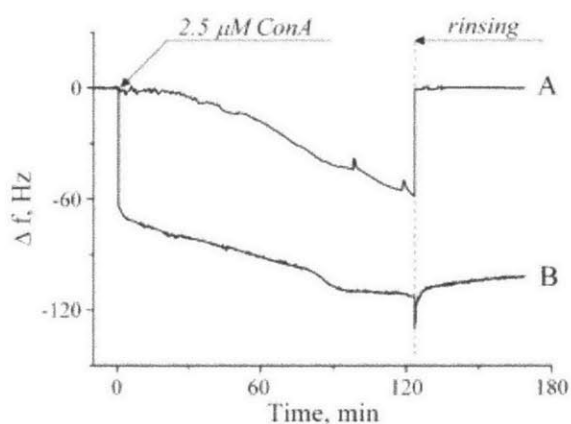
### 1.2.2.2 Construction of the lectin-SAM layer

Various methods of making lectin-SAM layers have been studied. In Faria’s work<sup>118</sup>, cystamine was first introduced to the gold surface and followed by glutaraldehyde molecules; the two candidate lectins were

introduced separately (Con A and Jacalin). The gold surface was then exposed to glycine solution to block the remaining aldehyde groups. Other studies have fabricated lectin-SAM surfaces using a similar strategy.<sup>119</sup>

In the work by Bueno, *et al.*,<sup>120</sup> EDC/NHS reaction was performed to covalently attach the lectins to the surface of the gold. In this work, the gold quartz crystal was incubated first in a solution of  $1 \times 10^{-4}$  Molar 11-mercaptoundecanoic acid (MUA) and 10 Molar 2-mercaptoethanol in ethanol. After the acid was activated by the EDC/NHS reagent, the gold crystal was exposed to the lectins, to form a lectin-SAM. This platform was used to compare the activity of the native and recombinant forms of Artin M protein to horseradish peroxidase glycoprotein (HRP).

Covalently attaching proteins to the gold surface is necessary for most QCM studies, as a simple adsorption-type SAM layer will not be able to withstand the many rinsing steps that follow. Lekka and co-workers<sup>119</sup> compared two types of immobilization methods. Direct adsorption of ConA on the gold surface resulted in the removal of the SAM layer in a later rinsing step, whereas covalent binding of Con A (using thiol chemistry) resulted in a more stable SAM layer, even after a rinsing step identical to that which removed the non-specifically adsorbed Con A (Fig. 1.9)



**Fig. 1.9** The time-dependent frequency shift recorded during the adsorption of  $2.5 \mu\text{M}$  Con A onto the surface of quartz crystal electrodes: pure gold (A) and thiol-glutaraldehyde-modified electrodes (B). Arrows indicate the moments when Con A was added and when electrodes were rinsed with pure buffer. Reproduced with permission.<sup>119</sup>

### 1.2.3 Applications of using QCM on SAMs

Frequency modulation of QCM is typically more closely monitored than damping resistance to understand the binding events. For a first-order Langmuir isotherm model, the frequency change can be described as follows.

$$\frac{\Delta f}{\Delta f_{\max}} = \frac{K_A C}{K_A C + 1} \quad (\text{Eq. 1.2})$$

$$\text{Or } \frac{C}{\Delta f} = \frac{C}{\Delta f_{\max}} + \frac{1}{\Delta f_{\max}} \frac{1}{K_A} \quad (\text{Eq. 1.3})$$

where  $K_A$  is the binding constant, and  $C$  is the concentration of the counter-molecule. Fitting  $\frac{C}{\Delta f}$  as a function of  $C$  allows for calculation of  $K_A$ .

With regard to the study of kinetics, the most common approach is to assume the binding of the analyte to the SAM layer is the first-order and described by two rate constants related to association ( $k_1$ ) and dissociation ( $k_{-1}$ ) of the molecules with the surface. The relaxation time constant,  $\tau$  is defined as,

$$\tau^{-1} = k_1[\text{concentration}] + k_{-1} \quad (\text{Eq. 1.4})$$

The percentage of the ligand bound to the receptor is defined as  $\theta_t$  and the equilibrium percentage as  $\theta_{\max}$ . At any given time,  $t$ ,

$$\theta_t = \theta_{\max} (1 - e^{-t/\tau}) \quad (\text{Eq. 1.5})$$

If one assumes the mass change of the crystal corresponds to the formation of the ligand-receptor complex and that the frequency change is proportional to the mass change, the following equation also holds:

$$\Delta f_t = \Delta f_{\max} (1 - e^{-t/\tau}) \quad (\text{Eq. 1.6})$$

Therefore, by monitoring the frequency modulation as a function of time and concentration,  $k_1$ ,  $k_{-1}$ , and  $K_A$  can be readily evaluated, elucidating the interaction between the lectins and the carbohydrates or



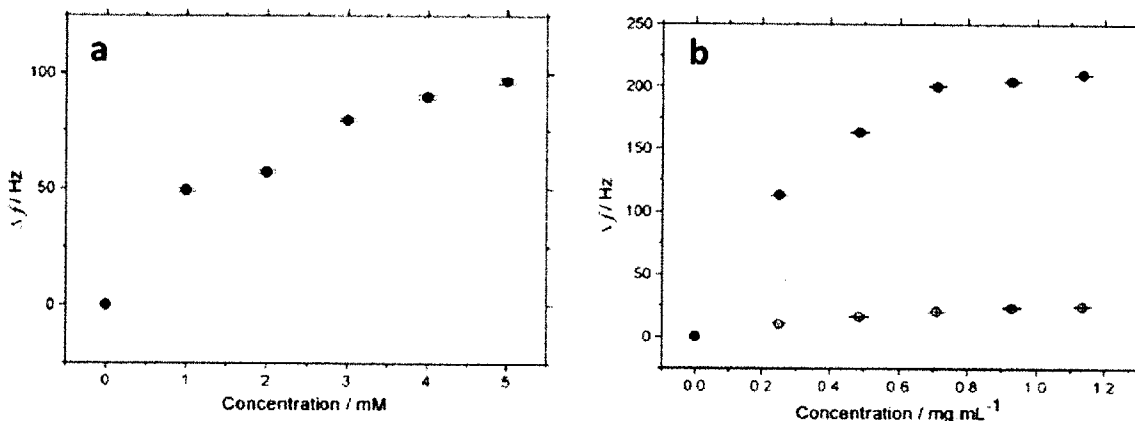
glycoproteins. In the following section, we will summarize a few examples on how to apply QCM to study this type of interaction. Reproduced with permission.<sup>119</sup>

### **1.2.3.1 Lectin detection using sugar-SAM**

In Wang and co-workers' study<sup>116</sup>, through click chemistry, three sugar-SAMs were constructed, including mannose-SAM, lactose-SAM, and galactose. The mannose-SAM was found to have a high specificity for ConA, but also had slight cross-activity with ECL, which binds to galactose and galactosyl ( $\beta$ -1,4) *N*-acetylglucosamine. The value of  $K_A$  for ConA binding with mannose was determined to be  $(8.7 \pm 2.8) \times 10^5 \text{ M}^{-1}$  by QCM technique and  $(3.9 \pm 0.3) \times 10^6 \text{ M}^{-1}$  by SPR; and  $(4.6 \pm 2.4) \times 10^6$  for lactose binding with ECL and  $(6.7 \pm 3.3) \times 10^6 \text{ M}^{-1}$  for  $\alpha$ -Gal binding with anti-Gal by QCM. The deviation between QCM and SPR results were explained by avidity effects, in which the gold chip used for SPR may contain more binding sites than the gold chip made for QCM. In this study, as well as others included in this review, multivalency has significant pros and cons. On the one hand, it can increase the apparent binding affinity (as in the case of glycosylated nanoparticles reviewed in Section 1.2.3.4). On the other hand, multivalency also results in a non-intrinsic binding constant, which may bias our understanding of the strength and kinetics of the binding between certain glycans and lectins, as given in the example above.

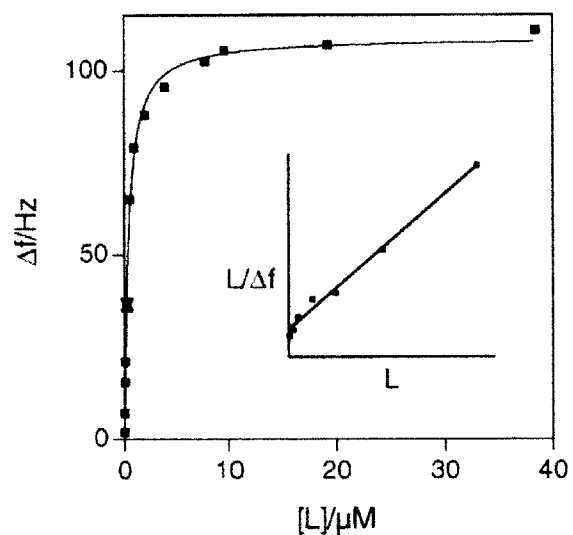
### **1.2.3.2 Glycan detection using lectin-SAM**

In the work by N. C. Pesquero *et al.*, a lectin-SAM layer was constructed for both Con A and Jacalin. Different concentrations of maltose and Fetuin were introduced to Con A- and Jacalin-coated electrodes, respectively, in a QCM setup, and equilibrium binding curves were measured (Fig. 1.10). Binding constants for Jacalin binding to Fetuin ( $(6.4 \pm 0.2) \times 10^4 \text{ M}^{-1}$ ) and Con A to maltose ( $(4.5 \pm 0.1) \times 10^2 \text{ M}^{-1}$ ) were found.

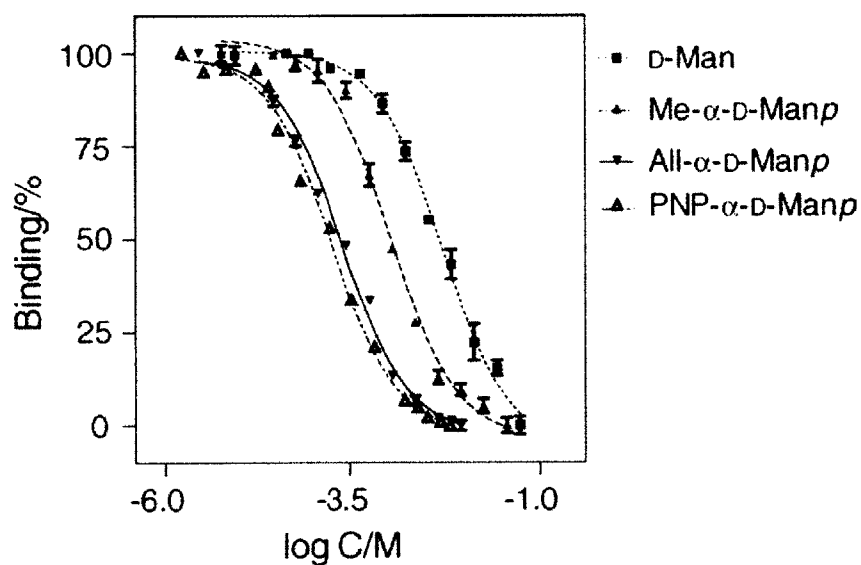


**Fig. 1.10** a) QCM frequency change monitored during the addition of different concentrations of maltose to a Con A-SAM. b) QCM frequency change monitored during the addition of different concentrations of fetuin (close circle) and BSA (open circle) to a Jacalin-SAM. Reproduced with permission.<sup>120</sup>

Rather than immobilizing Con A on the quartz crystal electrode, Ramström *et al.*<sup>121</sup> instead coated gold-plated electrodes with polystyrene followed by yeast mannan, a mannose-rich polysaccharide. Bovine serum albumin, a non-specific protein, was used to cover the non-polar regions of the surface to which the mannan didn't adsorb. Con A was then flowed over the surface at varying concentrations (Fig. 1.11) and a saturation curve was obtained. Using a Langmuir isotherm model, the dissociative constant for the studied mannan-Con A binding interaction was  $K_D = 0.4 \mu\text{M}$ , corresponding to an affinity constant  $K_a = 2.5 \times 10^6 \text{ M}^{-1}$ . After Con A adsorption, a series of competition experiments were conducted using four known Con A carbohydrate ligands: D-mannose, methyl- $\alpha$ -D-mannopyranoside, allyl- $\alpha$ -D-mannopyranoside, and 4-nitrophenyl- $\alpha$ -D-mannopyranoside. Increasing the concentration of the ligands led to a decrease in the oscillation frequency shift ( $\Delta f$ ), indicating a loss of surface-bound Con A. The competition plots (Fig. 1.12) were analyzed with a non-linear regression to obtain the  $EC_{50}$  values listed in Table 1.2.  $EC_{50}$  is defined as *half maximal effective concentration*, denoting to the concentration of the ligand that is needed to introduce a response halfway between the baseline and maximum. The ability of the QCM to assess the binding characteristics of glycans to lectins is limited by the mass of the analyte; if it is too small, a frequency shift will not be observed. By attaching the glycan, rather than the lectin, to the QCM surface, Ramström *et al.*<sup>121</sup> demonstrated a system that can potentially be modified to study monosaccharides or small glycans the QCM is unable to detect when they are introduced as analytes.

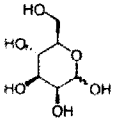
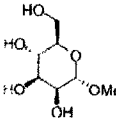
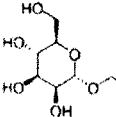
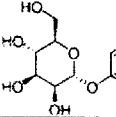


**Fig. 1.11** QCM frequency variation with increasing concentrations of the lectin, Con A. The inset represents the linearized data. Reproduced with permission.<sup>121</sup>



**Fig. 1.12** Competition plots for Con A binding, D-Man: D-mannose, Me- $\alpha$ -D-Manp: methyl- $\alpha$ -D-mannopyranoside, All- $\alpha$ -D-Manp: allyl- $\alpha$ -D-mannopyranoside, and PNP- $\alpha$ -D-Manp: 4-nitrophenyl- $\alpha$ -D-mannopyranoside. Reproduced with permission.<sup>121</sup>

**Table 1.2** Comparison of  $EC_{50}$ -values for the QCM lectin biosensor and reported values from an enzyme-linked lectin assay (ELLA). Reproduced with permission.<sup>121</sup>

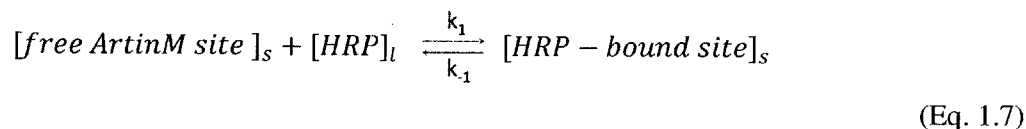
Ligand		$EC_{50}$ (mM) QCM (95% CI)	$EC_{50}$ (mM) (ELLA)
D-Mannose ( $\alpha/\beta$ ) <sup>a</sup>		5.3 (4.4–6.5)	>2.5
Methyl- $\alpha$ -D-mannopyranoside		1.1 (0.90–1.3)	0.92
Allyl- $\alpha$ -D-mannopyranoside		0.25 (0.21–0.29)	0.26
4-Nitrophenyl- $\alpha$ -D-mannopyranoside		0.18 (0.15–0.21)	0.11

Figures within brackets represent 95% confidence intervals (95% CI).

<sup>a</sup> 67%  $\alpha$ -, and 33%  $\beta$ -D-mannose according to <sup>1</sup>H NMR-analysis in D<sub>3</sub>PO<sub>4</sub>-NaOD buffered saline, pH 7.4.

### 1.2.3.3 Glycoprotein activity study using lectin-SAM

QCM has become a very useful tool to study glycoprotein and lectin interactions without labeling. In the work by Bueno, *et al.*,<sup>120</sup> comparative studies were performed on the binding activity of native and recombinant forms of Artin M protein to horseradish peroxidase glycoprotein (HRP). HRP is N-glycosylated protein that contains the trimannoside Man $\alpha$ 1–3[Man $\alpha$ 1–6]Man, which is a known ligand for jArtinM (a D-mannose binding lectin). In this work, QCM measurements were performed to assess HRP binding at different concentrations, monitoring both the kinetics of the binding and also the equilibrium constants. In the kinetics study, the following equations were assumed



$$r_b = k_1 \Gamma_f [HRP] \quad (\text{Eq. 1.8})$$

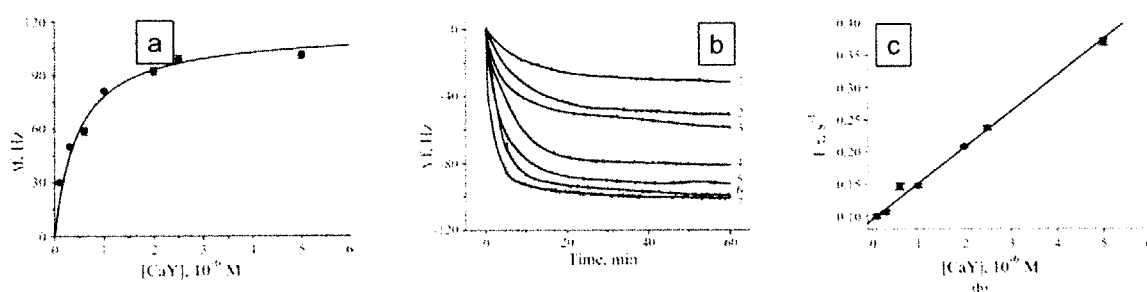
$$r_d = k_{-1} \Gamma_b \quad (\text{Eq. 1.9})$$

where  $k_1$  is the binding constant, and  $k_{-1}$  is the dissociation constant. For both forms of ArtinM, the kinetics rates and affinity equilibrium constants were determined to be statistically equivalent (Table 1.3).

**Table 1.3** Kinetic constant rates of binding and dissociation and kinetic equilibrium affinity constant for jArtinM–HRP and rArtinM–HRP. Reproduced with permission.<sup>120</sup>

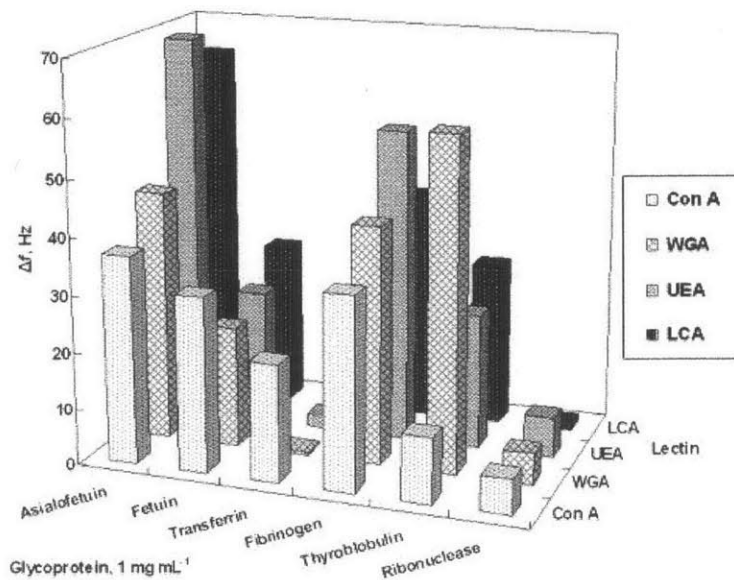
Interaction	$k_1$ (l/mol/s)	$k_{-1}$ (l/s)	$K_a$ (l/mol)
jArtinM–HRP	$28 \pm 4$	$(4 \pm 2) \times 10^{-3}$	$(7 \pm 3) \times 10^3$
rArtinM–HRP	$27 \pm 2$	$(4 \pm 1) \times 10^{-3}$	$(7 \pm 2) \times 10^3$

In other work by the Lekka group<sup>119</sup>, QCM was applied to study the interaction between immobilized Con A and carboxypeptidase Y (CaY - a glycoprotein that catalyzes the hydrolysis of peptidyl-L-amino acids). Frequency modulation was monitored in real-time at different concentrations of CaY (Fig. 1.13). The results showed that  $k_1 = (5.6 \pm 0.1) \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  and  $k_{-1} = 0.095 \pm 0.002 \text{ s}^{-1}$  respectively, and binding constant,  $K_a$ , was calculated as a ratio of the association and the dissociation rate constant  $(0.59 \pm 0.01) \times 10^6 \text{ M}^{-1}$ .



**Fig. 1.13** (a) The real-time response of the motional resistance measured for Con A-modified crystal in TBS (A), for Con A-modified crystal during binding of CaY to Con A. (b) Real-time frequency response curves obtained for Con A–CaY binding process recorded for different concentrations of CaY: (1) 0.1, (2) 0.3, (3) 0.6, (4) 1.0, (5) 2.0, (6) 2.5, and (7) 5  $\mu\text{M}$ . (c) The linear dependence between the relaxation constant and CaY concentration. Datapoints are average values of three independent experiments. The correlation coefficient of the fit was 0.994. Reproduced with permission.<sup>119</sup>

In the work by Danielsson *et al.*,<sup>113</sup> the cross-reactivity of a set of lectins and glycoproteins was studied using a lectin-SAM surface with a QCM. The results are presented in the following figure (Fig. 1.14) using the frequency change as a figure of merit.

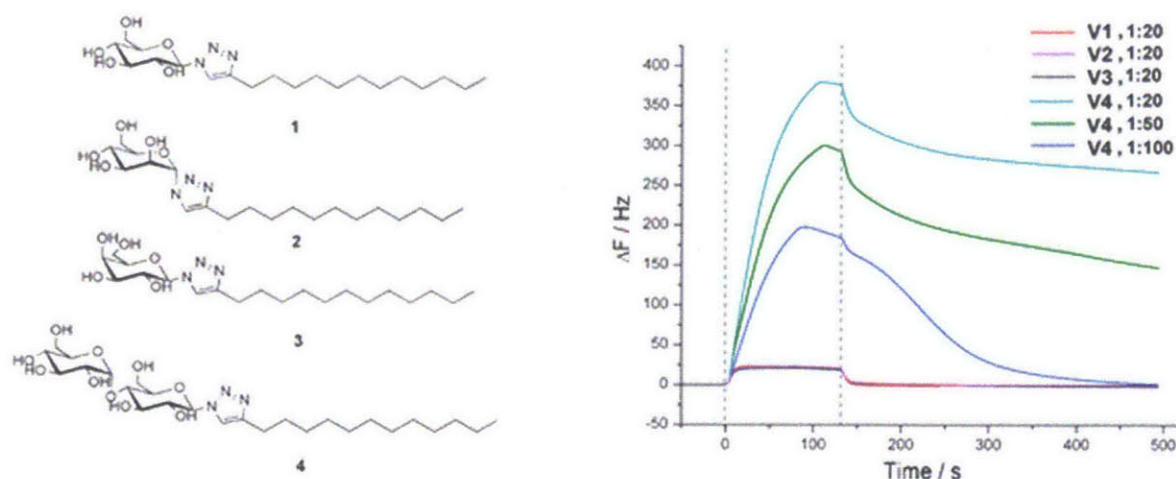


**Fig. 1.14** Affinity of glycoproteins to different lectin surfaces (Con A, WGA, UEA, LCA). PBS (10 mM, pH 7.4) containing 1 mM metal ions ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  and  $\text{Mn}^{2+}$ ) was pumped through the system at a flow rate  $50 \mu\text{L min}^{-1}$ . Glycoproteins ( $1 \text{ mg mL}^{-1}$ ) were injected for 40 min. Surfaces were regenerated with glycine-HCl (pH 2.5). Reproduced with permission.<sup>113</sup>

#### 1.2.3.4 QCM signal amplifier

According to QCM theory (Eq. 1.1), the observed frequency change scales with the change in mass. The current commercially-available QCM systems lack small molecule sensitivities and require additional ‘synthetic amplifiers’ to increase the signal. In addition, the theory assumes that the analyte mass couples tightly to the resonator, which is not always the case. When bacterial cells bind to the surface of the gold, they trap water, which damps oscillation and confounds the signal.<sup>122</sup> Thus, more complicated surface chemistry techniques are needed to ensure that the bacteria bind tightly to the gold surface.

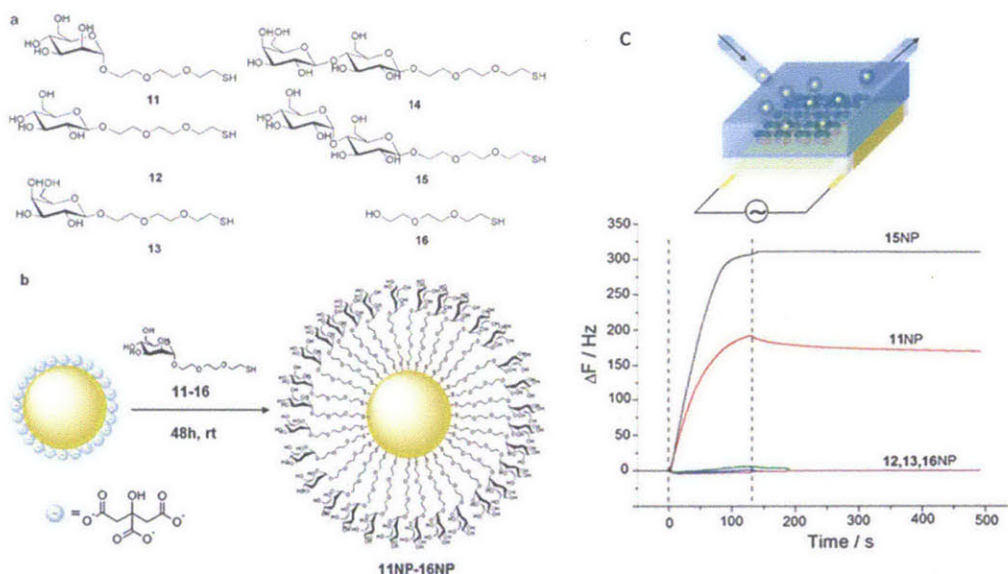
One way to increase the binding is to introduce Con A to a gold surface that is pre-treated with polymannoside and then introduce the alkyl glycosides (Fig. 1.15, left panel)<sup>123</sup>. Although Con A is well known to selectively bind  $\alpha$ -mannosides and  $\alpha$ -glucosides, in this setup only the ‘maltovesicles’ V4 adsorbed specifically while others displayed little adsorption (Fig. 1.15, right panel). These observations were attributed to conformational changes or the orientation of the glycosides on the surface.



**Fig. 1.15 Left: Scheme 1** Alkyl glycosides for vesicle functionalisation:  $\beta$ -glucoside **1**,  $\alpha$ -mannoside **2**,  $\beta$ -galactoside **3**, and  $\beta$ -maltoside **4**. **Right:** QCM adsorption profiles for alkyl glycoside functionalised vesicles at different molar ratio of alkyl glycosides **1–4** to phospholipids (1 : 20/mol : mol = 10  $\mu$ M in glycoside ). Reproduced with permission.<sup>123</sup>

As follow-up, Barboiu *et al.* analyzed the binding constant for Con A with  $\alpha$ -mannosides and  $\alpha$ -glucosides using gold glyconanoparticles to increase the mass of the analytes and their valence.<sup>123</sup> The Con A was immobilized on the surface by two methods. The first method consisted of immersion in an octadecanethiol solution followed by exposure to Con A. The second method introduced an intermediate step whereby polysaccharide Mannan was adsorbed on the octadecanethiol-functionalized surface before Con A deposition, resulting in denser lectin layers. Six gold nanoparticle (NP) solutions were tested with the Con A-treated system, containing respectively mannoside (**11**), glucoside (**12**), galactoside (**13**), lactoside (**14**), maltoside (**15**), and the control, 2-(2-(2-mercapto-ethoxy)ethoxy)ethanol (**16**). Only mannoside and maltoside displayed selective adsorption to the lectin-surface. The lectin layer was resistant to the glucoside and galactoside nanoparticles (Fig. 1.16). Application of a Langmuir model gave affinity constants of  $K_a = 1.6 \times 10^7 \text{ M}^{-1}$  and  $K_a = 7.2 \times 10^7 \text{ M}^{-1}$  for mannoside and maltoside, respectively. In discussion of these results, Barboiu *et al.* stated that the glucoside nanoparticles attach glucose in the  $\beta$ -configuration, whereas the maltose presents terminal glucose in the  $\alpha$ -configuration, which explains the lack of affinity of the glucoside nanoparticles to the Con A surface. The multivalency effect, as a result of the incorporation of the gold nanoparticles, was shown to increase the affinity constant by 2000 times for mannoside-NP relative to methyl- $\alpha$ -mannoside, and by 25,000 times for maltoside-NP relative to D-maltose. Barboiu *et al.* attributed the significant increase in binding affinity to a chelating affect as well as an entropic affect. The decrease in conformational entropy of the system when the ligands are clustered

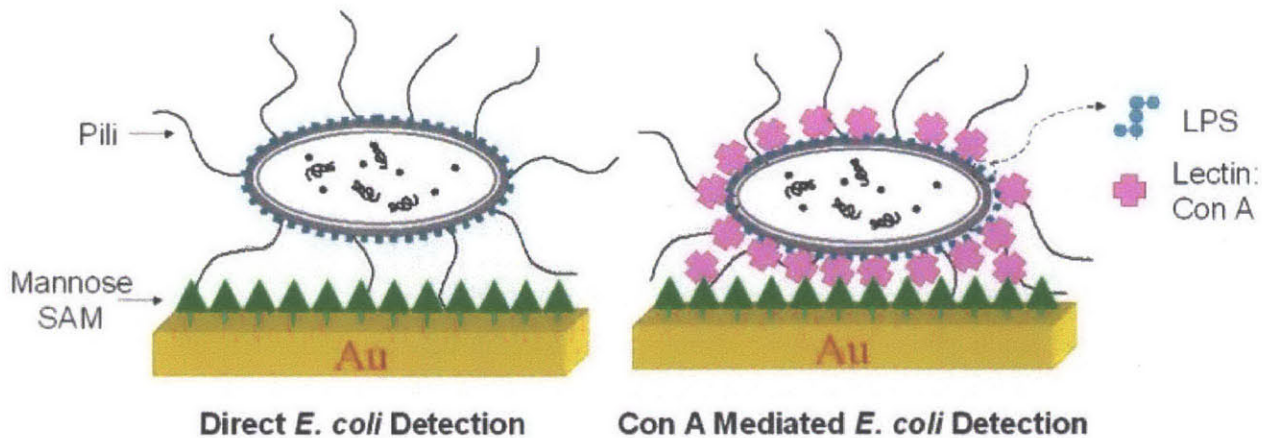
into nanoparticles relative to their free state reduces the entropic cost of a binding event. In their work, the ability of nanoparticles to significantly amplify the affinity of some of the carbohydrate-protein interactions was demonstrated. Yet, in some cases, the incorporation of the nanoparticle changed the configuration of the carbohydrate, as with glucoside, thereby preventing any binding.



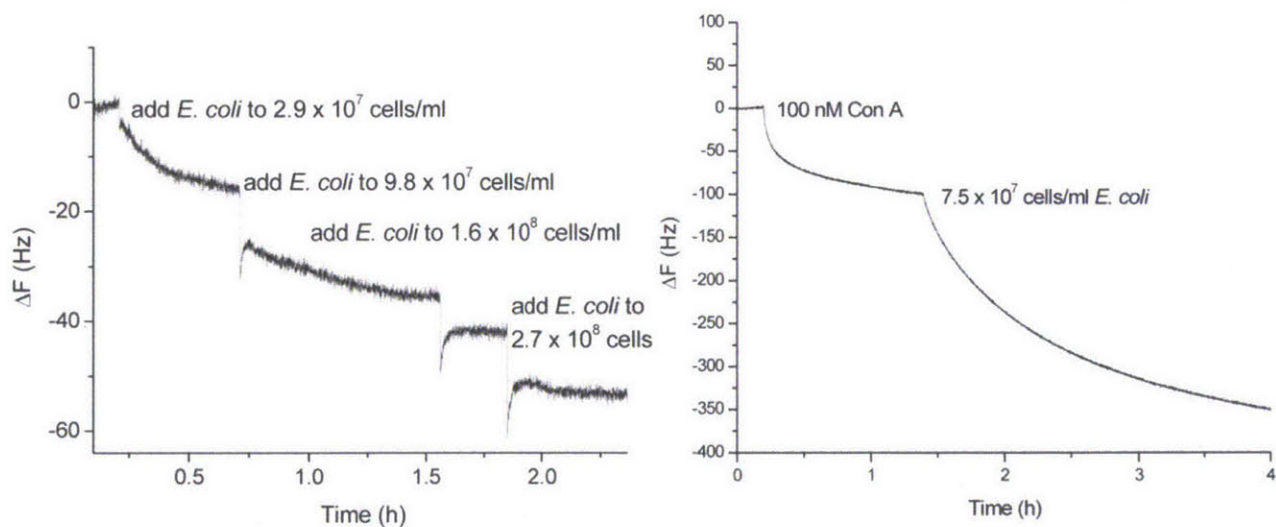
**Fig. 1.16** a) Glycoligands 11–16 used for (b) glycoNPs 11NP-16NP synthesis. c) QCM frequency variation during addition of the glyconanoparticles (8 nM, core side  $d = 12$  nm) to the Con A-treated QCM electrodes. Reproduced with permission.<sup>123</sup>

In order to increase the frequency modulation signal for bacterial adhesion on the gold surface, Wang and co-workers developed an innovative approach to strengthen the binding between the cell and the gold surface<sup>122</sup>. In a conventional approach, cells adhere to the mannose-SAM layer through a fimbriae-mediated adhesion mechanism which is relatively weak (Fig. 1.17, left), and the resultant response curve (Fig. 1.18, left) indicates a detection limit of  $3.0 \times 10^7$  cells/mL. In this approach, Con A was first introduced to adhere on the specific terminal carbohydrates of the bacterial lipopolysaccharides, which increased the contact area between the cell and the surface. The detection limit of the new approach decreased significantly to 750 cells/mL.





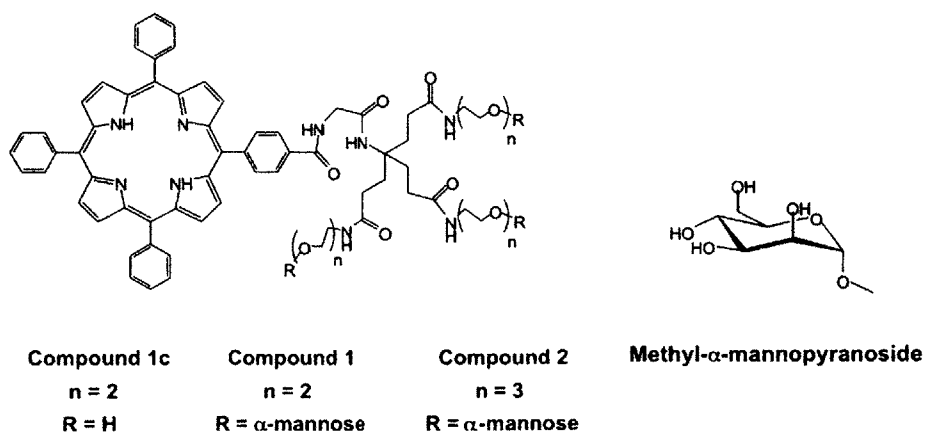
**Fig. 1.17** Schematic presentations of direct *E. coli* detection and Con A mediated *E. coli* detection. Reproduced with permission.<sup>122</sup>



**Fig. 1.18** Left: Frequency change vs. time curve when the mannose-QCM electrode during exposure to different concentrations of *E. coli* W1485 in 1.0 mL stirred PBS buffer (pH = 7.2) with 1mM  $\text{Ca}^{2+}$  and 1 mM  $\text{Mn}^{2+}$ . (The final concentrations of *E. coli* were:  $2.9 \times 10^7$ ,  $9.8 \times 10^7$ ,  $1.6 \times 10^8$ , and  $2.7 \times 10^6$  cells/mL.) Right: Frequency change vs. time curve when the mannose-QCM sensor was first exposed to 100 nM Con A, followed by the addition of *E. coli* W1485 ( $7.5 \times 10^7$  cells/mL) in 1.0 mL stirred PBS buffer (pH = 7.2) with 1mM  $\text{Ca}^{2+}$  and 1 mM  $\text{Mn}^{2+}$ . Reproduced with permission.<sup>122</sup>

### 1.2.3.5 Glycoconjugated porphyrins target receptors of cells

While QCM is useful in quantifying the interactions between lectin and carbohydrates or glycosylated proteins, it can also be used in more applied medical applications, such as designing new therapeutic drugs or detection tools that mimic physiologic interactions between lectins and carbohydrates. Researchers have synthesized sugar-functionalized nanoparticles to target cell receptors that contain sugar-binding domains, taking advantage of the lectin-sugar interaction for cell target studies.<sup>124</sup> For example, in a new therapy for retinoblastoma (a malignant tumor)<sup>125</sup>, a sensitizer that lacks mutagenic properties under visible light is administered to the cancerous cells. Upon the sensitizer's photo-activation, reactive oxygen species are generated that cause irreparable damage to these cells, resulting in cell death.<sup>126</sup> Most of the sensitizers are porphyrin-based compounds and are poorly soluble in water, yet glyconjugation is considered to be a potentially effective way to increase the solubility. In Rosilio's work<sup>124</sup>, glycoconjugated porphyrins are designed to specifically target the receptors to  $\alpha$ -mannose. These have been previously found to be the receptors exposed at the cell surface of neoplastic retinal tissues<sup>127</sup>. In this work, new *meso*-(tetraphenyl) glycodendrimeric porphyrins were synthesized (Fig. 1.19). 1c was a nonglycoconjugated porphyrin with the same structure as 1 and was used as a control for 1, and to a lesser extent, 2. The binding affinity of these porphyrins with concanavalin A (Con A), a mannose-specific lectin, was analyzed with QCM. The Au-coated crystals were coated with a 5 mM ethanolic 11-mercaptopundecanoic acid (MUA) solution. The carboxylic acid functions of MUA were converted to N-hydroxysuccinimide esters by reaction with *N*-hydroxysuccinimide, followed by exposure to Con A. The porphyrins were solubilized in DMPC liposomes to overcome their poor solubility in water. DMPC, DMPC-1c, DMPC-1, and DMPC-2 were introduced to the immobilized Con A surface, and the oscillation frequency shift ( $\Delta f$ ) and the energy dissipation change ( $\Delta D$ ) of the quartz crystal were monitored in the QCM over time. The binding constant values for the porphyrin derivatives with Con A and human serum albumin were determined, and the glycoconjugated porphyrin evidenced a much stronger binding to Con A than the original porphyrin. The study concluded that glycodendrimeric porphyrins could be applied for specific recognition of certain CRD-containing tumors with a higher selectivity than sensitizers or porphyrins that are not conjugated to sugars.



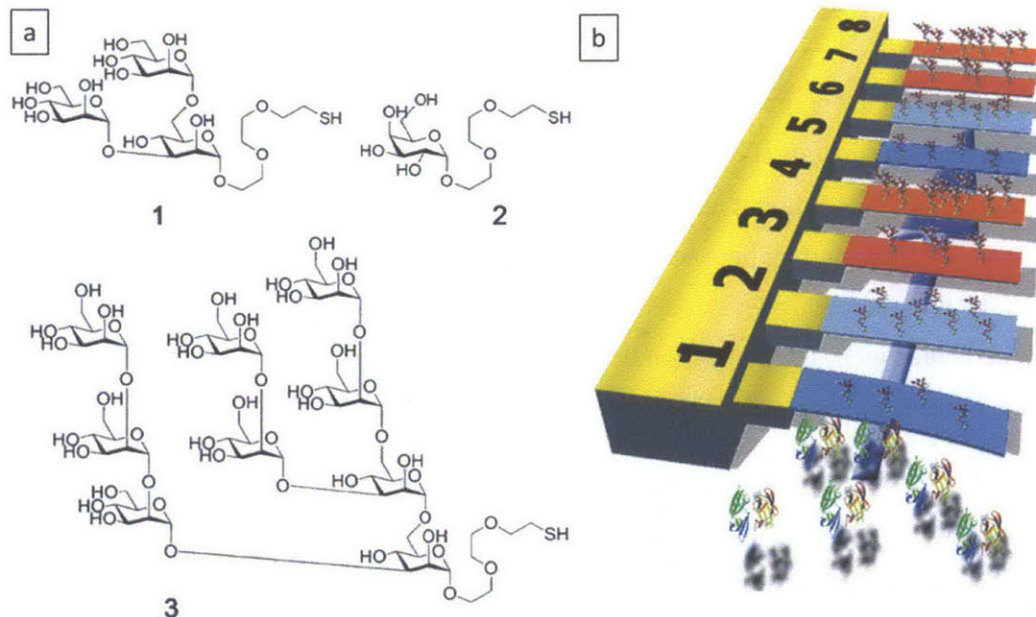
**Fig. 1.19** Chemical structures of the studied porphyrins and of methyl- $\alpha$ -D-mannopyranoside. Reproduced with permission.<sup>127</sup>

### 1.2.4 Cantilever array sensors detecting carbohydrate-protein interactions

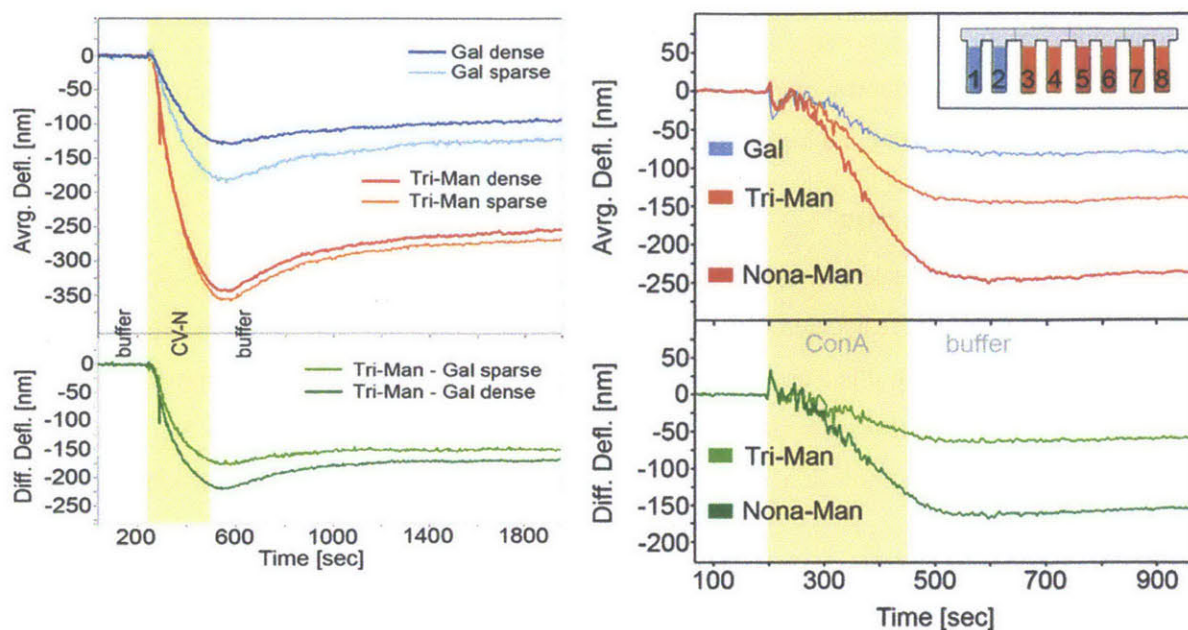
Cantilever microarrays are another emerging class of nanoscale tools for detecting molecular binding using mass change transduction. For example, a label-free, cantilever-array sensor was used to measure picomolar levels of mRNA without target amplification; moreover, these results could be obtained within a few minutes.<sup>128</sup> Upon binding between a receptor and a molecule, the cantilever bends, resulting in the deflection of the laser, allowing for direct detection of the binding events. In the work performed by Hermann<sup>129</sup> and colleagues, the interaction between mannose and an antiviral protein cyanovirin-N (CV-N) was studied, the latter being an 11 kDa protein that has recognizes oligomannosides. Here, the cantilever array was functionalized with trimannose (Figure 1.20a, **1**), nonamannose (Figure 1.20a, **3**), and galactose (Fig. 1.20a, **2**) at different densities using thiol-gold chemistry; galactose served as an internal control. Upon injection of CV-N, the deflection signal of the trimannose cantilevers (Fig. 1.21, red and orange) was approximately 3-4 times larger than that of the galactose cantilevers (Fig. 1.21, dark and light blue).

The nonamannose-functionalized cantilevers exhibited a 20% stronger deflection than trimannose-functionalized cantilevers upon CV-N addition. The galactose-subtracted differential signal was assumed to come from the specific adsorption of the CV-N on the galactose cantilever. Similar measurements were also performed using Con A in place of CV-N. The resulting deflection of the nonamannose was more significant than that of the trimannose, which result from multivalent binding of Con A to nonamannose. Using this technique, a 9.6 nM concentration of Con A (1  $\mu$ g/mL) can be detected even when immediately followed by a Con A pulse of higher concentration. This sensitivity is comparable to other surface techniques like SPR (Section 1.4.2) and QCM.

Cantilever sensors have potential as a quantitative tool for deducing interactions between glycans and CRDs; however, limitations do exist. Thiol-functionalized galactosides require significant synthetic efforts. Additionally, although CV-N and ConA are not galactose lectins<sup>130</sup>, in this study they demonstrate binding to the galactose-modified cantilever; this unexpected binding indicates that the surface chemistry used in this work may have altered some of the native structures of the sugars, introducing non-specific adsorption, which would confound selective sensing applications. For glycoprotein profiling, the detection scheme must be reversed; it would be necessary to immobilize the CRDs on the array surface and use the glycosylated reagent as the analyte. This reverse scheme may have the same mass limitations as SPR (Section 1.4.2) as the large lectins would induce a larger deflection than would a smaller glycosylated product.



**Fig. 1.20** a) Carbohydrates studied in this work, including trimannoside (1), galactoside (2), and nonamannoside (3). b) Scheme of a cantilever array functionalized with thiol-terminated carbohydrates where the galactose-modified cantilever serves as an internal control. While cantilevers 3, 4, 7, and 8 are coated with trimannose to detect CV-N, cantilevers 1, 2, 5, and 6 are coated with galactose as reference. Carbohydrate densities were adjusted using two different incubation times (1, 3, 5, 7 sparse and 2, 4, 6, 8 dense). Upon contact between the sensor surface and CV-N, protein binding results in intermolecular interactions that induce surface stress on the cantilever surface that in turn is relieved by cantilever bending. Reproduced with permission.<sup>129</sup>



**Fig. 1.21** a) *Averaged/differential* cantilever deflection as a function of time during CV-N detection (0.1 mg/mL (9.09  $\mu$ M), shaded area). Top: Upon injection of CV-N, trimannose cantilevers (red and orange) show an about 3–4 times larger deflection signal than the galactose cantilevers (dark and light blue). The more sparsely functionalized trimannose layer (9 min, orange) shows a somewhat larger deflection than the more densely functionalized layer (14 min, red). Each graph represents an *average* signal of two identically functionalized cantilevers. Bottom: Corresponding *differential* signals for two different incubation times (9 min, light green and 14 min, dark green), calculated by subtracting the nonspecific galactose signals from the specific trimannose signals. The *differential* signal is assumed to reflect specific binding events. At the return of the running buffer, the differential signal recovers more slowly than the averaged signal, indicating dissociation of nonspecific binding. B) Average and differential ConA binding signals. Upper panel: Following injection of ConA (2 mg/mL; 19.2  $\mu$ M) the average deflection of the tri- and nonamannose-coated cantilevers was considerably larger than the average deflection of the galactose reference cantilevers. Multivalent and multisite binding by nonamannose resulted in increased deflection. The inset illustrates cantilever array functionalization. Lower panel: The differential deflections represent the specific binding of Con-A to trimannose and nonamannose cantilevers after correction for the nonspecific binding using the galactose reference cantilever. Reproduced with permission.<sup>129</sup>

In summary, mechanical methods, especially QCM, have particular advantages for glycoprofiling, including: 1) the concept and operation of QCM is very simple, and the detection is fast (less than 30 min per run); 2) the detection is quite sensitive; typical binding constants are on the order of  $10^6 \text{ M}^{-1}$ , although detection is also surface-chemistry and analyte dependent; 3) simple but efficient surface chemistry can be readily applied, including gold-thiol chemistry, NHS/EDC chemistry, ‘click’ chemistry, and widely used conjugation chemistry; 4) nonspecific adsorption is a concern but it can be

overcome with careful engineering. There are other limitations in this technology, the main concern being the difficulty in transducing binding events between smaller molecules. Signal amplification, if used, is generally complicated and highly system-dependent.

### 1.3. Electrical Platforms

The concept of electrical transduction is primarily based on the detection or recognition of electrical signals, including resistance, impedance, conductance, potential, and voltage. An advantage of electrical transduction is that it is a nondestructive analytical technique, which makes it a promising method over the above-mentioned traditional methods such as mass spectrometry or HPLC. There are three important electrical methods that are recognized and reviewed herein. First, electrochemical sensors employing nanomaterials have been widely studied as a liquid-phase sensing system for probing intermolecular interactions.<sup>131</sup> Electrochemical impedance spectroscopy is especially utilized for the detection of a variety of biological analytes.<sup>131</sup> Second, field effect transistor (FET) sensors based on carbon nanotubes<sup>132-136</sup> or nanowires<sup>137</sup> have attracted much attention since their first report in 1998.<sup>138</sup> The constrained chemistry of the nanotubes and nanowires results in an observable decrease in conductivity along their length when an analyte binds to their surface. Nanotube FETs (NTFETs) are promising candidates for electronic detection of biological species by measuring the electrical conductance change of nanotubes before and after binding to a biomolecule. Third, biological pores<sup>139</sup> and solid-state nanopores<sup>4</sup> are another emerging class for electrical platforms; they have current applications in biomolecule detection or DNA sequencing by single-channel electrical recording technology. Following, we present a critical review of recent research towards the development of glycoprofiling technology based on these three electrical platforms.

#### 1.3.1 Electrochemical sensors

The development of nanoscale electrochemical platforms has been extensively investigated as an inexpensive, rapid approach to sensitively detect a variety of biological analytes<sup>132,140,141</sup> and the interactions between different biomolecules, including lectin-glycan<sup>133,142,143</sup>, antigen-antibody<sup>144-146</sup>, protein-cell<sup>147,148</sup>, avidin-biotin<sup>135,149</sup>, enzyme-biomolecule<sup>140</sup>, or cell-matrix.<sup>134</sup> The common electrochemical methods include voltammetry<sup>132,140,141,144</sup>, amperometry<sup>132</sup>, potentiometry<sup>132,150</sup>, and electrical impedance spectroscopy (EIS).<sup>131,142-144,146</sup> Generally, nanomaterial-modified electrodes are initially characterized using cyclic voltammetry (CV). Differential pulse voltammetry (DPV), square-wave voltammetry (SWV), amperometry, and EIS are then employed to quantify analytes, in which DPV produces less charging current, which is often applied to discriminate multiple compounds and improve sensitivity.<sup>140</sup> Among others, EIS is a powerful tool for probing the interface features of surface-modified

electrodes and is particularly used to study the stepwise assembly of the composite biosensors. Electrochemical impedance is generally measured by applying a small sinusoidal AC potential (2-10 mV) to an electrochemical cell and measuring the current that crosses through the cell. Thus, if the applied sinusoidal excitation potential  $E_t$  is expressed as

$$E_t = E_0 \cdot \sin(2\pi ft) \quad (\text{Eq. 1.10})$$

the corresponding current signal  $I_t$  could be expressed as

$$I_t = I_0 \cdot \sin(2\pi ft + \phi) \quad (\text{Eq. 1.11})$$

where  $E_0$  is the amplitude of the voltage signal,  $f$  is the frequency expressed in Hertz,  $I_0$  is the amplitude of the current signal,  $\phi$  is the phase angle,  $t$  is the time. In analogy to Ohm's law, the impedance of the system is given by:

$$Z = \frac{E_t}{I_t} = \frac{E_0 \cdot \sin(2\pi ft)}{I_0 \cdot \sin(2\pi ft + \phi)} = Z_0 \frac{\sin(2\pi ft)}{\sin(2\pi ft + \phi)} \quad (\text{Eq. 1.12})$$

For the mathematical treatment of data, a common way to represent the impedance vector model is to use complex notation; the in-phase current response determines the real (resistive) component of the impedance, and the out-of-phase current response determines the imaginary (capacitive) component. In this way all components that generate a phase shift (i.e. the capacitor) will contribute to the imaginary part of the impedance, while the ones that do not produce any phase shift (i.e. the resistance) will contribute to the real part.

$$Z = Z_{real} + iZ_{imag} \quad (\text{Eq. 1.13})$$

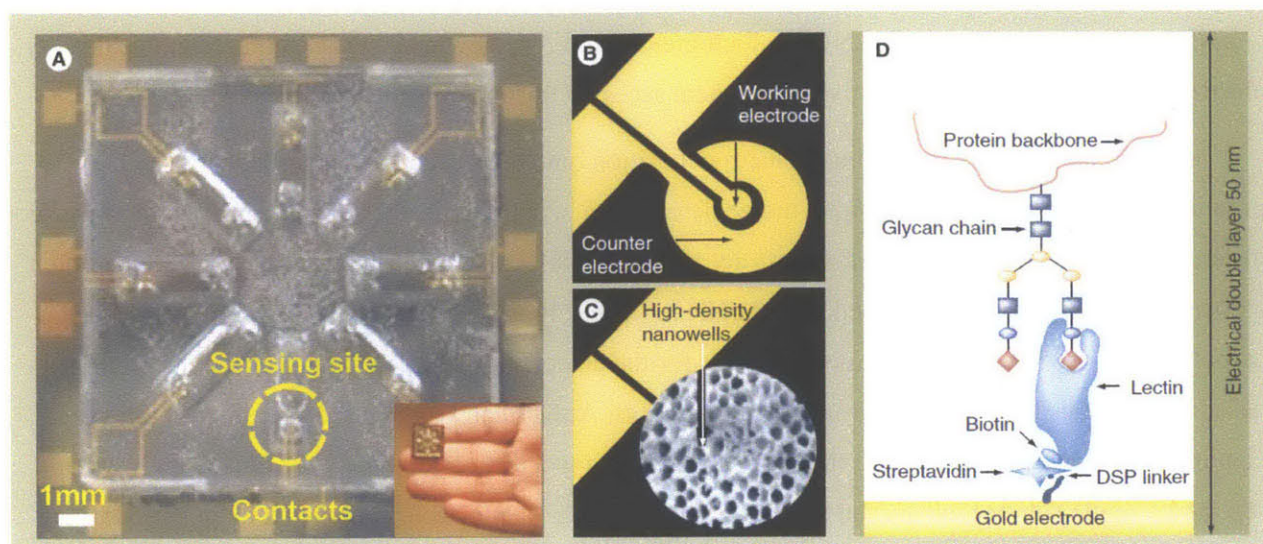
The impedance spectra of EIS are often presented in a Nyquist plot, which includes a semicircle part and a linear part. The semicircle part at higher frequencies corresponds to the electron-transfer limited process, in which the semicircle diameter equals the electron-transfer resistance, while the linear part at lower frequencies describes the diffusion-limited process.

The fabrication of electrodes normally involves a multilayer structured assemble process. First, a "sandwich" electrode is made from selected materials, including silicon chips, carbon or metal electrodes,

carbon nanotubes, graphene, or noble metal nanoparticles. Then the bioactive molecules are “hooked up” to construct a recognition interface by click chemistry or some simple conjugation reactions. For instance, Prasad’s group reported a NanoMonitor device consisting of a silicon chip with an array of gold electrodes overlaid on a nanoporous alumina membrane, in which lectins were conjugated to the surface of the electrode for glycan biomarker detection.<sup>143</sup>

As shown in Fig. 1.22, the NanoMonitor is composed of a microfabricated silicon platform with eight microscale gold-sensing sites, each of which includes a 25  $\mu\text{m}$  diameter working electrode and a 125  $\mu\text{m}$  diameter counter electrode. The eight sensing sites are physically isolated from each other so that parallel measurements can be performed at the same time. The construction of the recognition interface was achieved by first coating the electrode surface with the crosslinking agent, dithiobis succinimidyl propionate (DSP). The disulphide linkage of DSP chemisorbs quickly to the gold surface forming DSP monolayers while the N-hydroxysuccinimide groups are available for binding to the primary amine group of proteins. Then, streptavidin was added and incubated onto the electrode surface containing DSP by bioconjugation binding. Finally, the immobilization of lectins to the functionalized gold surface was achieved by using the biotin-avidin linker chemistry. In this study, three glycoforms including asialofetuin (ASF), 3’-linked terminal sialic acid (3SF), 6’-linked terminal sialic acid (6SF) of the glycoprotein fetuin were analyzed on the NanoMonitor through EIS, based on the principle of double layer capacitive measurement. The experimental results demonstrate that the NanoMonitor has excellent potential for development as a handheld electrochemical sensor for routine detection of glycan biomarkers. Table 1.4 summarizes of the performance of the NanoMonitor relative to that of the lectin-based enzyme linked immunosorbent assay (ELISA) for the detection of these three glycoform variants of fetuin. Compared with the traditional ELISA technique, the NanoMonitor is five orders of magnitude more sensitive as well as highly selective over a broad dynamic range of glycoprotein concentrations. The NanoMonitor is a rapid, completely label-free, and very promising detection platform. However, as discussed in the introduction of this review, the three glycoforms screened in this work are just a few of the many glycoforms that are possible. In this case, Fetuin can adopt over one hundred glycoforms with its three *N*-linked oligosaccharides (each with two to three branches) and one to two *O*-linked glycosylation sites that can also be terminally sialylated. Thus for this platform to resolve more of the glycoforms, especially differentiation between nearly identical species, more sensor sites would need to be employed each with a more accurate measure of binding on a unique CRD (in this case, lectins). This would effectively expand the possible ‘bar codes’ of the fetuin species and assign each a unique code.





**Fig. 1.22** (A) The NanoMonitor consists of a base silicon microplatform with eight gold metal sensing sites to form a hand-held diagnostic device (insert). (B) Each sensing site is composed of a gold working electrode (25µm diameter) and a counter electrode (125µm). (C) The nanoporous alumina membrane is used to create a high density array of nanowells on the sensing site. (D) Schematic representation of the interaction of biomolecules at the electrical double layer. Reproduced with permission.<sup>143</sup>

**Table 1.4** Summary of the performance of the NanoMonitor in comparison to the lectin-based ELISA for the detection of the glycoform variants of fetuin. Reproduced with permission.<sup>143</sup>

Performance parameter	Lectin-based ELISA	NanoMonitor
<b>Detection sensitivity</b>		
3' sialylated fetuin	0.156 µg/ml	1 pg/ml
6' sialylated fetuin	0.312 µg/ml	1 pg/ml
<b>Detection specificity</b>		
3' sialylated fetuin	≥0.156 µg/ml	≥1 pg/ml
6' sialylated fetuin	≥0.312 µg/ml	≥100 pg/ml
Linear range of detection	0.312–5 µg/ml	1 pg/ml–10 ng/ml
Volume per assay	50 µl minimum	10 µl maximum
Response time per sample	Approximately 4 h	<15 min
Detection method	Enzyme-linked colorimetric assay	Label-free impedance

Compared with a bare carbon electrode, the nanocomposite-modified electrode offers a more homogeneous surface for biomolecule or living cell loading, which enhances the reproducibility of both the biosensor and the obtained signals. The most widely employed nanomaterials are gold nanoparticles

(AuNPs) and carbon nanotubes (CNTs), which greatly improve the sensor surface area, biocompatibility, electrical conductivity and connectivity, chemical accessibility and electrocatalysis.<sup>131</sup>

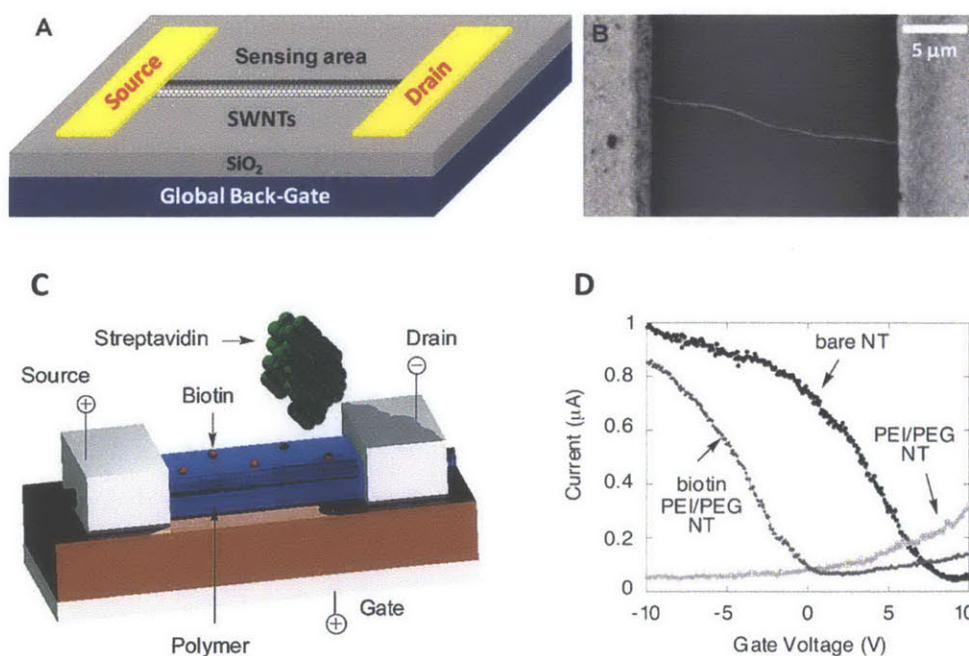
AuNPs can be layered through electrodeposition onto the surface of electrodes, on which a protein, oligonucleotide or other probe molecule can then be immobilized as the recognition interface. For instance, Penadés' group reported a nanostructured disposable impedimetric sensor using screen-printed carbon electrodes (SPCEs) modified with AuNPs.<sup>142</sup> The self-assembled monolayer (SAM) of thiolated carbohydrate was conjugated on AuNPs/SPCEs for the detection of biomolecular specific interactions based on reactions between carbohydrates and proteins. AuNPs can also be used as a solution-phase conjugation reagent for amplifying the signal or optimizing detection frequency for EIS. Belle's group reported a single sensor multiplexed marker assay, in which AuNPs were conjugated to an antibody in solution-phase to help tune the impedance frequency of the proteins away from one another for better signal processing.<sup>146</sup>

CNTs are normally incorporated onto electrodes by directly growing CNTs on the electrode surface, adsorbing them on existing electrodes, polymer coating or wrapping, or using a binder to make a paste electrode. For instance, Ju's group reported poly(diallyldimethylammonium) (PDDA) polymer functionalized multi-wall CNTs (MWCNTs) on glassy carbon electrode (GCE), on which negatively charged concanavalin A (Con A) was adsorbed, binding to the positively charged PDDA, for the specific recognition of cell-surface glycans.<sup>148</sup> This ConA/PDDA/MWCNT/GCE composite electrode demonstrated a linear response to K562 cells ranging from  $1 \times 10^4$  to  $1 \times 10^7$  cells/mL. Yuan's group reported a biosensor based on tris(2,2'-bipyridyl)cobalt(III) and MWCNTs-Nafion composite for immunoassay of carcinoma antigen-125, in which both AuNPs and MWCNTs were applied to assemble the sensing platform.<sup>144</sup> In addition, CNT itself can be modified by doping nitrogen, boron, or other elements to improve sensitivity and selectivity for a specific analyte. Zhu's group reported a 3D-architectural biosensor by combining PDDA-functionalized nitrogen-doped MWCNTs having a nitrogen content of 3.2% (PDCNx), with thionine (TH<sup>+</sup>) and AuNPs via the layer-by-layer method.<sup>147</sup> This ConA/AuNPs/TH<sup>+</sup>/PDCNx/GCE demonstrated excellent analytical performance for the detection of HeLa cells ranging from  $8.0 \times 10^2$  to  $2.0 \times 10^7$  cells/mL with detection limit of 500 cells/mL.

### 1.3.2 Nanotube field effect transistor (NTFET) sensors

In NTFETs, a single SWNT or a network of SWNTs acts a conducting channel between two metal electrodes (source and drain) while the two electrodes are held at a constant bias voltage (Fig. 1.23).<sup>135,151</sup> The design principle of NTFET sensors is based on high conductance sensitivity of semiconducting SWNTs when exposed to the environment. In an NTFET sensor, each atom on the surface of the SWNTs

is exposed to the surroundings such that even small changes in the environment can cause drastic changes to the SWNTs' electrical properties. All kinds of organic, inorganic, and biological materials may be applied to functionalize the surface of SWNTs for improving their biocompatibility, sensitivity, and selectivity by covalent or non-covalent binding. Covalent modification of SWNTs tends to create defects in the SWNTs, which may hamper their electronic properties. To avoid these defects, non-covalent modification is commonly employed. By  $\pi$ - $\pi$  stacking or hydrophobic interaction, various synthetic polymers and biomolecules including proteins, DNAs, polysaccharides, enzymes, antibodies, bacterium, and living cells are adsorbed or incubated on NTFETs to investigate specific biological interactions.<sup>132,134,135,151-154</sup> Widely used synthetic polymers include Tween 20, poly(ethylene imine), poly(ethylene glycol) and their derivatives. These polymers are employed not only as the recognition interface but also as a coating to protect the SWNTs from non-specific binding of interferences.<sup>132</sup> The suggested biosensing mechanism for NTFET sensors may involve electrostatic gating<sup>136,155-157</sup>, changes in gate coupling<sup>154</sup>, carrier mobility changes<sup>158,159</sup>, and Schottky barrier effects.<sup>136,157</sup> One promising application of NTFET sensors for the detection of glycan-CRD interactions was reported by Star's group who made a glycoconjugated SWNT network FET sensor, which demonstrated very good specific binding of bacterial lectins.<sup>133</sup> However, semiconducting SWNTs with higher purity are required to achieve better signal quality as a further research goal.



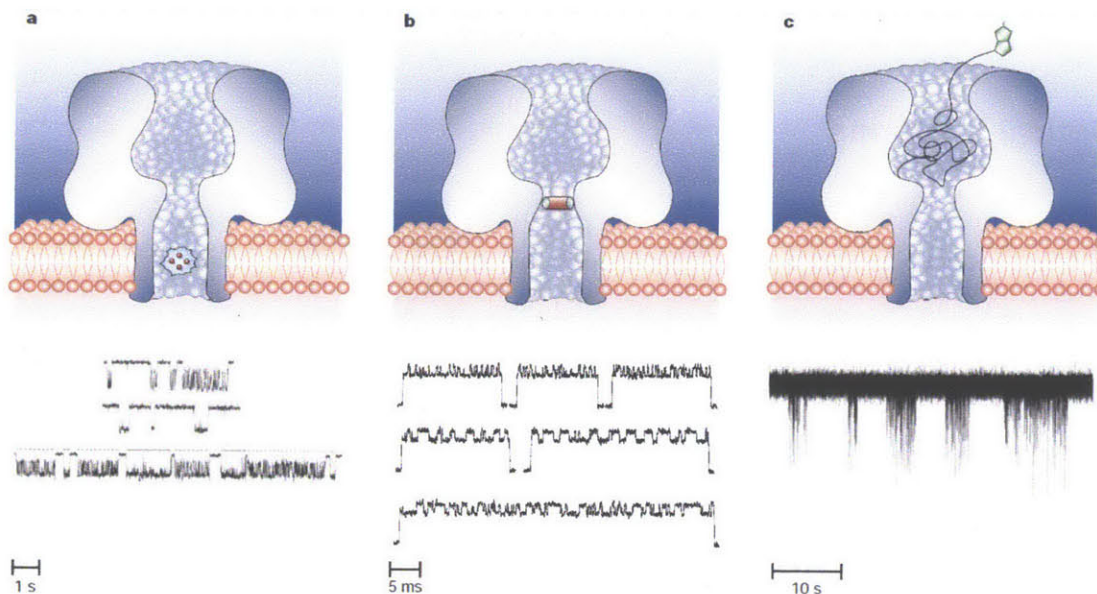
**Fig. 1.23** (A) Schematic representation of a NTFET sensor.<sup>151</sup> (B) A representative scanning electron micrograph (SEM) of an individual straight SWNT connecting source and drain electrodes on a silicon oxide (SiO<sub>2</sub>) surface.<sup>151</sup>

(C) Working scheme of the NTFET sensor. A polymeric functional layer, which coats the SWNT, is functionalized with a molecular receptor, biotin, a protein that recognizes a biomolecule, streptavidin.<sup>135</sup> (D) I-V characterization of the NTFET sensor before and after PEI/PEG polymer coating and after biotin attachment to the polymer layer. Reproduced with permission.<sup>135</sup>

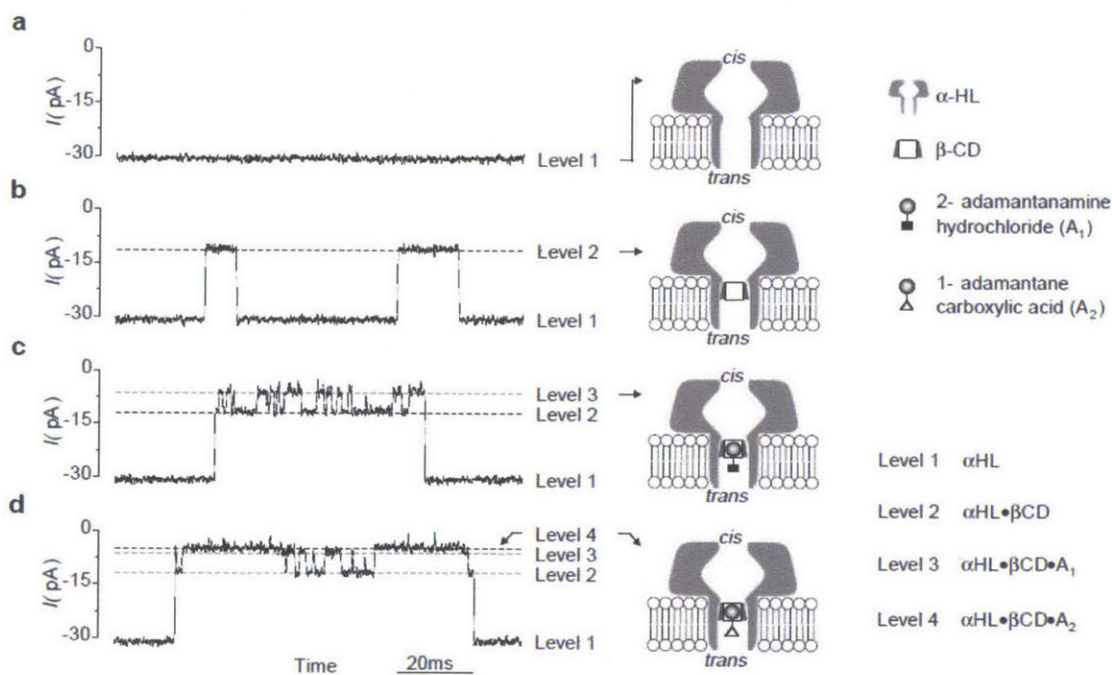
### 1.3.3 Nanopore sensors

Nanopore sensors with single-channel electrical recording is an emerging technology for the stochastic sensing of a variety of molecules.<sup>160</sup> As shown in Fig. 1.24, when different analytes pass through the nanopore, the current signal will report binding events due to the partial blocking of the ionic current.<sup>160</sup> Furthermore, different molecular adaptors can be placed in the lumen of the pore for periods sufficiently long to observe host-guest interactions. The binding events between adaptors and target molecules cause current fluctuations that permit the quantification and identification of the target molecules (Fig. 1.25).<sup>161</sup> Very recently, Bayley's group reported two new nanopore biosensors based on  $\alpha$ -hemolysin ( $\alpha$ -HL) protein pore combined with disulfide cyclodextrin adaptor for molecular recognition.<sup>139</sup> Their results demonstrate that with skeleton-modified cyclodextrins, the sensing selectivity could be tuned toward different target molecules, which may indicate a promising application for chirality sensing. For instance, since their data suggest a successful asymmetric alteration of both the conformation and the cavity of the cyclodextrin adaptor, we could extend this technology to make a series of adaptors with different chiral features, which may find application in chiral recognition or even separation of chiral biomolecules or pharmaceutical molecules.

Compared with mechanical platforms, especially QCM, these electrical platforms possess better anti-interference ability and biocompatibility, and they are easier to fit in a present sensing platform. However, they have disadvantages. For NTFETs, the sensitivity of the whole system is directly dependent on the purity of the semiconducting SWNTs as the presence of metallic SWNTs will cause an inferior conduction channel. However, the separation and purification of semiconducting SWNTs is a challenging research goal that is being actively pursued. As for the nanopore platform, the characterization of each new sensor is not easy and requires lots of expertise and time. Nevertheless, one of these electrical platforms may be integrated in a future system for label-free and non-destructive glycoprofiling.



**Fig. 1.24** Detection of a variety of analytes by stochastic sensing. (a) metal ions; (b) organic molecules; (c) proteins. Reproduced with permission.<sup>160</sup>



**Fig. 1.25**  $\alpha$ -hemolysin ( $\alpha$ -HL) protein pore with  $\beta$ -cyclodextrin ( $\beta$ -CD) adaptor as a sensing platform for the detection of model molecules 2-adamantanamine ( $A_1$ ) and 1-adamantanecarboxylic acid ( $A_2$ ). Reproduced with permission.<sup>161</sup>

## 1.4. Optical platforms

A few optical platforms are currently used to transduce glycan-CRD binding. They can be broadly separated into two categories: (1) non-equilibrium methods and (2) kinetic or near-equilibrium methods.<sup>20</sup> Equilibrium platforms include the microarray methods mentioned in the introduction (i.e. lectin arrays or lectin sandwich arrays) as well as other similar, fluorescent detection schemes like ELISA.<sup>162-165</sup> In these methods, no absolute quantitative binding information is given ( $k_b$ ,  $k_r$ , or  $K_D$ ) but rather relative binding information. A microarray screening can determine which glycan-CRD pair has the greatest affinity and how it relates to the other pairs screened on the same array. This limits their application to a glycoprofiling platform. Conversely, non-equilibrium methods can yield quantitative information; these include optical methods such as frontal affinity chromatography, surface plasmon resonance, and optical nanotransducers. A critical review of these technologies follows.

### 1.4.1 Fluorescent frontal affinity chromatography for kinetic parameters

Frontal affinity chromatography (FAC) is a widely used method for determining glycan-lectin kinetics since the 1970s. As the name connotes, FAC is an advance to the chromatographic processes mentioned in the introduction. Unlike simple affinity chromatography, whose aim is selective binding of glycan entities for separation into different elution volumes, FAC is a quantitative analysis of the advancing front in an affinity column. In their foundational paper<sup>166</sup>, Kasai and Oda demonstrate how analyzing the binding of an advancing front can yield  $K_D$  for a given glycan-CRD pair. Their initial method was nanoscale in its immobilization of lectin transducers and truly label-free, making acquisition of the proper volume fractionations time-intensive and laborious. Subsequent enhancements, including HPLC automation, data-analysis algorithms, and a coupled fluorescent detector, helped speed up and simplify the process, at the cost of requiring fluorescently-labeled glycan reagents.<sup>167</sup> An alternative to identifying fluorescently labeled analytes is to couple the FAC system with mass-spectrometry and analyze each of the species as they elute through the column; in this way both kinetic affinities and chemical identities can be achieved for a large population of analytes.<sup>168</sup> The FAC-MS method has not yet been widely adopted for glycan-CRD analysis, with one literature demonstration being interactions of blood trisaccharides with the lectin from *Marasmius oreades*.<sup>169</sup>

The current fluorescent-based FAC system for glycan-protein interactions is described well in this methods paper.<sup>170</sup> Dilute solutions of glycan are passed through a Lectin-immobilized column (Fig.

1.26a) at a constant concentration  $[A_o]$ . If the glycan has no affinity, it passes through the column (blue in Fig. 1.26) and the glycan front is fluorescently detected faster than the front of an affined glycan (red in Fig. 1.26), which is delayed by binding to the lectins in the column. The dissociation constant of the glycan lectin pair is determined by the governing equation of FAC (Eq. 1.14):

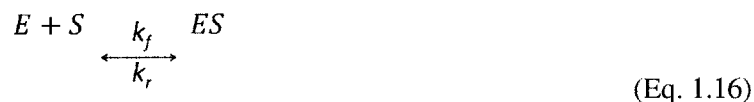
$$[A_o](V - V_o) = \frac{B_t[A_o]}{([A_o] + K_D)} \quad (\text{Eq. 1.14})$$

Where  $[A_o]$  is the initial glycan concentration applied to the column,  $B_t$  is the approximate mole amount of lectin in the column, and the  $(V-V_o)$  term is the difference in the elution front volume of the affined and non-affined glycan, as described in Fig. 1.26 b-d. The elution concentration of the glycan eventually approaches the concentration that is being applied, as the lectins become fully loaded by the glycan. This signal plateau (Fig. 1.26d) must be clearly defined in order to determine the elution volume.

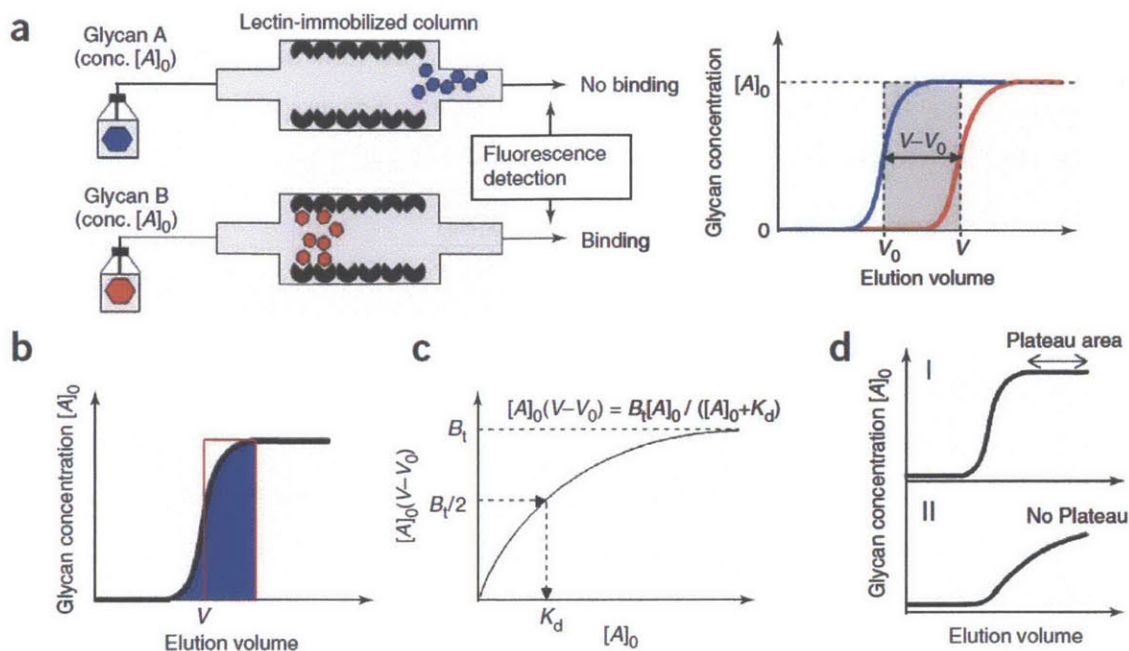
To better understand the equation of FAC analysis (Eq. 1.14), one should inspect its analogue: the equilibrium form of the Michaelis–Menten equation (Eq. 1.15).

$$[ES] = \frac{[E]_o[S]}{K_D + [S]} \quad (\text{Eq. 1.15})$$

This is the governing equation for a classic enzymatic reaction (Eq. 1.16) where a number of enzymes (E) binds with sample (S) to form a bound species (ES)



This reaction is governed by a forward and reverse reaction, whose divisor ( $k_r/k_f$ ) is an inverse measure of the affinity of the species, known as the dissociation constant ( $K_D$ , inverse of the  $K_a$  constant used in Section 1.2). The smaller the  $K_D$ , the more affined the two species are to one another. For reference, lectin-glycan pairs<sup>98</sup> typically have a  $K_D$  in the range of  $10^{-3}$  to  $10^{-7}$ . Comparing Eq. 1.14 with Eq. 1.15 illustrates that FAC analysis is simply the kinetic analysis of surface binding. The number of lectin binding sites is finite ( $B_t$ ) and the left hand term of equation 2 ( $[A_o](V - V_o)$ ) is a measure of the bound lectin sites at equilibrium.



**Fig. 1.26** Experimental scheme for frontal affinity chromatography (a) and method of calculating the elution front volume (b-d). Reproduced with permission.<sup>170</sup>

Hirabayashi and his team established fluorescent FAC as a tool for oligosaccharide-lectin interactions. In their founding works they screened various oligosaccharides to lectin-immobilized columns, as they did for 12 glycans to LEC6 (a galectin or galactose binding lectin from the nematode *Caenorhabditis elegans*) in this work.<sup>167</sup> The glycans were fluorescently labeled and detected using a FAC system coupled with fluorescent detection as detailed above (Fig. 1.26). The FAC traces (Fig. 1.27) provided the loading curves necessary to determine the elution volumes and enabled determination of the dissociation constants of the 12 pairs (Table 1.5). This paper is a good representation of the type of data fluorescent FAC systems can provide. The Hirabayashi group has continued to probe galectin interactions using fluorescent FAC.<sup>171-173</sup> Other groups have used fluorescent FAC for medically relevant glycan interaction studies in the areas of immune cells,<sup>174</sup> ER associated degradation of proteins,<sup>175-177</sup> HIV,<sup>178</sup> disease biomarkers,<sup>179</sup> and glycan binding specificity of many lectins.<sup>180-183</sup>



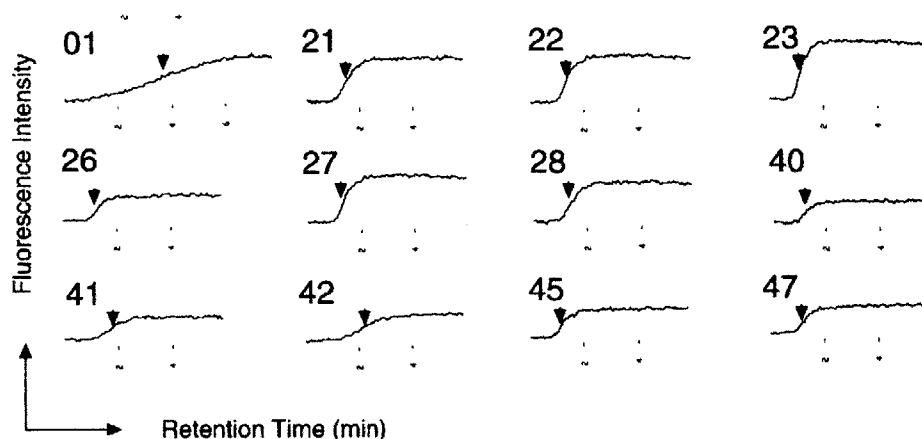


Fig. 1.27 Showing the FAC elution curves necessary to determine the front volume. Reproduced with permission.<sup>167</sup>

Table 1.5 Calculated  $K_D$  for the 12 glycan-LEC6 pairs screened in the Hirabayashi study. Reproduced with permission.<sup>167</sup>

Structure of PA-oligosaccharides used in this study and the obtained dissociation constants ( $K_D$ 's) for LEC-6

No.	Common name	Structure	V (ml)	V - $V_h$ (ml)	$K_D$ for LEC-6 (M)	Relative affinity <sup>a</sup>
01	Biantennary complex	Galβ1-4GlcNAcβ1-2Manα1-6 Galβ1-4GlcNAcβ1-2Manα1-3 <sup>b</sup> ManβR <sup>b</sup>	0.92	0.65	$4.2 \cdot 10^{-5}$	6.4
21	Monosialylated biantennary complex	NeuAcα2-6Galβ1-4GlcNAcβ1-2Manα1-6 Galβ1-4GlcNAcβ1-2Manα1-3 <sup>b</sup> ManβR <sup>b</sup>	0.38	0.11	$2.5 \cdot 10^{-4}$	1.1
22	Monosialylated biantennary complex	Galβ1-4GlcNAcβ1-2Manα1-6 NeuAcα2-6Galβ1-4GlcNAcβ1-2Manα1-3 <sup>b</sup> ManβR <sup>b</sup>	0.35	0.08	$3.2 \cdot 10^{-4}$	0.83
23	Disialylated biantennary complex	NeuAcα2-6Galβ1-4GlcNAcβ1-2Manα1-6 NeuAcα2-6Galβ1-4GlcNAcβ1-2Manα1-3 <sup>b</sup> ManβR <sup>b</sup>	0.29	0.02	$1.4 \cdot 10^{-3}$	0.2
26	Asialo GM3 (lactose)	Galβ1-4Glc-PA	0.34	0.07	$3.9 \cdot 10^{-4}$	0.69
27	Asialo GM2	GalNAcβ1-4Galβ1-4Glc-PA	0.31	0.04	$6.9 \cdot 10^{-4}$	0.39
28	Asialo GM1	Galβ1-3GalNAcβ1-4Galβ1-4Glc-PA	0.37	0.10	$2.7 \cdot 10^{-4}$	0.98
40	Forsman pentasaccharide	GalNAcα1-3GalNAcβ1-3Galα1-4Galβ1-4Glc-PA	0.32	0.05	$5.5 \cdot 10^{-4}$	0.49
41	Lacto-N-neotetraose	Galβ1-4GlcNAcβ1-3Galβ1-4Glc-PA	0.43	0.16	$1.7 \cdot 10^{-4}$	1.6
42	Lacto-N-tetraose	Galβ1-3GlcNAcβ1-3Galβ1-4Glc-PA	0.53	0.26	$1.1 \cdot 10^{-3}$	2.6
45	Lacto-N-fucopeptase III	Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ-PA Fucα1-3	0.31	0.03	$9.1 \cdot 10^{-4}$	0.29
47	A-Hexasaccharide	GalNAcα1-3Galβ1-3GlcNAcβ1-3Galβ1-4Glc-PA Fucα1-3	0.33	0.06	$4.6 \cdot 10^{-4}$	0.59

<sup>a</sup> Relative affinity calculated on the basis of  $K_D$  for *p*-aminophenyl β-lactoside.

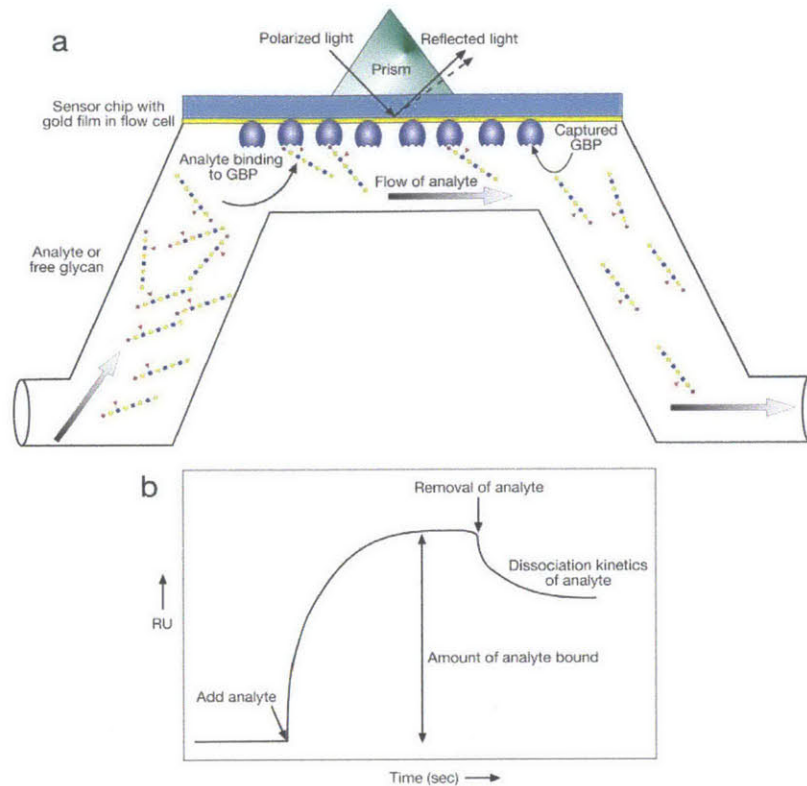
<sup>b</sup> R = 1-4GlcNAcβ1-4GlcNAc-PA.

The advantages of FAC have been delineated<sup>170</sup> as 1) the equilibrium kinetic analysis is a simple, clear equation; 2) the method is capable of measuring weak interactions ( $K_D > 10^{-7}$ ); 3) data is reliably reproduced; 4) only small amount of the glycan is necessary (0.5 pmol/assay); and 5) the lectin concentration does not need to be known to a high precision. Possible drawbacks were also noted as 1)

lectin immobilization may affect its binding properties; 2) proteins may be unstable during the assay; 3) the complete screening takes a few days; 4) a large amount of lectin is needed for full analysis (0.5 mg); and 5) crude, heterogeneous samples cannot be analyzed. Another apparent drawback is the labeling requirement; the glycosylated analytes must be fluorescently tagged, most often with pyridylamine groups,<sup>167,184</sup> to produce the frontal binding curves required for kinetic analysis.

#### **1.4.2 Surface plasmon resonance for kinetic parameters**

Another nanoscale optical technique also analyzes the loading of surface sensors for kinetic analysis of bimolecular interactions, but without the need of labeled reagents: surface plasmon resonance (SPR). First applied by Liedberg *et al.*<sup>185</sup> in 1983, this technique has become an essential laboratory tool; there are many recent reviews of SPR techniques for measuring bimolecular interactions<sup>186,187</sup> as well as a review<sup>188</sup> and book chapter<sup>189</sup> specifically for plasmonic techniques used for oligosaccharide-CRD interactions. In SPR measurements, one binding ligand is immobilized to a gold surface and the other binding pair is transported across the surface (Fig. 1.28a). The refractive index of the surface changes as the surface ligands become occupied. The surface loading can thus be transduced with a careful optical setup, resulting in a loading curve (Fig. 1.28b) similar to those seen for QCM measurements (SRC) and fluorescent FAC.<sup>167</sup> A typical SPR experiment consists of three phases: 1) association, 2) dissociation, and 3) regeneration. In the association phase, the analyte is introduced at a constant concentration and the surface is fully loaded. In the dissociation phase, the analyte flow is stopped, and the bound analytes are washed away. In the regeneration phase the surface is given time to recalibrate to its original refractive index in preparation for the next measurement. These phases are replicated with varying analyte concentration to determine the binding kinetics of analyte to surface-bound ligand. This technique has achieved widespread use due to simple-to-use commercial instruments, like the BiaCore™ system.



**Fig. 1.28** Surface plasmon resonance detector for glycan-lectin interactions. A) Lectins are immobilized on a thin gold film and glycosylated analytes are flown across the surface. As analytes bind to the surface, the refractive index changes altering the path of a polarized light source. B) The resulting light signal is reported in a “sensogram” showing the characteristic loading curve after analyte addition. Reproduced with permission.<sup>20</sup>

The kinetic analysis of SPR loading curves is also a straightforward analogue of a Michaelis–Menten reaction.<sup>190</sup> In the case of lectin-glycan interactions, if we assume the same surface reaction form as FAC experiments (Eq. 1.16) we can write a reaction equation as



where  $G$  is the glycosylated analyte,  $L$  is the lectin binding sites and  $GL$  is the bound complex. Given this model, we would expect the following kinetic model to express the rate of change of bound complex:

$$\frac{d[GL]}{dt} = k_f[G][L] - k_r[GL] \quad (\text{Eq. 1.18})$$

The concentration of the free Lectin sites at a given time can be expressed as:

$$[L]_t = [L]_{t_0} - [GL]_t$$

(Eq. 1.19)

Thus, equation 1.18 can be rewritten as:

$$\frac{d[GL]}{dt} = k_f[G]([L]_{to} - [GL]_t) - k_r[GL]$$

(Eq. 1.20)

We assume that the analyte concentration, [G], is constant (as the bulk of fluid above the SPR surface is large in comparison to the number of lectin binding sites). The change in our SPR signal (measured in resonance units, RU) is a measure of the change in bound complex [GL], so an analogous form of equation 1.20 for an SPR system would be:

$$\frac{dRU}{dt} = k_f C_g (RU_{max} - RU_t) - k_r (RU_t)$$

(Eq. 1.21)

This equation can be rearranged to lump the kinetic parameters together into one kinetic variable ( $k_s$ ) to fit SPR data:

$$\frac{dRU}{dt} = k_f C_g RU_{max} - k_s (RU_t)$$

(Eq. 1.22)

$$k_s = k_f C_g + k_r$$

(Eq. 1.23)

Integrating equation 1.22 yields the equation for the absorption curve:

$$RU_t = M(1 - e^{-k_s t}) + RU_o$$

(Eq. 1.24)

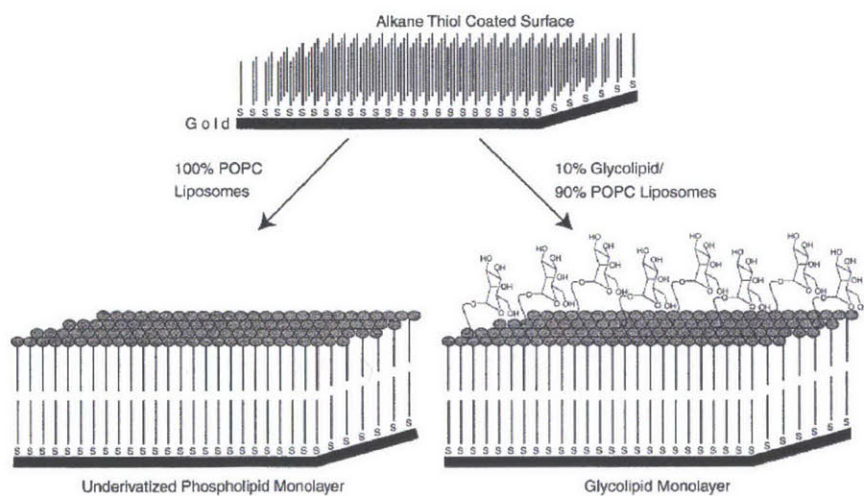
$$M = \frac{k_f C_g RU_m}{k_f C_g + k_d}$$

(Eq. 1.25)

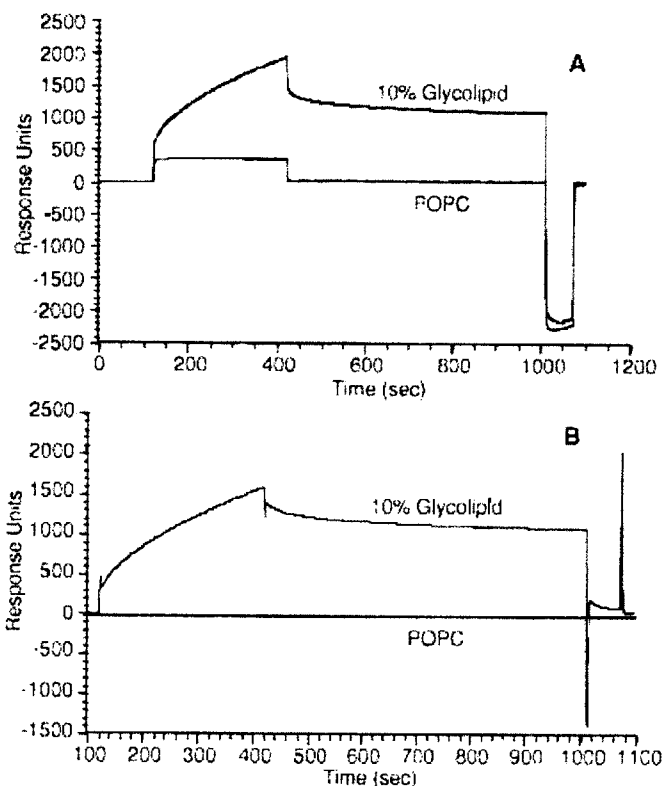
Thus, by obtaining absorption curves at three to four different concentrations of glycosylated analytes ( $C_g$ ) and fitting them to Equation 1.24, one can plot  $k_s$  versus  $C_g$ . If the reaction is governed by the first-order surface reaction (Eq. 1.16), a linear trend should emerge which, when fitted, will yield  $k_f$  and  $k_r$  as the slope and y-intercept, respectively (Eq. 1.23). This analysis method, called integrate rate analysis, is the basis of the software packages usually coupled with SPR machines. It is better at estimating the

error<sup>191</sup> and deviation from the ideal surface reaction (Eq. 20) than the prior method of SPR analysis, called the linearization method.<sup>192-194</sup>

There are many research studies spanning the last two decades using SPR for measuring glycan-CRD interactions, those included herein are only a sampling of these works to illustrate how SPR is used. First is the Mann *et al.* study of 1998 where they probed ligands for Concanavalin A (ConA) using SPR.<sup>195</sup> As discussed in Section 2, Con A is a tetrameric plant lectin expressed from jack-bean, *Canavalia ensiformis* and shows specificity for terminal mannosyl and glucosyl groups.<sup>130</sup> Its binding behavior is well-known and thus often a candidate for new systems. The investigators note the difficulty of measuring the binding of low molecular weight carbohydrate (200 Da) to immobilized lectin via SPR, so they use Con A as the analyte and form a synthetic lipid-mannose layer on the gold sensor surface (Fig. 1.29). By controlling the molar amount of mannose, they could tailor surfaces with varying densities of carbohydrate. For their control, they created surfaces with only lipid (phosphatidylcholine or POPC). They tested these surfaces with monovalent Con A derivatives and stronger, multivalent Con A. The expected loading curve of SPR measurements was observed (Fig. 1.30). Their study demonstrates the control of SPR surface functionalization to probe the effect of multivalent and monovalent interactions. They also illustrated how SPR could be used to analyze the effect of various inhibitors on glycan-lectin interactions, having significant implications in pharmacology.

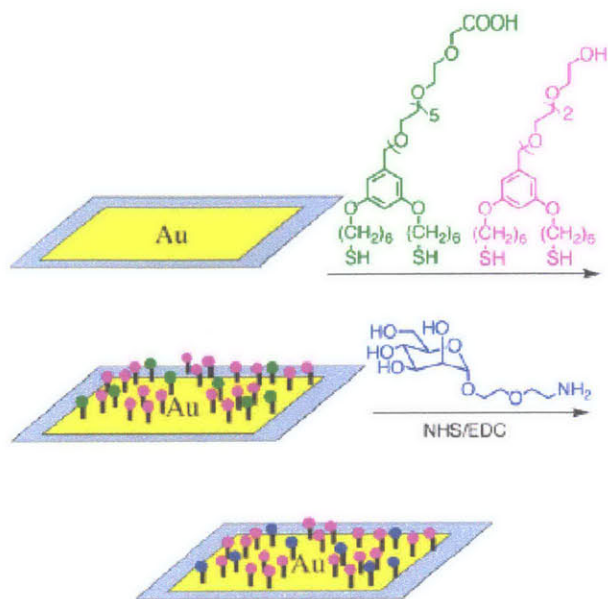


**Fig. 1.29** Lipid carbohydrate preparation of gold surface for SPR measurements. Reproduced with permission.<sup>195</sup>

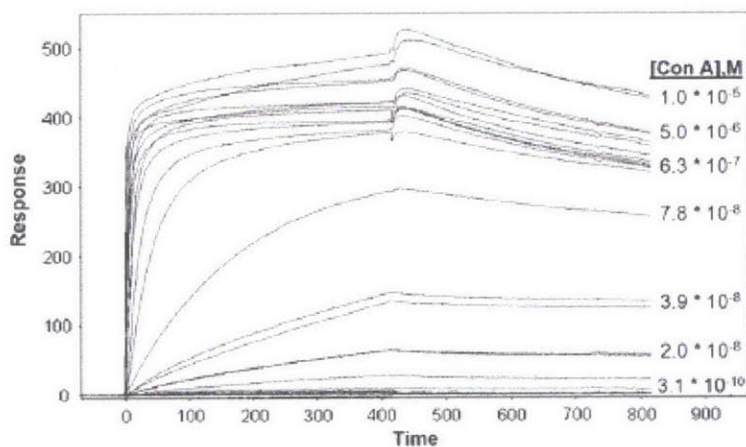


**Fig. 1.30** SPR response of ConA to glycolipid surface and control (POPC) as raw signal (A) and subtracting background bulk refractive index change (B) showing the expected loading curve. Reproduced with permission.<sup>195</sup>

A more recent study by Schlick *et al.* also utilized ConA and mannose in SPR experiments.<sup>196</sup> The mannose was immobilized on the gold surface via a self-assembled monolayer of dithiols coupled to amine-functionalized mannose derivatives (Fig. 1.31). This functionalization scheme resulted in a very stable surface that could be used for multiple assays. Their reported loading curves are similar in shape to the Mann *et al.* study conducted more than a decade prior, but illustrate the replication possible with an automated system. The loading curves are reported at two replicates of seven different analyte (Con A) concentrations in a “sensogram” (Fig. 1.32). The automated SPR software yields a  $K_D$  of 78 nM. Inhibition was also studied in this work by adding glycodendrimers to the solution which competitively bound to the Con A in solution. These SPR competition assays help confirm the affinity enhancements due to multivalency.



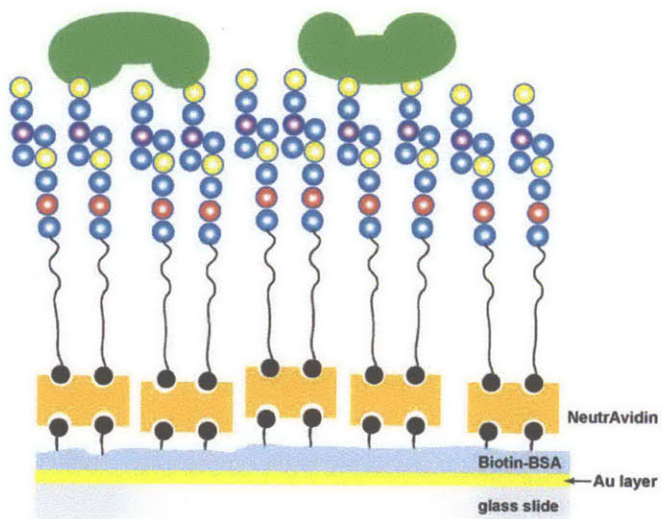
**Fig. 1.31** Gold surface functionalization by self-assembled monolayers bearing dithiols coupled to amine functionalized mannose derivatives. Reproduced with permission.<sup>196</sup>



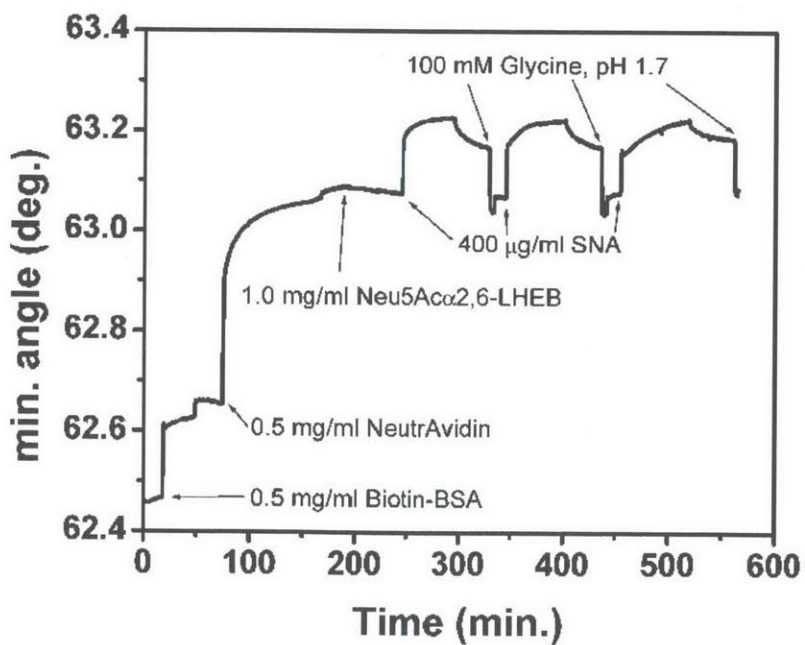
**Fig. 1.32** Sensogram illustrating the SPR loading curves of ConA to mannose surface injected at 7 series dilutions. The measured  $K_D$  is found to be 78nM. Reproduced with permission.<sup>196</sup>

Another method of SPR surface immobilization of glycans was found to boost the sensitivity of the weak carbohydrate-lectin interactions.<sup>197</sup> In this scheme, biotinylated sialosides (glycans) are connected to the SPR gold surface via NeutrAvidin linkage to a deposited BSA layer (Fig. 1.33). The addition of an inert hydrophilic hexaethylene glycol spacer (HEG) between the glycan and the surface reportedly enables a more defined surface that has the correct orientation and flexibility necessary for interactions. Again, in this scheme the lectins are the analyte (*Sambucus nigra* agglutinin (SNA),

*Maackia amurensis* lectin (MAL), ConcanavalinA (Con A), and wheat germ agglutinin (WGA)). The resulting SPR loading curves are similar to the two previous studies, showing the characteristic loading, release, and wash steps of an automated SPR system (Fig. 1.34).



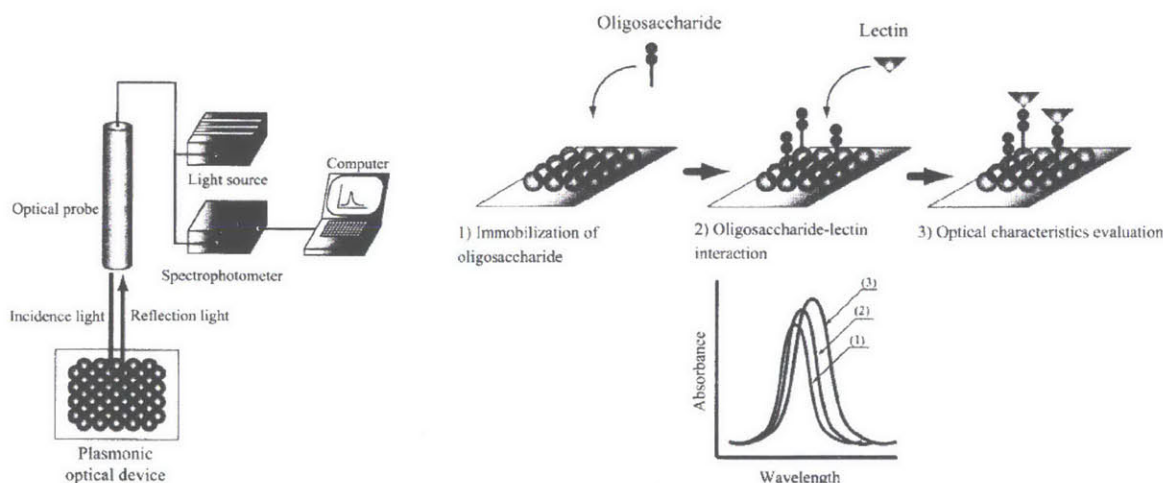
**Fig. 1.33** Use of flexible synthetic tethers to immobilize glycans to SPR surface for signal improvement. Reproduced with permission.<sup>197</sup>



**Fig. 1.34** SPR sensogram showing signal response to construction of the BSA-Biotin-NeutrAvidin-glycan surface and repeated addition/wash cycles of the lectin (SNA). Reproduced with permission.<sup>197</sup>



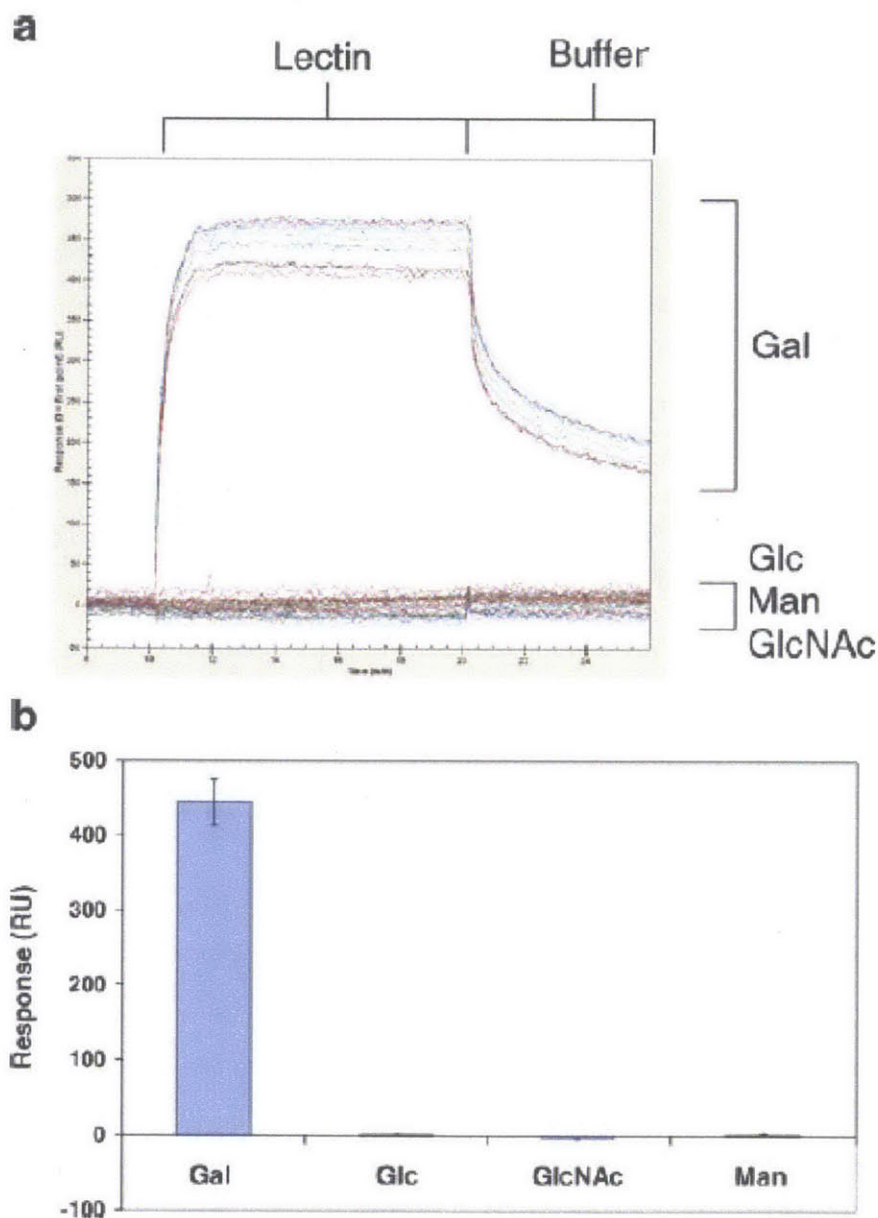
Finally, there have been two recent advances in SPR technology that are worthy of review for glycoprofiling applications. The first is the development of localized SPR readings for glycan-CRD interaction measurement.<sup>198</sup> This device is built as an array of core-shell nanoparticles deposited on a surface (Fig. 1.35). Rather than analyzing the surface plasmon over a large chip area, these devices probe the plasmon generated on nanosurfaces. The result is a simpler, more efficient device that requires fewer reagents than traditional SPR or QCM. The functionality of this device was demonstrated using immobilized LacNAc glycoside on the surface of the nanoparticles and flowing WGA lectin as the analyte. As the lectin binds to the particles, there is a distinct shift in the absorption spectra of the particle. The paper only reports the relative binding efficiencies of these new probes, but one can envision calculation of kinetic parameters from the fluorescent signal, as is done with traditional SPR. These nanoscale particles may develop into exciting arrays of unique sensors capable of reporting binding many lectin-glycan pair independently. Such a nanoarray could be used for glycoprofiling applications.



**Fig. 1.35** Concept of localized SPR for glycan-lectin interactions. The surface plasmon of nanoparticles are investigated for binding of analyte lectins to immobilized oligosaccharides. Reproduced with permission.<sup>198</sup>

A second recent advance is the commercial production of multiplexed or arrayed SPR machines. These can quantify the binding kinetics of many spots of immobilized molecules on the surface to the analytes in solution. Such devices radically reduce time and reagent costs. These machines have been used in studies of protein and nucleic acid arrays,<sup>199,200</sup> and more recently with glycan arrays.<sup>201</sup> Karamanska *et al.* used a commercially available array-SPR machine (Biacore Flexchip SPR) which can measure loading curves from a 20x20 array of independent spots on a 1 cm<sup>2</sup> chip. Each arrayed spot has an accompanying control region to monitor changes in the bulk refractive index and to correct for any instrumental drift. First, the SPR array fidelity was tested by spotting ten replicate spots of Gal, Glc, Man, and GlcNAc and interrogating them with RCA<sub>120</sub> (lectin that shows specificity for Galactose<sup>202</sup>).

The results (Fig. 1.36) reveal the predicted response to Galactose and good replication among spots. Next, 40 sialylated and non-sialylated glycans were spotted and their interactions with a recombinant Fc fusion protein (human Siglec 7 (h-Siglec7-Fc)) were assessed to determine which binding partners were the strongest. Although this study does not include any kinetic analysis, Karamanska *et al.* do acknowledge that such an array could be used for screening lectins.



**Fig. 1.36** Taken from the Karamanska *et al* study that uses an arrayed SPR tool for glycan-lectin interactions. The investigators array replicates of Gal, Glc, Man, and GlcNAc and interrogate with RCA<sub>120</sub> lectin. They see the expected response to Gal for all replicates (a) and good reproducibility of the max RU signal (b). Reproduced with permission.<sup>201</sup>

The advantages of SPR for determining kinetics of glycan-CRD pairs are: 1) a wide range of affinities can be assessed (pM to mM); 2) forward and reverse kinetic parameters can be determined (not just  $K_D$  as in fluorescent FAC); 3) only 1-5  $\mu\text{g}$  of sample is needed for immobilization; 4) the amount of analyte is approximately 0.1 to 100 times the  $K_D$  of the pair, with flow volumes of 50 to 150  $\mu\text{l}$  per test; and 5) established, automated measurement tools and packaged analysis software make for rapid determination of kinetics.<sup>20</sup> The same source also delineated some significant limitations of SPR analysis: 1) the analyte must have significant mass to transduce the change in refractive index necessary for SPR analysis, thus the more substantial lectins are routinely used as the analyte and the glycans are bound to the surface; 2) the mass-transport of the analyte to the surface may affect the measured kinetic parameters; and 3) glycan coupling directly to the gold surface is inadequate, so neoglycoproteins<sup>203</sup> (i.e. glycans synthetically bound to a protein backbone) and the more substantial lectins are used as the binding analytes. This can bias the analysis of single CRD-glycan interactions as neoglycoproteins alter the presentation and density of the glycans, a critical parameter in binding kinetics.<sup>204</sup> In terms of application to glycoprofiling, SPR tools are beginning to emerge in a microarray format, allowing one to get quantitative parameters between hundreds of glycan-CRD pairs in a single assay. This could conceivably be turned into a profiling tool for lectins, but not for glycans or small glycoproteins due to the large analyte mass constraint. However, an emerging technology that uses a different means of signal transduction may overcome this limitation.

#### **1.4.3 Optical carbon nanotube based sensor for measuring glycan-lectin kinetics**

A platform using fluorescent single walled carbon nanotubes (SWNT) for measuring kinetic parameters between glycans and lectins<sup>205</sup> is the subject of this thesis (detailed further in chapters 3-6). One of its greatest advantages is that the clear, rapid signal transduction of SPR is preserved, but the sensor platform has no mass constraints. Thus, the lectins or synthetic CRDs can be arrayed on the surface and glycans or glycoproteins can be the analytes. There is no need for labeling, and the ability to pinpoint nanoscale strong transducers will serve well in creating a nanoarray of lectins or CRDs to report kinetic parameters. Rather than having hundreds of sensors on a  $1\text{cm}^2$  chip, as emerging multiplexed SPR devices are able to do, one could have thousands of reporters on a  $1\text{mm}^2$  spot. This could radically reduce the required lectin or CRD reagent requirement and push the sensitivity of the optical array to the sub-nanomolar level, capable of probing glycosylated biomarkers in their native state and physiological concentration.

## 1.5 Summary

As carbohydrates dictate much of a protein's mass, structure, and function, accurate glycoprofiling is a major research focus both in academia and in industry. Current methods of profiling include separation techniques, like liquid chromatography or capillary electrophoresis, and atomistic determination methods, like NMR or mass spectrometry, used separately or in tandem. These methods require glycan liberation, via chemical or enzymatic methods, to ensure that the protein content does not bias the glycan analysis. Additionally, fluorescence labeling is often required to increase sensitivity. More recent microarray methods have eliminated the need for glycan liberation but still require labeling for binding detection. Additionally, the current microarray methods yield relative binding information, so their profiling resolution is limited. Many emerging nanoengineered sensors allow for quantitative measurement of glycan-CRD interactions (executive summary provided in Table 1.6). Mechanical systems include QCM and cantilever arrays which have good sensitivity but are unable to transduce the binding of low mass species. Electronic systems include electrochemical sensors, field effect transistors, and pore impedance sensors. These tools have demonstrated quantified binding of glycans and lectins but are difficult to construct and may be limited in their microarray applications. Finally, optical systems include FAC, SPR, and fluorescent SWNT. FAC has enabled the determination of the dissociation constants of many Lectin-Glycan pairs, but the most sensitive form of FAC requires fluorescent labeling. SPR is a regularly used tool due to simple, commercialized platforms and straightforward signal analysis but is limited in its ability to detect smaller glycosylated products. The SWNT-based methods developed in this thesis are not labeling or mass limited, but to date have only measured glycoprotein-lectin  $K_D$  in the  $\mu\text{M}$  range and require further work to optimize the sensor signal. In the near future one of these reviewed nanoengineered platforms, or one yet to be developed, may be incorporated into an array of quantitative, glycan-CRD sensors. With enough unique CRDs, a unique "kinetic code" could be determined for each glycan of interest. Thus profiling could occur as it does regularly in nature, without liberation or labeling of the glycans. Modeling of this profiling method is the subject of the next chapter.

**Table 1.6** – Executive summary table overviewing the nanoengineered sensors detailed in this review

<b>Sensor Type</b>	<b>Technology</b>	<b>Limitations</b>	<b>Advantages</b>
<b>Mechanical</b>	(a) Quartz crystal microbalance (QCM). (b) Cantilever arm detectors.	(a) The mass of the analyte limits the ability of the QCM. (b) The assumption that the analyte mass couples tightly to the resonator is not always the case. (c) Surface chemistry introduces non-specific adsorption. (d) Signal amplification is complicated and highly system-dependent.	(a) Label-free and liberation-free detection platform. (b) Very simple concept and operation. (c) Reasonable detection time (<30 min.). (d) Higher sensitivities could be achieved by additional signal amplifier. (e) Simple but efficient surface chemistry can be readily applied to improve the binding events.
<b>Electrical</b>	(a) Electrical impedance spectroscopy (EIS). (b) Nanotube field effect transistors (NTFETs). (c) Nanopore Sensors.	(a) Complicated assemble process of multilayer electrodes. (b) Semiconducting SWNTs with higher purity are required to achieve better signal quality in NTFETs. (c) Complicated characterization of nanopore sensors.	(a) Label-free and liberation-free detection platform. (b) Fast detection time (<15 min.). (c) Lower detection sensitivity (1 pg/mL). (d) Excellent potential for development as the handheld sensors. (e) Great potential for the detection of chiral molecules using nanopore sensors.
<b>Optical</b>	(a) Frontal affinity chromatography (FAC). (b) Surface plasmon resonance (SPR). (c) Nanotransducers based on fluorescent SWNTs.	(a) Slow analysis time and required large amount of analyte with higher purity when using FAC. (b) FAC is not a label-free platform unless coupled with mass spec. (c) The mass of the analyte limits the ability of the SPR. (d) Mass-transport of the analyte to the surface affects the measured kinetic parameters using SPR. (e) Neoglycoproteins used as the binding analytes may bias the analysis results of SPR. (f) The binding events must cause a change of fluorescence of SWNTs, which needs careful engineering design.	(a) Label-free and liberation-free detection platform except FAC. (b) FAC is capable of measuring weak interactions and the equilibrium kinetic analysis. (c) Wide range of affinities (pM to mM), and forward and reverse kinetic parameters can be determined using SPR. (d) No mass constraints with nanotransducer platform (single molecular detection is achievable). (e) Great potential for in vivo detection with nanotransducer platform.

## 1.6 References

- 1 Reuel, N. F., Mu, B., Zhang, J., Hinckley, A. & Strano, M. S. Nanoengineered glycan sensors enabling native glycoprofiling for medicinal applications: towards profiling glycoproteins without labeling or liberation steps. *Chemical Society Reviews* **41**, 5744-5779, doi:10.1039/c2cs35142k (2012).
- 2 *Transforming Glycoscience: A Roadmap for the Future*. (The National Academies Press, 2012).
- 3 Roco, M. C. Nanotechnology: convergence with modern biology and medicine. *Current Opinion in Biotechnology* **14**, 337-346, doi:10.1016/s0958-1669(03)00068-5 (2003).
- 4 Dekker, C. Solid-state nanopores. *Nature Nanotechnology* **2**, 209-215, doi:10.1038/nnano.2007.27 (2007).
- 5 Davis, M. E., Katz, A. & Ahmad, W. R. Rational catalyst design via imprinted nanostructured materials. *Chemistry of Materials* **8**, 1820-1839 (1996).
- 6 Vriezema, D. M. *et al.* Self-assembled nanoreactors. *Chemical Reviews* **105**, 1445-1489, doi:10.1021/cr0300688 (2005).
- 7 Rudkevich, D. A. Nanoscale molecular containers. *Bulletin of the Chemical Society of Japan* **75**, 393-413 (2002).
- 8 Polizu, S., Savadogo, O., Poulin, P. & Yahia, L. Applications of carbon nanotubes-based biomaterials in biomedical nanotechnology. *Journal of Nanoscience and Nanotechnology* **6**, 1883-1904, doi:10.1166/jnn.2006.197 (2006).
- 9 Randall, C. L., Leong, T. G., Bassik, N. & Gracias, D. H. 3D lithographically fabricated nanoliter containers for drug delivery. *Advanced Drug Delivery Reviews* **59**, 1547-1561, doi:10.1016/j.addr.2007.08.024 (2007).
- 10 Savic, R., Eisenberg, A. & Maysinger, D. Block copolymer micelles as delivery vehicles of hydrophobic drugs: Micelle-cell interactions. *Journal of Drug Targeting* **14**, 343-355, doi:10.1080/10611860600874538 (2006).
- 11 Tarahovsky, Y. S. "Smart" Liposomal Nanocontainers in Biology and Medicine. *Biochemistry-Moscow* **75**, 811-824, doi:10.1134/s0006297910070023 (2010).
- 12 Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry* **73**, 1019-1049, doi:10.1146/annurev.biochem.73.011303.073752 (2004).
- 13 Spiro, R. G. Role of N-linked polymannose oligosaccharides in targeting glycoproteins for endoplasmic reticulum-associated degradation. *Cellular and Molecular Life Sciences* **61**, 1025-1041, doi:10.1007/s00018-004-4037-8 (2004).
- 14 Sitia, R. & Braakman, I. Quality control in the endoplasmic reticulum protein factory. *Nature* **426**, 891-894, doi:10.1038/nature02262 (2003).
- 15 Sheridan, C. Commercial interest grows in glycan analysis. *Nature Biotechnology* **25**, 145-146, doi:10.1038/nbt0207-145 (2007).
- 16 Ito, Y., Hagihara, S., Matsuo, I. & Totani, K. Structural approaches to the study of oligosaccharides in glycoprotein quality control. *Current Opinion in Structural Biology* **15**, 481-489, doi:10.1016/j.sbi.2005.08.012 (2005).
- 17 Vanderschaeghe, D., Festjens, N., Delanghe, J. & Callewaert, N. Glycome profiling using modern glycomics technology: technical aspects and applications. *Biological Chemistry* **391**, 149-161, doi:10.1515/bc.2010.031 (2010).
- 18 Kunz, H. Emil Fischer - Unequaled classicist, master of organic chemistry research, and inspired trailblazer of biological chemistry. *Angewandte Chemie-International Edition* **41**, 4439-+ (2002).
- 19 Speiser, P. & Smekal, F. G. *Karl Landsteiner, the discoverer of the blood-groups and a pioneer in the field of immunology : biography of a Nobel Prize winner of the Vienna Medical School*. (Hollinek, 1975).
- 20 Varki, A. *Essentials of glycobiology*. 2nd edn, (Cold Spring Harbor Laboratory Press, 2009).
- 21 Rademacher, T. W., Parekh, R. B. & Dwek, R. A. GLYCOBIOLOGY. *Annual Review of Biochemistry* **57**, 785-838 (1988).

- 22 Hart, G. W. Editorial. *Glycobiology* **1**, 1, doi:10.1093/glycob/1.1.1 (1990).
- 23 Wang, L. Awards - Carolyn Bertozzi wins Lemelson-MIT Prize. *Chem. Eng. News* **88**, 2, doi:10.1021/cen-v088n024.p047 (2010).
- 24 Raman, R., Raguram, S., Venkataraman, G., Paulson, J. C. & Sasisekharan, R. Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nature Methods* **2**, 817-824, doi:10.1038/nmeth807 (2005).
- 25 Feizi, T. Progress in deciphering the information content of the 'glycome' - a crescendo in the closing years of the millennium. *Glycoconjugate Journal* **17**, 553-565 (2000).
- 26 Paulson, J. C., Blixt, O. & Collins, B. E. Sweet spots in functional glycomics. *Nature Chemical Biology* **2**, 238-248, doi:10.1038/nchembio785 (2006).
- 27 Pilobello, K. T. & Mahal, L. K. Deciphering the glycode: the complexity and analytical challenge of glycomics. *Current Opinion in Chemical Biology* **11**, 300-305, doi:10.1016/j.cbpa.2007.05.002 (2007).
- 28 Gabius, H. R. a. H.-J. in *The Sugar Code: fundamentals of glycosciences* (ed Hans-Joachim Gabius) 3-14 (Wiley-VCH 2009).
- 29 Laine, R. A. A CALCULATION OF ALL POSSIBLE OLIGOSACCHARIDE ISOMERS BOTH BRANCHED AND LINEAR YIELDS  $1.05 \times 10^{12}$  STRUCTURES FOR A REDUCING HEXASACCHARIDE - THE ISOMER-BARRIER TO DEVELOPMENT OF SINGLE-METHOD SACCHARIDE SEQUENCING OR SYNTHESIS SYSTEMS. *Glycobiology* **4**, 759-767 (1994).
- 30 Stenutz, R., Weintraub, A. & Widmalm, G. The structures of Escherichia coli O-polysaccharide antigens. *Fems Microbiology Reviews* **30**, 382-403, doi:10.1111/j.1574-6976.2006.00016.x (2006).
- 31 Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Molecular Biosystems* **5**, 1087-1104, doi:10.1039/b907931a (2009).
- 32 McNaught, A. D. in *Advances in Carbohydrate Chemistry and Biochemistry, Vol 52* Vol. 52 *Advances in Carbohydrate Chemistry and Biochemistry* 43-177 (Academic Press Inc, 1997).
- 33 Kornfeld, S., Li, E. & Tabas, I. SYNTHESIS OF COMPLEX-TYPE OLIGOSACCHARIDES .2. CHARACTERIZATION OF PROCESSING INTERMEDIATES IN SYNTHESIS OF COMPLEX OLIGOSACCHARIDE UNITS OF VESICULAR STOMATITIS VIRUS-G PROTEIN. *Journal of Biological Chemistry* **253**, 7771-7778 (1978).
- 34 Harvey, D. J. *et al.* Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics* **9**, 3796-3801, doi:10.1002/pmic.200900096 (2009).
- 35 Varki, A. *et al.* Symbol nomenclature for glycan representation. *Proteomics* **9**, 5398-5399, doi:10.1002/pmic.200900708 (2009).
- 36 Chung, C. H. *et al.* Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *New England Journal of Medicine* **358**, 1109-1117 (2008).
- 37 Qian, J. *et al.* Structural characterization of N-linked oligosaccharides on monoclonal antibody cetuximab by the combination of orthogonal matrix-assisted laser desorption/ionization hybrid quadrupole-quadrupole time-of-flight tandem mass spectrometry and sequential enzymatic digestion. *Analytical Biochemistry* **364**, 8-18, doi:10.1016/j.ab.2007.01.023 (2007).
- 38 Kawasaki, N. *et al.* The Significance of Glycosylation Analysis in Development of Biopharmaceuticals. *Biological & Pharmaceutical Bulletin* **32**, 796-800 (2009).
- 39 Rudd, P. M. & Dwek, R. A. Glycosylation: Heterogeneity and the 3D structure of proteins. *Critical Reviews in Biochemistry and Molecular Biology* **32**, 1-100, doi:10.3109/10409239709085144 (1997).
- 40 Li, H. J. & d'Anjou, M. Pharmacological significance of glycosylation in therapeutic proteins. *Current Opinion in Biotechnology* **20**, 678-684, doi:10.1016/j.copbio.2009.10.009 (2009).
- 41 Hamilton, S. R. *et al.* Production of complex human glycoproteins in yeast. *Science* **301**, 1244-1246, doi:10.1126/science.1088166 (2003).

- 42 Rich, J. R. & Withers, S. G. Emerging methods for the production of homogeneous human  
glycoproteins. *Nature Chemical Biology* **5**, 206-215, doi:10.1038/nchembio.148 (2009).
- 43 Schwarz, F. *et al.* A combined method for producing homogeneous glycoproteins with eukaryotic  
N-glycosylation. *Nature Chemical Biology* **6**, 264-266, doi:10.1038/nchembio.314 (2010).
- 44 Beck, A., Cochet, O. & Wurch, T. GlycoFi's technology to control the glycosylation of  
recombinant therapeutic proteins. *Expert Opinion on Drug Discovery* **5**, 95-111,  
doi:10.1517/17460440903413504 (2010).
- 45 Alavi, A. & Axford, J. S. Glyco-biomarkers: Potential determinants of cellular physiology and  
pathology. *Disease Markers* **25**, 193-205 (2008).
- 46 Watson, M., Rudd, P. M., Bland, M., Dwek, R. A. & Axford, J. S. Sugar printing rheumatic  
diseases - A potential method for disease differentiation using immunoglobulin G  
oligosaccharides. *Arthritis and Rheumatism* **42**, 1682-1690, doi:10.1002/1529-  
0131(199908)42:8<1682::aid-anr17>3.0.co;2-x (1999).
- 47 Zhao, J., Patwa, T. H., Lubman, D. M. & Simeone, D. M. Protein biomarkers in cancer: Natural  
glycoprotein microarray approaches. *Current Opinion in Molecular Therapeutics* **10**, 602-610  
(2008).
- 48 Saldova, R., Wormald, M. R., Dwek, R. A. & Rudd, P. M. Glycosylation changes on serum  
glycoproteins in ovarian cancer may contribute to disease pathogenesis. *Disease Markers* **25**,  
219-232 (2008).
- 49 Arnold, J. N., Saldova, R., Hamid, U. M. A. & Rudd, P. M. Evaluation of the serum N-linked  
glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **8**, 3284-3293,  
doi:10.1002/pmic.200800163 (2008).
- 50 Arnold, J. N., Wormald, M. R., Sim, R. B., Rudd, P. M. & Dwek, R. A. The impact of  
glycosylation on the biological function and structure of human immunoglobulins. *Annual Review  
of Immunology* **25**, 21-50, doi:10.1146/annurev.immunol.25.022106.141702 (2007).
- 51 Narimatsu, H. *et al.* A strategy for discovery of cancer glyco-biomarkers in serum using newly  
developed technologies for glycoproteomics. *Febs Journal* **277**, 95-105, doi:10.1111/j.1742-  
4658.2009.07430.x (2010).
- 52 Lowe, J. B. & Marth, J. D. A genetic approach to mammalian glycan function. *Annual Review of  
Biochemistry* **72**, 643-691, doi:10.1146/annurev.biochem.72.121801.161809 (2003).
- 53 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced  
from analysis of the SWISS-PROT database. *Biochimica Et Biophysica Acta-General Subjects*  
**1473**, 4-8, doi:10.1016/s0304-4165(99)00165-8 (1999).
- 54 Jung, E., Veuthey, A. L., Gasteiger, E. & Bairoch, A. Annotation of glycoproteins in the SWISS-  
PROT database. *Proteomics* **1**, 262-268, doi:10.1002/1615-9861(200102)1:2<262::aid-  
prot262>3.0.co;2-# (2001).
- 55 Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**,  
855-867, doi:10.1016/j.cell.2006.08.019 (2006).
- 56 Gamblin, D. P., Scanlan, E. M. & Davis, B. G. Glycoprotein Synthesis: An Update. *Chemical  
Reviews* **109**, 131-163, doi:10.1021/cr078291i (2009).
- 57 Mitra, N., Sharon, N. & Surolia, A. Role of n-linked glycan in the unfolding pathway of Erythrina  
coraliodendron lectin. *Biochemistry* **42**, 12208-12216, doi:10.1021/bi035169e (2003).
- 58 Seeberger, P. H. & Werz, D. B. Synthesis and medical applications of oligosaccharides. *Nature*  
**446**, 1046-1051, doi:10.1038/nature05819 (2007).
- 59 Miyamoto, S. Clinical applications of glycomic approaches for the detection of cancer and other  
diseases. *Current Opinion in Molecular Therapeutics* **8**, 507-513 (2006).
- 60 Lau, K. S. & Dennis, J. W. N-Glycans in cancer progression. *Glycobiology* **18**, 750-760,  
doi:10.1093/glycob/cwn071 (2008).
- 61 Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation. Potential for therapeutics and  
diagnostics. *Nature Reviews Drug Discovery* **4**, 477-488, doi:10.1038/nrd1751 (2005).



- 62 Marth, J. D. & Grewal, P. K. Mammalian glycosylation in immunity. *Nature Reviews*  
63 *Immunology* **8**, 874-887, doi:10.1038/nri2417 (2008).
- 64 Brockhausen, I., Schutzbach, J. & Kuhns, W. Glycoproteins and their relationship to human  
65 disease. *Acta Anatomica* **161**, 36-78 (1998).
- 66 Dennis, J. W., Granovsky, M. & Warren, C. E. Protein glycosylation in development and disease.  
67 *Bioessays* **21**, 412-421, doi:10.1002/(sici)1521-1878(199905)21:5<412::aid-bies8>3.0.co;2-5  
68 (1999).
- 69 El-Boubbou, K. & Huang, X. F. Glyco-Nanomaterials: Translating Insights from the "Sugar-  
70 Code" to Biomedical Applications. *Current Medicinal Chemistry* **18**, 2060-2078 (2011).
- 71 Krishnamoorthy, L. & Mahal, L. K. Glycomic Analysis: An Array of Technologies. *Acc*  
72 *Chemical Biology* **4**, 715-732, doi:10.1021/cb900103n (2009).
- 73 Rudd, P. M. & Dwek, R. A. Rapid, sensitive sequencing of oligosaccharides from glycoproteins.  
74 *Current Opinion in Biotechnology* **8**, 488-497, doi:10.1016/s0958-1669(97)80073-0 (1997).
- 75 Geyer, H. & Geyer, R. Strategies for analysis of glycoprotein glycosylation. *Biochimica Et*  
76 *Biophysica Acta-Proteins and Proteomics* **1764**, 1853-1869, doi:10.1016/j.bbapap.2006.10.007  
77 (2006).
- 78 Yamada, K. & Takechi, K. Recent advances in the analysis of carbohydrates for biomedical use.  
79 *Journal of Pharmaceutical and Biomedical Analysis* **55**, 702-727, doi:10.1016/j.jpba.2011.02.003  
80 (2011).
- 81 Lee, Y. C. Carbohydrate analyses with high-performance anion-exchange chromatography.  
*Journal of Chromatography A* **720**, 137-149, doi:10.1016/0021-9673(95)00222-7 (1996).
- Yagi, H. & Kato, K. Multidimensional HPLC mapping method for the structural analysis of  
anionic N-glycans. *Trends in Glycoscience and Glycotechnology* **21**, 95-104,  
doi:10.4052/tigg.21.95 (2009).
- Ruhaak, L. R., Deelder, A. M. & Wuhrer, M. Oligosaccharide analysis by graphitized carbon  
liquid chromatography-mass spectrometry. *Analytical and Bioanalytical Chemistry* **394**, 163-174,  
doi:10.1007/s00216-009-2664-5 (2009).
- Wuhrer, M., Deelder, A. M. & Hokke, C. H. Protein glycosylation analysis by liquid  
chromatography-mass spectrometry. *Journal of Chromatography B-Analytical Technologies in*  
*the Biomedical and Life Sciences* **825**, 124-133, doi:10.1016/j.jchromb.2005.01.030 (2005).
- Dell, A. & Morris, H. R. Glycoprotein structure determination mass spectrometry. *Science* **291**,  
2351-2356, doi:10.1126/science.1058890 (2001).
- Morris, H. R. *et al.* ANTIFREEZE GLYCOPROTEINS FROM BLOOD OF AN ANTARCTIC  
FISH - STRUCTURE OF PROLINE-CONTAINING GLYCOPEPTIDES. *Journal of Biological*  
*Chemistry* **253**, 5155-5162 (1978).
- Wuhrer, M., Catalina, M. I., Deelder, A. M. & Hokke, C. H. Glycoproteomics based on tandem  
mass spectrometry of glycopeptides. *Journal of Chromatography B-Analytical Technologies in*  
*the Biomedical and Life Sciences* **849**, 115-128, doi:10.1016/j.jchromb.2006.09.041 (2007).
- Zamfir, A. & Peter-Katalinic, J. Capillary electrophoresis-mass spectrometry for glycoscreening  
in biomedical research. *Electrophoresis* **25**, 1949-1963, doi:10.1002/elps.200405825 (2004).
- Amon, S., Zamfir, A. D. & Rizzi, A. Glycosylation analysis of glycoproteins and proteoglycans  
using capillary electrophoresis-mass spectrometry strategies. *Electrophoresis* **29**, 2485-2507,  
doi:10.1002/elps.200800105 (2008).
- Bao, Y. W. & Newburg, D. S. Capillary electrophoresis of acidic oligosaccharides from human  
milk. *Electrophoresis* **29**, 2508-2515, doi:10.1002/elps.200700873 (2008).
- Starr, C. M., Masada, R. I., Hague, C., Skop, E. & Klock, J. C. Fluorophore-assisted carbohydrate  
electrophoresis in the separation, analysis, and sequencing of carbohydrates. *Journal of*  
*Chromatography A* **720**, 295-321, doi:10.1016/0021-9673(95)00749-0 (1996).
- Harvey, D. J. ANALYSIS OF CARBOHYDRATES AND GLYCOCONJUGATES BY  
MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRY: AN

- UPDATE FOR THE PERIOD 2005-2006. *Mass Spectrometry Reviews* **30**, 1-100, doi:10.1002/mas.20265 (2011).
- 82 Harvey, D. J. ANALYSIS OF CARBOHYDRATES AND GLYCOCONJUGATES BY MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRY: AN UPDATE FOR 2003-2004. *Mass Spectrometry Reviews* **28**, 273-361, doi:10.1002/mas.20192 (2009).
- 83 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update covering the period 2001-2002. *Mass Spectrometry Reviews* **27**, 125-201, doi:10.1002/mas.20157 (2008).
- 84 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update covering the period 1999-2000. *Mass Spectrometry Reviews* **25**, 595-662, doi:10.1002/mas.20080 (2006).
- 85 Harvey, D. J. Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *Proteomics* **5**, 1774-1786, doi:10.1002/pmic.200401248 (2005).
- 86 Harvey, D. J. Proteomic analysis of glycosylation: structural determination of N- and O-linked glycans by mass spectrometry. *Expert Review of Proteomics* **2**, 87-101, doi:10.1586/14789450.2.1.87 (2005).
- 87 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates and glycoconjugates. *International Journal of Mass Spectrometry* **226**, 1-35, doi:10.1016/s1387-3806(02)00968-5 (2003).
- 88 Harvey, D. J. Identification of protein-bound carbohydrates by mass spectrometry. *Proteomics* **1**, 311-328, doi:10.1002/1615-9861(200102)1:2<311::aid-prot311>3.3.co;2-a (2001).
- 89 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrometry Reviews* **18**, 349-450, doi:10.1002/(sici)1098-2787(1999)18:6<349::aid-mas1>3.3.co;2-8 (1999).
- 90 Lazar, I. M., Lazar, A. C., Cortes, D. F. & Kabulski, J. L. Recent advances in the MS analysis of glycoproteins: Theoretical considerations. *Electrophoresis* **32**, 3-13, doi:10.1002/elps.201000393 (2011).
- 91 Cortes, D. F., Kabulski, J. L., Lazar, A. C. & Lazar, I. M. Recent advances in the MS analysis of glycoproteins: Capillary and microfluidic workflows. *Electrophoresis* **32**, 14-29, doi:10.1002/elps.201000394 (2011).
- 92 Manzi, A. E. *et al.* Exploring the glycan repertoire of genetically modified mice by isolation and profiling of the major glycan classes and nano-NMR analysis of glycan mixtures. *Glycobiology* **10**, 669-689, doi:10.1093/glycob/10.7.669 (2000).
- 93 Muller, D., Domon, B., Karas, M., Vanoostrum, J. & Richter, W. J. CHARACTERIZATION AND DIRECT GLYCOFORM PROFILING OF A HYBRID PLASMINOGEN-ACTIVATOR BY MATRIX-ASSISTED LASER-DESORPTION AND ELECTROSPRAY MASS-SPECTROMETRY - CORRELATION WITH HIGH-PERFORMANCE LIQUID-CHROMATOGRAPHIC AND NUCLEAR-MAGNETIC-RESONANCE ANALYSES OF THE RELEASED GLYCANS. *Biological Mass Spectrometry* **23**, 330-338, doi:10.1002/bms.1200230606 (1994).
- 94 Hsu, K. L. & Mahal, L. K. Sweet tasting chips: microarray-based analysis of glycans. *Current Opinion in Chemical Biology* **13**, 427-432, doi:10.1016/j.cbpa.2009.07.013 (2009).
- 95 Tateno, H. K., A., Hirabayashi, J. in *The Sugar Code: Fundamentals of Glycosciences* (ed H.J. Gabius) 247-260 (Wiley-VCH, 2009).
- 96 Gabius, H. J., Siebert, H. C., Andre, S., Jimenez-Barbero, J. & Rudiger, H. Chemical biology of the sugar code. *Chembiochem* **5**, 740-764, doi:10.1002/cbic.200300753 (2004).
- 97 Gabius, H. J., Andre, S., Jimenez-Barbero, J., Romero, A. & Solis, D. From lectin structure to functional glycomics: principles of the sugar code. *Trends in Biochemical Sciences* **36**, 298-313, doi:10.1016/j.tibs.2011.01.005 (2011).

- 98 Hirabayashi, J. Concept, strategy and realization of lectin-based glycan profiling. *Journal of Biochemistry* **144**, 139-147, doi:10.1093/jb/mvn043 (2008).
- 99 Angeloni, S. *et al.* Glycoprofiling with micro-arrays of glycoconjugates and lectins. *Glycobiology* **15**, 31-41, doi:10.1093/glycob/cwh143 (2005).
- 100 Pilobello, K. T., Krishnamoorthy, L., Slawek, D. & Mahal, L. K. Development of a lectin microarray for the rapid analysis of protein glycopatterns. *Chembiochem* **6**, 985-989, doi:10.1002/cbic.200400403 (2005).
- 101 Rosenfeld, R. *et al.* A lectin array-based methodology for the analysis of protein glycosylation. *Journal of Biochemical and Biophysical Methods* **70**, 415-426, doi:10.1016/j.jbbm.2006.09.008 (2007).
- 102 Abbott, K. L. & Pierce, J. M. in *Methods in Enzymology, Vol 480: Glycobiology* Vol. 480 *Methods in Enzymology* (ed M. Fukuda) 461-476 (2010).
- 103 Zheng, T., Peelen, D. & Smith, L. M. Lectin arrays for profiling cell surface carbohydrate expression. *Journal of the American Chemical Society* **127**, 9982-9983, doi:10.1021/ja0505550 (2005).
- 104 Tao, S. C. *et al.* Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers. *Glycobiology* **18**, 761-769, doi:10.1093/glycob/cwn063 (2008).
- 105 Yue, T. T. *et al.* The Prevalence and Nature of Glycan Alterations on Specific Proteins in Pancreatic Cancer Patients Revealed Using Antibody-Lectin Sandwich Arrays. *Molecular & Cellular Proteomics* **8**, 1697-1707, doi:10.1074/mcp.M900135-MCP200 (2009).
- 106 Gregg, K. J., Finn, R., Abbott, D. W. & Boraston, A. B. Divergent modes of glycan recognition by a new family of carbohydrate-binding modules. *Journal of Biological Chemistry* **283**, 12604-12613, doi:10.1074/jbc.M709865200 (2008).
- 107 Betanzos, C. M. *et al.* Bacterial Glycoprofiling by Using Random Sequence Peptide Microarrays. *Chembiochem* **10**, 877-888, doi:10.1002/cbic.200800716 (2009).
- 108 Chen, S. M. *et al.* Multiplexed analysis of glycan variation on native proteins captured by antibody microarrays. *Nature Methods* **4**, 437-444, doi:10.1038/nmeth1035 (2007).
- 109 Hirabayashi, J., Kuno, A. & Tateno, H. Lectin-based structural glycomics: A practical approach to complex glycans. *Electrophoresis* **32**, 1118-1128, doi:10.1002/elps.201000650 (2011).
- 110 Propheter, D. C. & Mahal, L. K. Orientation of GST-tagged lectins via in situ surface modification to create an expanded lectin microarray for glycomic analysis. *Molecular Biosystems* **7**, 2114-2117, doi:10.1039/c1mb05047h (2011).
- 111 Sauerbrey, G. VERWENDUNG VON SCHWINGQUARZEN ZUR WAGUNG DUNNER SCHICHTEN UND ZUR MIKROWAGUNG. *Zeitschrift Fur Physik* **155**, 206-222, doi:10.1007/bf01337937 (1959).
- 112 Janshoff, A., Galla, H.-J. & Steinem, C. Piezoelectric Mass-Sensing Devices as Biosensors—An Alternative to Optical Biosensors? *Angewandte Chemie International Edition* **39**, 4004-4032, doi:10.1002/1521-3773(20001117)39:22<4004::aid-anie4004>3.0.co;2-2 (2000).
- 113 Yakovleva, M. E., Safina, G. R. & Danielsson, B. A study of glycoprotein-lectin interactions using quartz crystal microbalance. *Analytica Chimica Acta* **668**, 80-85, doi:10.1016/j.aca.2009.12.004 (2010).
- 114 Gooding, J. J., Mearns, F., Yang, W. & Liu, J. Self-Assembled Monolayers into the 21st Century: Recent Advances and Applications. *Electroanalysis* **15**, 81-96, doi:10.1002/elan.200390017 (2003).
- 115 Mrksich, M. A surface chemistry approach to studying cell adhesion. *Chemical Society Reviews* **29**, 267-273 (2000).
- 116 Zhang, Y. *et al.* Carbohydrate-Protein Interactions by “Clicked” Carbohydrate Self-Assembled Monolayers. *Analytical Chemistry* **78**, 2001-2008, doi:10.1021/ac051919+ (2006).
- 117 Devaraj, N. K. *et al.* Chemoselective Covalent Coupling of Oligonucleotide Probes to Self-Assembled Monolayers. *Journal of the American Chemical Society* **127**, 8600-8601, doi:10.1021/ja0514621 (2005).

- 118 Pedroso, M. M., Watanabe, A. M., Roque-Barreira, M. C., Bueno, P. R. & Faria, R. C. Quartz  
Crystal Microbalance monitoring the real-time binding of lectin with carbohydrate with high and  
low molecular mass. *Microchemical Journal* **89**, 153-158, doi:10.1016/j.microc.2008.02.001  
(2008).
- 119 Lebed, K., Kulik, A. J., Forró, L. & Lekka, M. Lectin-carbohydrate affinity measured using a  
quartz crystal microbalance. *Journal of Colloid and Interface Science* **299**, 41-48,  
doi:10.1016/j.jcis.2006.01.053 (2006).
- 120 Pesquero, N. C. *et al.* Real-time monitoring and kinetic parameter estimation of the affinity  
interaction of jArtinM and rArtinM with peroxidase glycoprotein by the electrogravimetric  
technique. *Biosensors & Bioelectronics* **26**, 36-42, doi:10.1016/j.bios.2010.04.047 (2010).
- 121 Pei, Z. C., Anderson, H., Aastrup, T. & Ramstrom, O. Study of real-time lectin-carbohydrate  
interactions on the surface of a quartz crystal microbalance. *Biosensors & Bioelectronics* **21**, 60-  
66, doi:10.1016/j.bios.2004.10.006 (2005).
- 122 Shen, Z. *et al.* Nonlabeled Quartz Crystal Microbalance Biosensor for Bacterial Detection Using  
Carbohydrate and Lectin Recognitions. *Analytical Chemistry* **79**, 2312-2319,  
doi:10.1021/ac061986j (2007).
- 123 Mahon, E., Aastrup, T. & Barboiu, M. Multivalent recognition of lectins by glyconanoparticle  
systems. *Chemical Communications* **46**, 5491-5493 (2010).
- 124 Makky, A. *et al.* Evaluation of the Specific Interactions between Glycodendrimeric Porphyrins,  
Free or Incorporated into Liposomes, and Concanavaline A by Fluorescence Spectroscopy,  
Surface Pressure, and QCM-D Measurements. *Langmuir* **26**, 12761-12768,  
doi:10.1021/la101260t (2010).
- 125 Stephan, H., Boeloeni, R., Eggert, A., Bornfeld, N. & Schueler, A. Photodynamic therapy in  
retinoblastoma: Effects of verteporfin on retinoblastoma cell lines. *Investigative Ophthalmology  
& Visual Science* **49**, 3158-3163, doi:10.1167/iovs.07-1016 (2008).
- 126 Vanhillegersberg, R., Kort, W. J. & Wilson, J. H. P. CURRENT STATUS OF  
PHOTODYNAMIC THERAPY IN ONCOLOGY. *Drugs* **48**, 510-527 (1994).
- 127 Ballut, S. *et al.* New strategy for targeting of photosensitizers. Synthesis of glycodendrimeric  
phenylporphyrins, incorporation into a liposome membrane and interaction with a specific lectin.  
*Chemical Communications*, 224-226 (2009).
- 128 ZhangJ *et al.* Rapid and label-free nanomechanical detection of biomarker transcripts in human  
RNA. *Nat Nano* **1**, 214-220 (2006).
- 129 Gruber, K. *et al.* Cantilever Array Sensors Detect Specific Carbohydrate-Protein Interactions  
with Picomolar Sensitivity. *ACS Nano* **5**, 3670-3678, doi:10.1021/nn103626q (2011).
- 130 Sumner, J. B., Gralén, N. & Eriksson-Quensel, I.-B. THE MOLECULAR WEIGHTS OF  
CANAVALLIN, CONCANAVALIN A, AND CONCANAVALIN B. *Journal of Biological  
Chemistry* **125**, 45-48 (1938).
- 131 Suni, II. Impedance methods for electrochemical sensors using nanomaterials. *Trac-Trends in  
Analytical Chemistry* **27**, 604-611, doi:10.1016/j.trac.2008.03.012 (2008).
- 132 Yanez-Sedeno, P., Riu, J., Pingarron, J. M. & Rius, F. X. Electrochemical sensing based on  
carbon nanotubes. *Trac-Trends in Analytical Chemistry* **29**, 939-953,  
doi:10.1016/j.trac.2010.06.006 (2010).
- 133 Vedala, H. *et al.* Nanoelectronic Detection of Lectin-Carbohydrate Interactions Using Carbon  
Nanotubes. *Nano Letters* **11**, 170-175, doi:10.1021/nl103286k (2011).
- 134 Sudibya, H. G. *et al.* Interfacing Glycosylated Carbon-Nanotube-Network Devices with Living  
Cells to Detect Dynamic Secretion of Biomolecules. *Angewandte Chemie-International Edition*  
**48**, 2723-2726, doi:10.1002/anie.200805514 (2009).
- 135 Star, A., Gabriel, J. C. P., Bradley, K. & Gruner, G. Electronic detection of specific protein  
binding using nanotube FET devices. *Nano Letters* **3**, 459-463, doi:10.1021/nl0340172 (2003).
- 136 Heller, I. *et al.* Identifying the mechanism of biosensing with carbon nanotube transistors. *Nano  
Letters* **8**, 591-595, doi:10.1021/nl072996i (2008).

- 137 Cui, Y., Wei, Q. Q., Park, H. K. & Lieber, C. M. Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. *Science* **293**, 1289-1292, doi:10.1126/science.1062711 (2001).
- 138 Tans, S. J., Verschueren, A. R. M. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49-52 (1998).
- 139 Li, W. W. *et al.* Tuning the Cavity of Cyclodextrins: Altered Sugar Adaptors in Protein Pores. *Journal of the American Chemical Society* **133**, 1987-2001, doi:10.1021/ja1100867 (2011).
- 140 Jacobs, C. B., Peairs, M. J. & Venton, B. J. Review: Carbon nanotube based electrochemical sensors for biomolecules. *Analytica Chimica Acta* **662**, 105-127, doi:10.1016/j.aca.2010.01.009 (2010).
- 141 Trojanowicz, M. Analytical applications of carbon nanotubes: a review. *Trac-Trends in Analytical Chemistry* **25**, 480-489, doi:10.1016/j.trac.2005.11.008 (2006).
- 142 Loaiza, O. A. *et al.* Nanostructured Disposable Impedimetric Sensors as Tools for Specific Biomolecular Interactions: Sensitive Recognition of Concanavalin A. *Analytical Chemistry* **83**, 2987-2995, doi:10.1021/ac103108m (2011).
- 143 Nagaraj, V. J. *et al.* Nano Monitor: a miniature electronic biosensor for glycan biomarker detection. *Nanomedicine* **5**, 369-378, doi:10.2217/nnm.10.11 (2010).
- 144 Chen, S. H. *et al.* Electrochemical sensing platform based on tris(2,2'-bipyridyl)cobalt(III) and multiwall carbon nanotubes-Nafion composite for immunoassay of carcinoma antigen-125. *Electrochimica Acta* **54**, 7242-7247, doi:10.1016/j.electacta.2009.07.035 (2009).
- 145 Bhattacharya, M., Hong, S., Lee, D., Cui, T. & Goyal, S. M. Carbon nanotube based sensors for the detection of viruses. *Sensors and Actuators B: Chemical* **155**, 67-74 (2011).
- 146 La Belle, J. T., Demirok, U. K., Patel, D. R. & Cook, C. B. Development of a novel single sensor multiplexed marker assay. *Analyst* **136**, 1496-1501, doi:10.1039/c0an00923g (2011).
- 147 Zhang, J. J., Cheng, F. F., Zheng, T. T. & Zhu, J. J. Design and Implementation of Electrochemical Cytosensor for Evaluation of Cell Surface Carbohydrate and Glycoprotein. *Analytical Chemistry* **82**, 3547-3555, doi:10.1021/ac9026127 (2010).
- 148 Xue, Y. D. *et al.* Noncovalent functionalization of carbon nanotubes with lectin for label-free dynamic monitoring of cell-surface glycan expression. *Analytical Biochemistry* **410**, 92-97, doi:10.1016/j.ab.2010.11.019 (2011).
- 149 Haddad, R., Cosnier, S., Maaref, A. & Holzinger, M. Non-covalent biofunctionalization of single-walled carbon nanotubes via biotin attachment by pi-stacking interactions and pyrrole polymerization. *Analyst* **134**, 2412-2418, doi:10.1039/b916774a (2009).
- 150 Zelada-Guillen, G. A., Riu, J., Duzgun, A. & Rius, F. X. Immediate Detection of Living Bacteria at Ultralow Concentrations Using a Carbon Nanotube Based Potentiometric Aptasensor. *Angewandte Chemie-International Edition* **48**, 7334-7337, doi:10.1002/anie.200902090 (2009).
- 151 Liu, S. *et al.* Chemical functionalization of single-walled carbon nanotube field-effect transistors as switches and sensors. *Coordination Chemistry Reviews* **254**, 1101-1116, doi:10.1016/j.ccr.2009.11.007 (2010).
- 152 Kauffman, D. R. & Star, A. Electronically monitoring biological interactions with carbon nanotube field-effect transistors. *Chemical Society Reviews* **37**, 1197-1206, doi:10.1039/b709567h (2008).
- 153 Allen, B. L., Kichambare, P. D. & Star, A. Carbon nanotube field-effect-transistor-based biosensors. *Advanced Materials* **19**, 1439-1451, doi:10.1002/adma.200602043 (2007).
- 154 Besteman, K., Lee, J. O., Wiertz, F. G. M., Heering, H. A. & Dekker, C. Enzyme-coated carbon nanotubes as single-molecule biosensors. *Nano Letters* **3**, 727-730, doi:10.1021/nl034139u (2003).
- 155 Artyukhin, A. B. *et al.* Controlled electrostatic gating of carbon nanotube FET devices. *Nano Letters* **6**, 2080-2085, doi:10.1021/nl061343j (2006).

- 156 Boussaad, S., Tao, N. J., Zhang, R., Hopson, T. & Nagahara, L. A. In situ detection of  
cytochrome c adsorption with single walled carbon nanotube device. *Chemical Communications*,  
1502-1503, doi:10.1039/b302681g (2003).
- 157 Gui, E. L. *et al.* DNA sensing by field-effect transistors based on networks of carbon nanotubes.  
*Journal of the American Chemical Society* **129**, 14427-14432, doi:10.1021/ja075176g (2007).
- 158 Hecht, D. S. *et al.* Bioinspired detection of light using a porphyrin-sensitized single-wall  
nanotube field effect transistor. *Nano Letters* **6**, 2031-2036, doi:10.1021/nl061231s (2006).
- 159 Maroto, A., Balasubramanian, K., Burghard, M. & Kern, K. Functionalized metallic carbon  
nanotube devices for pH sensing. *Chemphyschem* **8**, 220-223, doi:10.1002/cphc.200600498  
(2007).
- 160 Bayley, H. & Cremer, P. S. Stochastic sensors inspired by biology. *Nature* **413**, 226-230,  
doi:10.1038/35093038 (2001).
- 161 Gu, L. Q., Braha, O., Conlan, S., Cheley, S. & Bayley, H. Stochastic sensing of organic analytes  
by a pore-forming protein containing a molecular adapter. *Nature* **398**, 686-690 (1999).
- 162 Knibbs, R. N., Osborne, S. E., Glick, G. D. & Goldstein, I. J. BINDING DETERMINANTS OF  
THE SIALIC ACID-SPECIFIC LECTIN FROM THE SLUG LIMAX-FLAVUS. *Journal of  
Biological Chemistry* **268**, 18524-18531 (1993).
- 163 Blixt, O., Collins, B. E., van den Nieuwenhof, I. M., Crocker, P. R. & Paulson, J. C. Sialoside  
specificity of the siglec family assessed using novel multivalent probes - Identification of potent  
inhibitors of myelin-associated glycoprotein. *Journal of Biological Chemistry* **278**, 31007-31019,  
doi:10.1074/jbc.M304331200 (2003).
- 164 Collins, B. E. *et al.* High-affinity ligand probes of CD22 overcome the threshold set by cis  
ligands for binding, endocytosis, and killing of B cells. *Journal of Immunology* **177**, 2994-3003  
(2006).
- 165 Yamaji, T., Teranishi, T., Alpey, M. S., Crocker, P. R. & Hashimoto, Y. A small region of the  
natural killer cell receptor, Siglec-7, is responsible for its preferred binding to alpha 2,8-disialyl  
and branched alpha 2,6-sialyl residues - A comparison with Siglec-9. *Journal of Biological  
Chemistry* **277**, 6324-6332, doi:10.1074/jbc.M110146200 (2002).
- 166 Kasai, K., Oda, Y., Nishikata, M. & Ishii, S. FRONTAL AFFINITY-CHROMATOGRAPHY -  
THEORY FOR ITS APPLICATION TO STUDIES ON SPECIFIC INTERACTIONS OF  
BIOMOLECULES. *Journal of Chromatography* **376**, 33-47, doi:10.1016/s0378-4347(00)80822-  
1 (1986).
- 167 Hirabayashi, J., Arata, Y. & Kasai, K. Reinforcement of frontal affinity chromatography for  
effective analysis of lectin-oligosaccharide interactions. *Journal of Chromatography A* **890**, 261-  
271, doi:10.1016/s0021-9673(00)00545-8 (2000).
- 168 Ng, E. S. M., Chan, N. W. C., Lewis, D. F., Hindsgaul, O. & Schriemer, D. C. Frontal affinity  
chromatography-mass spectrometry. *Nature Protocols* **2**, 1907-1917, doi:10.1038/nprot.2007.262  
(2007).
- 169 Rempel, B. P., Winter, H. C., Goldstein, I. J. & Hindsgaul, O. Characterization of the recognition  
of blood group B trisaccharide derivatives by the lectin from *Marasmius oreades* using frontal  
affinity chromatography-mass spectrometry. *Glycoconjugate Journal* **19**, 175-180,  
doi:10.1023/a:1024297623445 (2002).
- 170 Tateno, H., Nakamura-Tsuruta, S. & Hirabayashi, J. Frontal affinity chromatography: sugar-  
protein interactions. *Nature Protocols* **2**, 2529-2537, doi:10.1038/nprot.2007.357 (2007).
- 171 Iwaki, J. *et al.* The Gal beta-(syn)-gauche configuration is required for galectin-recognition  
disaccharides. *Biochimica Et Biophysica Acta-General Subjects* **1810**, 643-651,  
doi:10.1016/j.bbagen.2011.04.001 (2011).
- 172 Takeuchi, T., Hayama, K., Hirabayashi, J. & Kasai, K. *Caenorhabditis elegans* N-glycans  
containing a Gal-Fuc disaccharide unit linked to the innermost GlcNAc residue are recognized by  
*C. elegans* galectin LEC-6. *Glycobiology* **18**, 882-890, doi:10.1093/glycob/cwn077 (2008).

- 173 Miyanishi, N. *et al.* Carbohydrate-recognition domains of galectin-9 are involved in intermolecular interaction with galectin-9 itself and other members of the galectin family. *Glycobiology* **17**, 423-432, doi:10.1093/glycob/cwm001 (2007).
- 174 Yabe, R., Tateno, H. & Hirabayashi, J. Frontal affinity chromatography analysis of constructs of DC-SIGN, DC-SIGNR and LSECtin extend evidence for affinity to agalactosylated N-glycans. *Febs Journal* **277**, 4010-4026, doi:10.1111/j.1742-4658.2010.07792.x (2010).
- 175 Mikami, K. *et al.* The sugar-binding ability of human OS-9 and its involvement in ER-associated degradation. *Glycobiology* **20**, 310-321, doi:10.1093/glycob/cwp175 (2010).
- 176 Yamamoto, K. Intracellular Lectins Involved in Folding and Transport in the Endoplasmic Reticulum. *Biological & Pharmaceutical Bulletin* **32**, 767-773 (2009).
- 177 Quan, E. M. *et al.* Defining the Glycan Destruction Signal for Endoplasmic Reticulum-Associated Degradation. *Molecular Cell* **32**, 870-877, doi:10.1016/j.molcel.2008.11.017 (2008).
- 178 Tanaka, H. *et al.* Mechanism by which the lectin actinohivin blocks HIV infection of target cells. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15633-15638, doi:10.1073/pnas.0907572106 (2009).
- 179 Isomura, R., Kitajima, K. & Sato, C. Structural and Functional Impairments of Polysialic Acid by a Mutated Polysialyltransferase Found in Schizophrenia. *Journal of Biological Chemistry* **286**, 21535-21545, doi:10.1074/jbc.M111.221143 (2011).
- 180 Kawsar, S. M. A. *et al.* Glycan-binding profile of a D-galactose binding lectin purified from the annelid, *Perinereis nuntia* ver. *vallata*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **152**, 382-389, doi:10.1016/j.cbpb.2009.01.009 (2009).
- 181 Kawsar, S. M. A. *et al.* Glycan-Binding Profile and Cell Adhesion Activity of American Bullfrog (*Rana catesbeiana*) Oocyte Galectin-1. *Protein and Peptide Letters* **16**, 677-684 (2009).
- 182 Kawsar, S. M. A. *et al.* Isolation, purification, characterization and glycan-binding profile of a D-galactoside specific lectin from the marine sponge, *Halichondria okadai*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **150**, 349-357, doi:10.1016/j.cbpb.2008.04.004 (2008).
- 183 Tateno, H., Nakamura-Tsuruta, S. & Hirabayashi, J. Comparative analysis of core-fucose-binding lectins from *Lens culinaris* and *Pisum sativum* using frontal affinity chromatography. *Glycobiology* **19**, 527-536, doi:10.1093/glycob/cwp016 (2009).
- 184 Nakakita, S. I. *et al.* Gas-phase pyridylamination of saccharides: Development and applications. *Analytical Chemistry* **79**, 2674-2679, doi:10.1021/ac0700878 (2007).
- 185 Liedberg, B., Nylander, C. & Lundstrom, I. SURFACE-PLASMON RESONANCE FOR GAS-DETECTION AND BIOSENSING. *Sensors and Actuators* **4**, 299-304, doi:10.1016/0250-6874(83)85036-7 (1983).
- 186 Cooper, M. A. Label-free screening of bio-molecular interactions. *Analytical and Bioanalytical Chemistry* **377**, 834-842, doi:10.1007/s00216-003-2111-y (2003).
- 187 Homola, J., Yee, S. S. & Gauglitz, G. Surface plasmon resonance sensors: review. *Sensors and Actuators B-Chemical* **54**, 3-15, doi:10.1016/s0925-4005(98)00321-9 (1999).
- 188 Larsen, K., Thygesen, M. B., Guillaumie, F., Willats, W. G. T. & Jensen, K. J. Solid-phase chemical tools for glycobiology. *Carbohydrate Research* **341**, 1209-1234, doi:10.1016/j.carres.2006.04.045 (2006).
- 189 Sato, C., Yamakawa, N. & Kitajima, K. in *Methods in Enzymology, Vol 478: Glycomics* Vol. 478 *Methods in Enzymology* (ed M. Fukuda) 219-232 (2010).
- 190 Oshannessy, D. J., Brighamburke, M., Soneson, K. K., Hensley, P. & Brooks, I. Determination of rate and equilibrium binding constants for macromolecular interactions using surface-plasmon resonance: use of non-linear least-squares analysis methods *Analytical Biochemistry* **212**, 457-468 (1993).
- 191 Morton, T. A., Myszka, D. G. & Chaiken, I. M. INTERPRETING COMPLEX BINDING-KINETICS FROM OPTICAL BIOSENSORS - A COMPARISON OF ANALYSIS BY

- LINEARIZATION, THE INTEGRATED RATE-EQUATION, AND NUMERICAL-  
INTEGRATION. *Analytical Biochemistry* **227**, 176-185, doi:10.1006/abio.1995.1268 (1995).
- 192 Malmborg, A. C., Michaelsson, A., Ohlin, M., Jansson, B. & Borrebaeck, C. A. K. REAL-TIME  
ANALYSIS OF ANTIBODY ANTIGEN REACTION-KINETICS. *Scandinavian Journal of*  
*Immunology* **35**, 643-650, doi:10.1111/j.1365-3083.1992.tb02970.x (1992).
- 193 Karlsson, R., Michaelsson, A. & Mattsson, L. KINETIC-ANALYSIS OF MONOCLONAL  
ANTIBODY-ANTIGEN INTERACTIONS WITH A NEW BIOSENSOR BASED  
ANALYTICAL SYSTEM. *Journal of Immunological Methods* **145**, 229-240, doi:10.1016/0022-  
1759(91)90331-9 (1991).
- 194 Fagerstam, L. G., Frostelkarlsson, A., Karlsson, R., Persson, B. & Ronnberg, I. BIOSPECIFIC  
INTERACTION ANALYSIS USING SURFACE-PLASMON RESONANCE DETECTION  
APPLIED TO KINETIC, BINDING-SITE AND CONCENTRATION ANALYSIS. *Journal of*  
*Chromatography* **597**, 397-410, doi:10.1016/0021-9673(92)80137-j (1992).
- 195 Mann, D. A., Kanai, M., Maly, D. J. & Kiessling, L. L. Probing low affinity and multivalent  
interactions with surface plasmon resonance: Ligands for concanavalin A. *Journal of the*  
*American Chemical Society* **120**, 10575-10582, doi:10.1021/ja9818506 (1998).
- 196 Schlick, K. H. & Cloninger, M. J. Inhibition binding studies of glycodendrimer/lectin interactions  
using surface plasmon resonance. *Tetrahedron* **66**, 5305-5310, doi:10.1016/j.tet.2010.05.038  
(2010).
- 197 Linman, M. J., Taylor, J. D., Yu, H., Chen, X. & Cheng, Q. Surface plasmon resonance study of  
protein-carbohydrate interactions using biotinylated sialosides. *Analytical Chemistry* **80**, 4007-  
4013, doi:10.1021/ac702566e (2008).
- 198 Endo, T. *et al.* Label-Free Detection of Oligosaccharide-Lectin Interaction Using Plasmonic  
Optical Device for Glycomics Application. *Sensors and Materials* **23**, 135-146 (2011).
- 199 Boozer, C., Kim, G., Cong, S. X., Guan, H. W. & Londergan, T. Looking towards label-free  
biomolecular interaction analysis in a high-throughput format: a review of new surface plasmon  
resonance technologies. *Current Opinion in Biotechnology* **17**, 400-405,  
doi:10.1016/j.copbio.2006.06.012 (2006).
- 200 Cooper, M. A. Optical biosensors: where next and how soon? *Drug Discovery Today* **11**, 1061-  
1067, doi:10.1016/j.drudis.2006.10.003 (2006).
- 201 Karamanska, R. *et al.* Surface plasmon resonance imaging for real-time, label-free analysis of  
protein interactions with carbohydrate microarrays. *Glycoconjugate Journal* **25**, 69-74,  
doi:10.1007/s10719-007-9047-y (2008).
- 202 Manimala, J. C., Roach, T. A., Li, Z. T. & Gildersleeve, J. C. High-throughput carbohydrate  
microarray analysis of 24 lectins. *Angewandte Chemie-International Edition* **45**, 3607-3610,  
doi:10.1002/anie.200600591 (2006).
- 203 Zhang, Y. L., Li, Q. A., Rodriguez, L. G. & Gildersleeve, J. C. An Array-Based Method To  
Identify Multivalent Inhibitors. *Journal of the American Chemical Society* **132**, 9653-9662,  
doi:10.1021/ja100608w (2010).
- 204 Oyelaran, O., Li, Q., Farnsworth, D. & Gildersleeve, J. C. Microarrays with Varying  
Carbohydrate Density Reveal Distinct Subpopulations of Serum Antibodies. *Journal of Proteome*  
*Research* **8**, 3529-3538, doi:10.1021/pr9002245 (2009).
- 205 Reuel, N. F. *et al.* Transduction of Glycan-Lectin Binding Using Near-Infrared Fluorescent  
Single-Walled Carbon Nanotubes for Glycan Profiling. *Journal of the American Chemical*  
*Society* **133**, 17923-17933, doi:10.1021/ja2074938 (2011).



## 2. Modeling Native Glycoprofiling with Glycan-CRD Sensors

Some of the work, text and figures presented in this chapter are reprinted or adapted from Reuel, N. F., Mu, B., Zhang, J., Hinckley, A. & Strano, M. S. “Nanoengineered glycan sensors enabling native glycoprofiling for medicinal applications: towards profiling glycoproteins without labeling or liberation steps.” *Chemical Society Reviews* 41, 5744-5779, (2012) - (reproduced under thesis use allowance from The Royal Society of Chemistry)

### 2.1 Overview

Lectin microarrays have been proposed as a platform for profiling glycoproteins, where each lectin has a few, highly-specific glycan pairs. It may be possible to profile using a much smaller set of lectins, if the entire affinity spectrum can be utilized. A Weak Affinity Dynamic Microarray (WADM) utilizes nanosensor transducers that allow for single molecule adsorption and desorption dynamics to be measured in real time, as opposed to equilibrium binding only. This resolves both weakly and strongly binding partners. This concept could be realized by one of the many nanoengineered sensors reviewed in the previous chapter, or ideally the SWNT-based sensors developed in this work. To model this concept, we use approximate binding constants from 75 lectins and 442 glycans and use the model to define the requirements and limitations of the dynamic array. A Kinetic Monte Carlo model of reduced glycoproteins can estimate the optimal number of lectin types, number of transducers, and glycoprotein concentration for three prototypical glycan profiling applications: 1) screening protein therapeutics, 2) clinical biomarker screen, and 3) complete profiling of glycoproteins without *a priori* knowledge of their synthesis pathway. Finally, experimental considerations for profiling homogenous and heterogeneous solutions of glycoproteins with a lectin WADM are discussed as well as using the dynamic array to obtain an accurate database of mono and multivalent interaction parameters between lectin-glycan pairs.

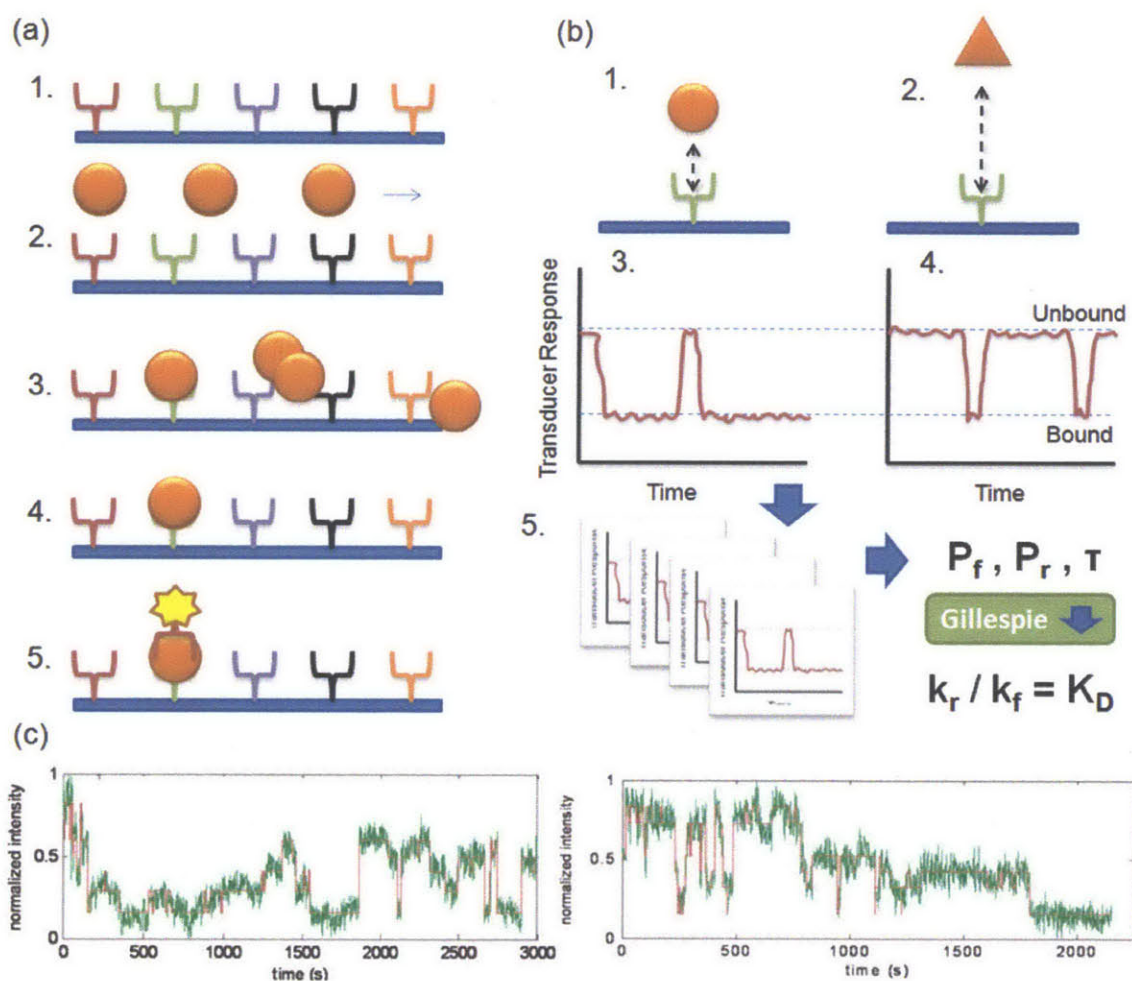
### 2.2 Introduction

In the past two years there has been an explosion of technologies to recognize sugar groups on proteins. Glycans participate in protein signaling, interaction, structure, and folding thus knowledge of the attached glycans helps better characterize the protein.<sup>1,2</sup> This is especially applicable to protein therapeutics where over two thirds of the rapidly-growing market is composed of glycoproteins.<sup>3</sup> Expression of these therapies in non-native cell lines yields a highly heterogeneous mixture of glycosylated proteins (although there have been recent studies showing the production of homogenous glycoproteins using eukaryotic cells<sup>4-6</sup>). The heterogeneity is unacceptable for highly-specific therapies, as well as passing any stringent drug review process.<sup>3</sup> Thus many profiling systems have been proposed:

magnetic nanoparticles,<sup>7</sup> reverse-phase chromatography tandem mass spectrometry,<sup>8</sup> nanoflow liquid chromatography coupled with Fourier-transform, ion-cyclotron, resonance mass spectrometry,<sup>9</sup> and magnetic resonance imaging.<sup>10</sup> Lectin arrays have also recently been proposed,<sup>11-15</sup> however these current platforms are limited to detection of more strongly-affined ligands. The concept explored in this work, a Weak Affinity Dynamic Microarray (WADM), may overcome these limitations by monitoring single-molecule adsorption and desorption dynamics in real time, as opposed to equilibrium binding only. In this way, the full spectrum of affinities, with an emphasis in the weakly binding regime, can be resolved, thereby reducing the number of lectin types needed for glycoprotein profiling.

Lectins, unlike antibodies, have a wider range of binding affinities, but their affinities are much weaker than typical antibody-antigen pairs (experimental values of  $K_D$  ranged from  $10^{-3}$  to  $10^{-7}$ ).<sup>15</sup> It has been proposed that a carefully selected array of stronger binding lectins might be able to profile glycoproteins.<sup>11-15</sup> Recent speculation has centered upon the case of four lectins that exhibit characteristic, fluorescent binding patterns to four different glycans.<sup>16</sup> By using the four lectins together in a microarray, they suggested that a unique fluorescent signature could be obtained for each glycan. The limitation in this approach is that it relies only on strong associations between glycan-lectin pairs to obtain a unique fluorescent signal. The weakly binding pairs are lost in the typical processing steps of conventional arrays (Fig 2.1a). When one considers the vast number of natural structures in the glycome (estimated at  $10^4 - 10^5$  structures<sup>16</sup>) and the infinitely varied synthetic structures, such a platform would require a substantial set of unique lectins to provide a clear fluorescent signature from bound-pairs. If however one could utilize the full spectrum of affinities to provide a clear readout signature, then the set of unique lectins required in a microarray would be greatly reduced. The concept of achieving this reduction by utilizing affinity information directly is the focus of this work.

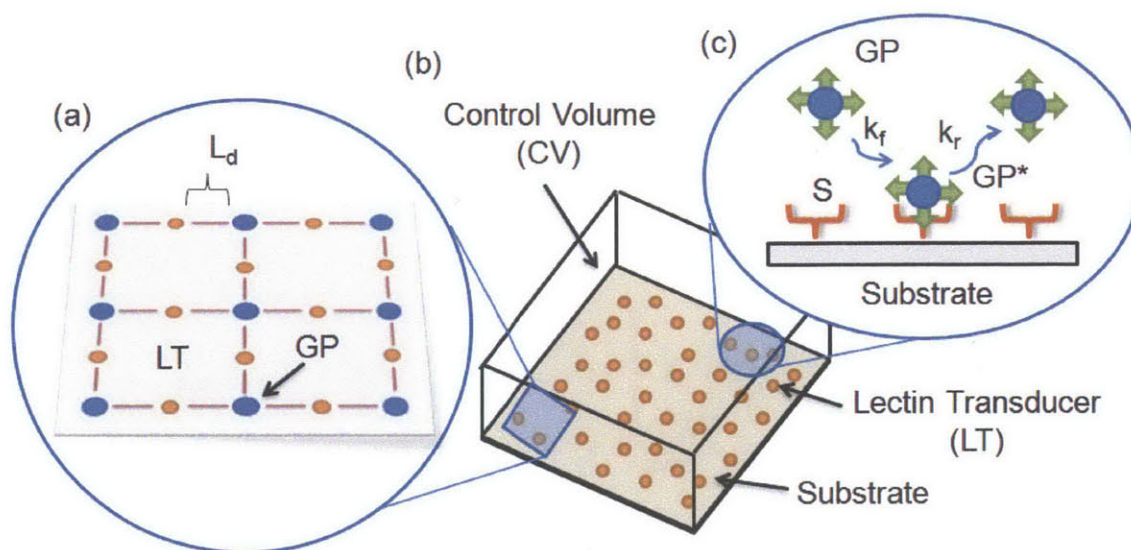
As an alternative to current microarray technology which uses strong-binding, ligand-receptor pairs for differentiation, the transient binding behavior of weakly-affined lectins (Fig 2.1b) can be utilized to identify the glycan. By monitoring the statistical behavior of a glycoprotein binding on and off the lectin receptor, the full spectrum of the lectin's affinities can be used to profile the glycoprotein in solution. Note that several emerging nanosensor platforms allow for single molecule adsorption/desorption dynamics to be recorded in real time.<sup>17-21</sup> The response of such nanosensors (Fig 2.1c) could be used to determine the kinetic parameters between glycan-lectin pairs. These parameters in turn, could help differentiate between glycans present on the protein surface. By carefully choosing the right lectin types and running the assay under optimal conditions, the profiling accuracy can be greatly improved.



**Figure 2.1** - Comparison of traditional lectin array and conceptual lectin WADM. Traditional arrays (a) consist of the following detection steps: 1) construct array with many unique lectin types, 2) flow homogenous glycoprotein sample over array, 3) incubate for binding, 4) wash unbound glycoproteins away, 5) tag bound glycoproteins to determine lectin location (if the glycoprotein is not tagged prior). In this scheme only the more affined interactions are resolved. In the WADM (b) both strong and weak interactions can be detected. The strong-binding glycoprotein has less dynamic response (1) than the weak-binding (2). This dynamic signal could be transduced electrically or fluorescently for the strong (3) and weak (4) glycoproteins. An ensemble average for the probabilities of binding ( $P_f$ ) and release ( $P_r$ ) as well as the average event time ( $\tau$ ) could be constructed from long-term monitoring of a single lectin site, or composing an average from an array of lectin transducers (5). These ensemble values could then be used to calculate the reverse and forward kinetic parameters by Gillespie's equations (Gillespie 1977). This has been done experimentally (c) on the single molecule level using the fluorescent response of carbon nanotubes (Jin *et al.* 2008). The right graph shows the quenching response of 20  $\mu\text{M}$   $\text{H}_2\text{O}_2$  and the left of HCl at a pH of 4. The green trace is the fluorescent response and the red is a Hidden Markov Model fit used for transducers that have more than one binding site.

## 2.3 Experimental Model

This dynamic array was modeled with a Kinetic Monte Carlo simulation of a Langmuir association/dissociation surface reaction (Fig 2.2c). Glycoproteins in solution (GP) bind to free lectin sites (S) to form glycoprotein-lectin pairs (GP\*). The pairs can then dissociate back to free sites and glycoproteins in solution. Association and dissociation are governed by the forward and reverse rates ( $k_f$  and  $k_r$ ) respectively. The ratio of these rates ( $k_r/k_f$ ) is the disassociation constant ( $K_D$ ). These interaction values can be determined experimentally from liquid chromatography (LC), mass spectrometry (MS), capillary electrophoresis (CE), and frontal affinity chromatography (FAC).<sup>15</sup> Due to growing importance of glycoproteins, databases of these affinity values are continually expanding, but access for general researchers is still limited.

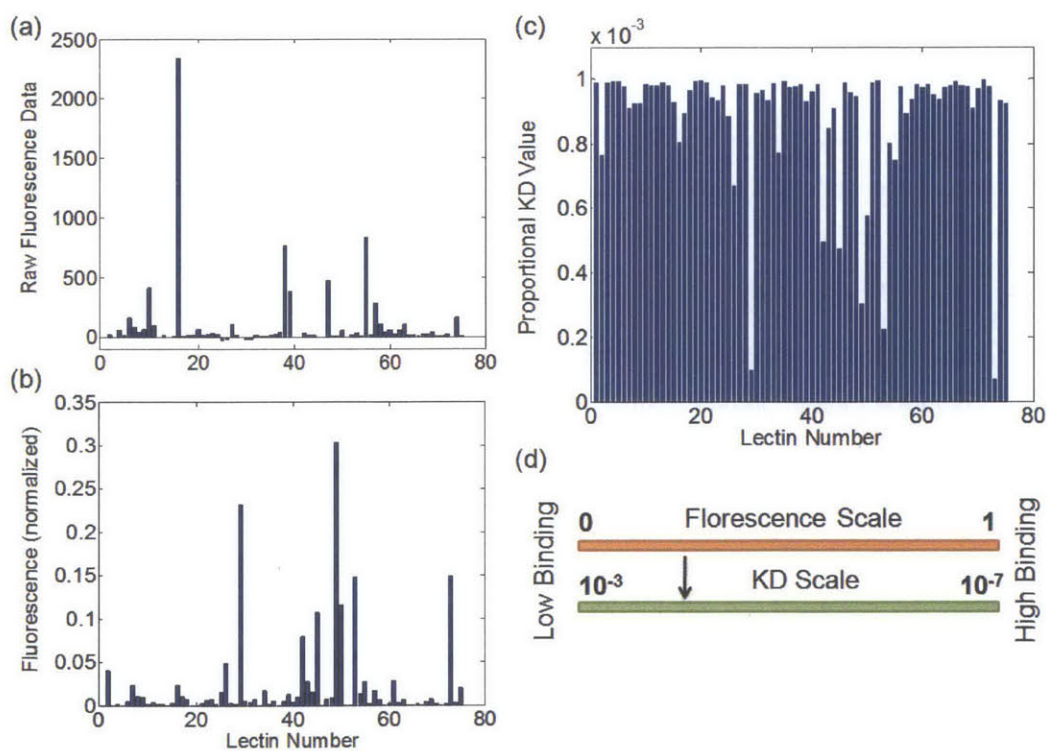


**Figure 2.2** – Setup of stochastic model. (a) Spacing of lectin transducers (LT) is based on the diffusion length scale ( $L_d$ ) of the glycoproteins in solution. This length scale is dynamically solved in the stochastic model and the overall spacing yields a control volume. (b) The control volume depth is set at  $L_d$ . (c) The surface reaction follows Langmuir equation. Free glycoprotein (GP) associates with free lectin surface points (S) to form a glycoprotein-lectin pair (GP\*) according to the forward reaction rate ( $k_f$ ). The glycoprotein also disassociates from the surface to reform free GP according to the reversible rate ( $k_r$ ).

The kinetic parameters for this model were supplied in the form of dissociation constants between 75 unique lectins and 442 glycans, approximated from the only public database of lectin-glycan interaction values.<sup>22</sup> The CFG collects lectin proteins submitted by consortium investigators and the CFG core staff assay the proteins on glass slides with immobilized glycans. Affinity of lectin-glycan pairs are measured by relative fluorescence. The most current printed array (Version 4.1) has 442 glycans. To

date over a hundred lectin submissions have been made by CFG members on this current array, with 75 of them being unique lectin types at varying low concentrations.

The approximate  $K_D$  interaction matrix for this model was made by combining the experimentally known  $K_D$  range from Hirabayashi's work<sup>15</sup> with the fluorescent data given by the CFG.<sup>22</sup> Hirabayashi found that glycan-lectin  $K_D$  values range from  $10^{-3}$  to  $10^{-7}$ . It is assumed that the 442 glycans present on the CFG array span this interaction range. By normalizing the relative fluorescence data of each CFG lectin submission to its respective fluorescence range, and then finding the proportional distance in the  $K_D$  range, a full matrix of Glycan-Lectin approximated  $K_D$  values can be formed for the 75 unique lectin entries (Fig 2.3 and Appendix 2.1). This database is much smaller than the number of naturally occurring glycans<sup>23</sup> and the kinetic estimates from the fluorescent data are rough, however the model still provides valuable insight into the working parameters of a dynamic array. More accurate kinetic parameters could be obtained from the dynamic array itself and fed into the model to provide better predictions of lectin subsets for specific screening and profiling problems.



**Figure 2.3** – Approximation of dissociation constants from CFG fluorescent database. Graphs shown for Glycan 1 (Gal $\alpha$ -Sp8): a) raw relative fluorescence data compiled from the CFG database, b) fluorescence normalized to range used by each lectin submission, c)  $K_D$  estimate made from proportionality of fluorescence range to measured  $K_D$  range (d).

The Gillespie algorithm<sup>24</sup> was used to find an ensemble average of occupied lectin sites for a given glycoprotein in solution. Stochastics were invoked in this model due to the very small species concentrations and simulated control volume. First the forward and reverse rate expressions were transformed into probability functions:

$$P(R_f) = \frac{k_f \cdot N_S \cdot N_{GP}}{A_N \cdot V} \quad (2.1)$$

$$P(R_r) = k_r \cdot N_{GP^*} \quad (2.2)$$

The concentration of glycoproteins in solution and the number of lectin sites (NS) were specified by the user. The control volume (V) was dynamically solved for using the diffusion length ( $L_d$ ) of protein in solution and the event time step ( $t_s$ ) predicted by the Gillespie algorithm:

$$L_D = \sqrt{4 \cdot D \cdot t_s} \quad (2.3)$$

The diffusion length was used to space the lectin transducers into a regular grid (Fig 2.2a) and the resulting area was multiplied by the reactive depth (again the diffusion length) to produce the control volume (V) (Fig 2.2b). The number of free glycoproteins in solution (NGP) was calculated from the specified protein concentration and the control volume. To ensure that both reactive probabilities have the units of molecular event per second, Equation 1 also includes Avogadro's number (AN). The forward reaction rate was assumed to be diffusion limited. The diffusion constants for glycoproteins vary by a small degree, so an average protein  $k_f$  value of  $2E8 \text{ (M}\cdot\text{s)}^{-1}$  was used.<sup>25</sup>

The Gillespie algorithm uses an ensemble average of probabilistic binding/dissociation events over a given simulation time to predict frequency of events and average occupancy number. Each experimental "run" is for 1000 binding events, and the ensemble average is composed of 1000 runs. The code loops through each of the lectins to determine which lectin types provide the greatest contrast for screening/profiling purposes. The general algorithm goes as follows: 1) read in glycan profile for protein in solution (either user specified or randomly generated – numbers should correspond to the glycan rows in the interaction  $K_D$  matrix (Appendix 2.1)), 2) generate a random number to determine which glycan site is presented on the glycoprotein, 3) generate a random number to determine which of the reactions occur, 4) adjust count of GP, S, and GP\*, 5) generate next time step according to Equation 2.4, 6) Add time step to time count, 7) loop through steps 2-6 until 1000 binding events are achieved, 8) conduct 1000 runs of the stochastic simulation to create ensemble average, and 9) perform an ensemble average for each of the lectins to determine which would be best for screening purposes. This algorithm was compiled in MATLAB.

$$\Delta Time = -\left(\frac{1}{R_f + R_r}\right) \cdot \log(rand) \quad (2.4)$$

To make the simulation tractable, the modeled glycoprotein has been significantly reduced. The natural occurrence of heterogeneous mixtures of glycoproteins (glycoforms) has been neglected in the current model. Only homogenous solutions of glycoproteins are considered, although a system for profiling heterogeneous mixtures is discussed below. In the first three simulation examples, each glycan on the surface of the simulated glycoprotein is assumed as equally accessible (a random number is used to determine which glycan is presented to the lectin); but in the fourth example, profiling a glycoprotein is modeled while taking into account spatial arrangement. As multivalent interactions would lead to obvious, non-dynamic responses (the glycoprotein would simply adhere strongly to the lectin transducer), these events are omitted in the dynamic model. These events are also rarer between a wide range of glycans and lectins due to the required, close-matched spacing of carbohydrate recognition domains and surface ligands needed for a multivalent interaction.<sup>26</sup> However the dynamic array could be used to measure more accurate mono and multivalent kinetic parameters. Interactions between the protein surfaces (charge, Van der Waals, and sterics) are assumed minimal.

To demonstrate how a dynamic microarray could be used for screening and profiling, the reduced experimental model was applied to three pertinent areas of glycoprotein research: 1) screening of protein therapeutics, 2) screening clinical biomarkers, and 3) complete profiling of unknown glycoproteins. A fourth example then explores profiling glycoproteins while taking into account spatial arrangement, or accessibility.

## 2.4 Results and Discussion

### 2.4.1 Screening of Protein Therapeutics – Galactose- $\alpha$ 1,3-Galactose

A recent example of recombinant glycoprotein therapeutics being expressed with harmful glycans is the drug Cetuximab used for colorectal cancer and squamouscell carcinoma. Chung *et al.* found that this murine-expressed, monoclonal antibody occasionally contained a glycan, Galactose- $\alpha$ 1,3-Galactose, which caused hypersensitivity in a third of the treated patients.<sup>27</sup> Another study found that this therapeutic had 21 unique glycan candidates.<sup>28</sup> Most of these glycans are not yet part of the kinetic parameter database; however two of the candidate, biantennary glycans are included: one benign (Glycan #50) and one with two deleterious Galactose- $\alpha$ 1,3-Galactose groups (Glycan #359) (Fig 2.4a insert). A dynamic microarray could be used to screen between these two and determine which cell line is

producing harmful strains of the drug. This first case study is the simplest form of screening: differentiation of one glycoprotein from another.

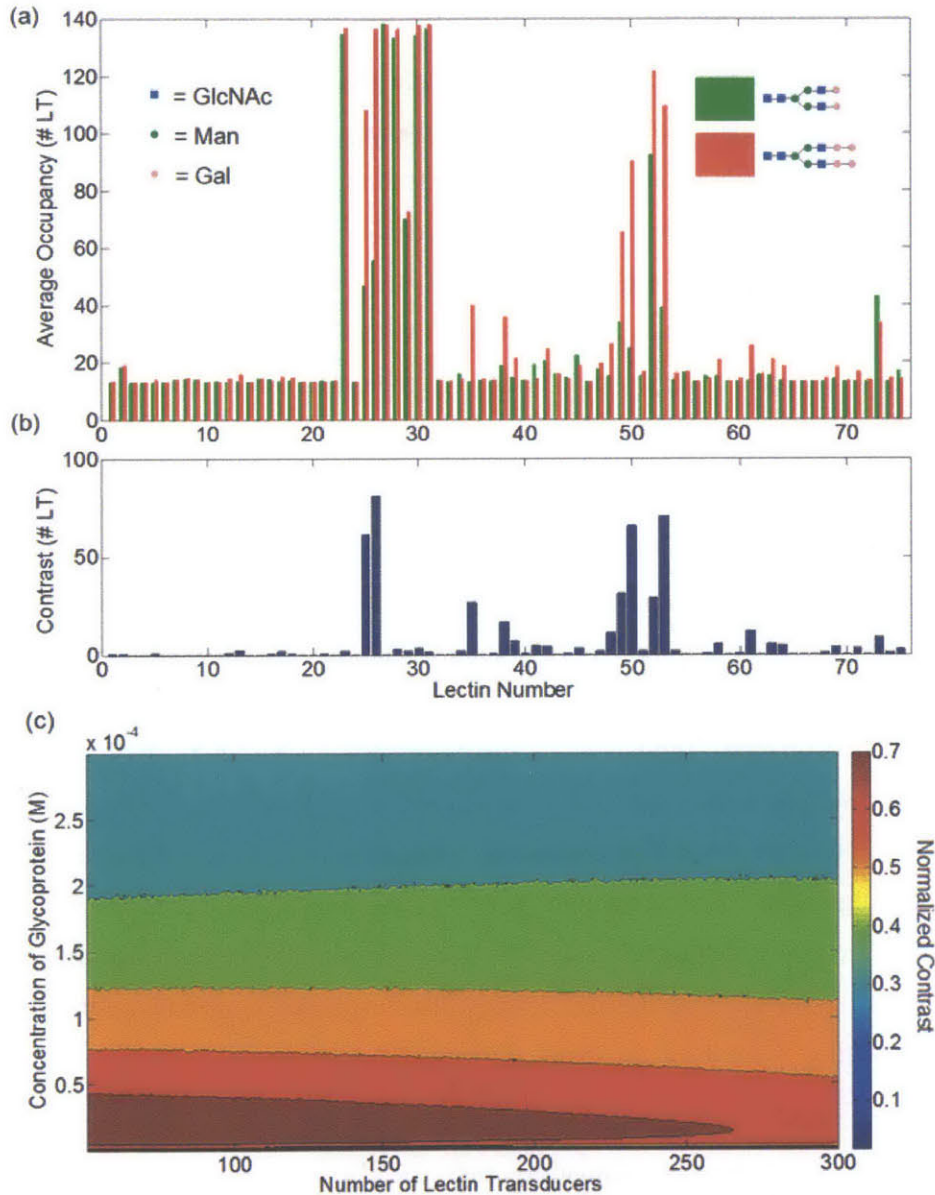
The kinetic model simulates the dynamic response of the unique glycoprotein in solution to find optimal conditions for screening. Utilizing the user-supplied kinetic parameters (in the form of dissociation constants), it determines the ensemble average of bound lectins over time. The responses of both glycoproteins are then determined for each lectin type (150 ensemble averages generated for the current 75-lectin database). The responses of the two glycoproteins are then compared for each lectin type to determine the optimal lectin. Once the optimal lectin has been found, the best operating conditions are determined by running the model at various concentrations and number of lectin binding sites (or lectin transducers). The resulting contrast values are normalized by the total number of lectin transducers used in the simulations. This yields an operating space to find the optimal number of transducers and concentration of glycoproteins in solution. This predictive process is illustrated with the approximate  $K_D$  database for the Cetuximab problem.

The model predicts that a single lectin type is sufficient for clear differentiation between two homogenous solutions of Cetuximab therapeutics and provides the practical operating conditions for the screen. By looking at the average number of bound species over time of both proteins on each lectin (Fig 2.4a), it was found that a single lectin is needed to show contrast between these two glycoproteins. Lectin 26 (Antibody BD6 made to rat basophilic leukemia cells (RBL-2H3)) was found to show maximum contrast (difference between the average number of bound species over time of the two proteins) (Fig 2.4b). By varying both number of lectin transducers and concentration of glycoprotein an operating space is generated to help maximize contrast in the screen (Fig 2.4c).

There exists an optimal protein concentration and number of transducers as too few of the latter will not yield a measurable response above the noise and too many proteins in solution will saturate the binding sites. At saturation, both protein types would be in excess and bind to the limited number of transducers, thus decreasing the ability to differentiate between the two solutions. For this example, glycoprotein concentrations less than 5  $\mu\text{M}$  and greater than 100  $\mu\text{M}$  inhibit differentiation. It is found that a glycoprotein concentration of 20  $\mu\text{M}$  and 150 lectin transducers exhibits a normalized contrast of 0.74; the protein with the benign group (#50) occupies 111 less (out of the 150) lectin transducer sites than the protein with the harmful glycan (#359). In this particular example, the range of allowable glycoprotein concentrations and number of lectin transducer sites is wide. In other situations where the affinities are weaker or the two glycoproteins are very similar there is a much tighter optimal range. Galactose- $\alpha$ 1,3-Galactose is just one of many potentially harmful or simply unwanted glycans that can be expressed on therapeutics or other laboratory glycoproteins. For each case of differentiation between two



known glycoprofiles there will be a unique, optimized lectin and operating space. However the generalized method described above holds for all differentiation-screen scenarios and the predictive power is increased with the quality of available kinetic data.



**Figure 2.4** – First simulation: screening for benign glycan #50 (green) versus harmful glycan #359 (red) present on Cetuximab drug. a) Average occupancy for both glycoproteins on each lectin type (number of occupied transducer sites), insert shows the structure of Glycan #50 and #359. b) Contrast (difference in average occupancy of two proteins on a lectin) to find the best lectin transducer (at 200  $\mu$ M concentration of GP and 200 LT) to differentiate these two glycoprotein profiles. Lectin #26 (Galectin-1:Galectin-1 N-term CRD-Galectin-9 linker-Galectin-1 C-term CRD (1-9-1)) yields a contrast of 81 lectin transducers. c) Effect of Lectin Transducer (LT) number and concentration of glycoprotein (GP) on contrast of best lectin (Lectin #26) for glycoprotein therapeutic screening.

Contrast refers to the difference in number of occupation sites between the two glycoproteins on Lectin 26. It is normalized by total number of LT to find optimal region.

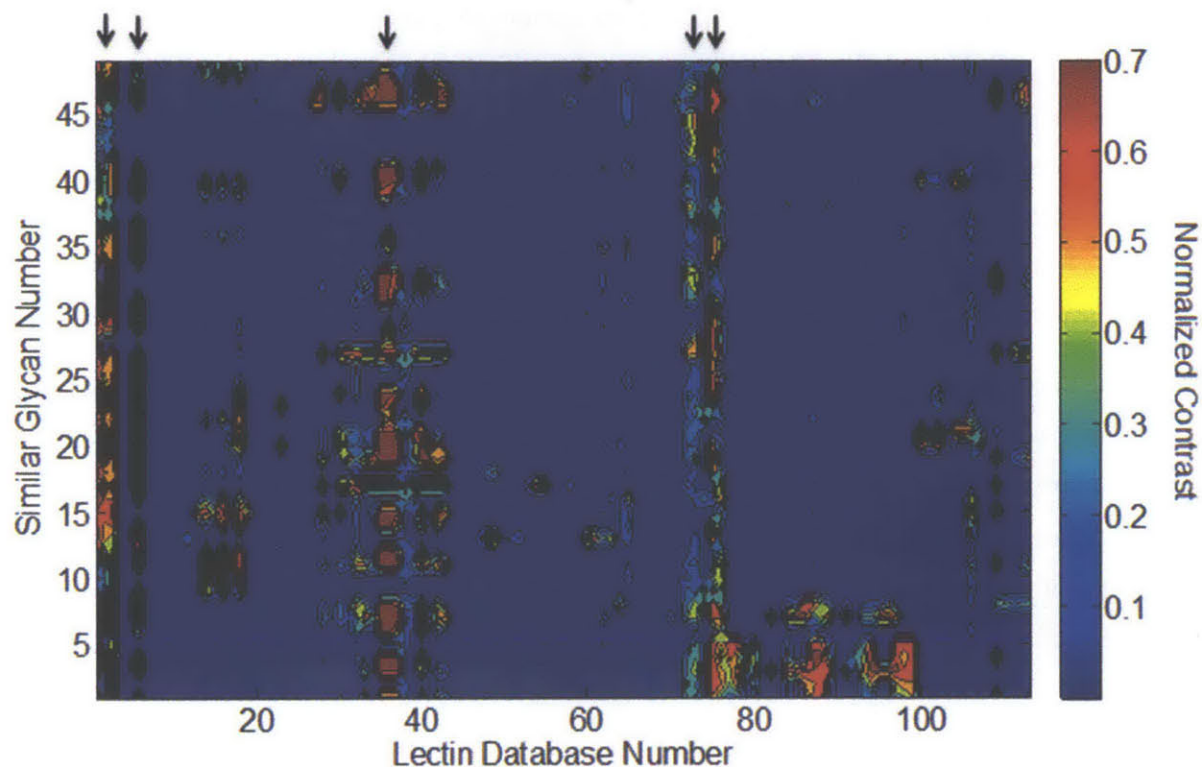
#### **2.4.2 Screening Biomarkers – Arthritic Disease and Glycosylation of IgG**

Arthritis has been linked with changes to the single glycosylation site in the constant domain of the IgG Fc region.<sup>29</sup> It is thought that by screening the glycan group on these proteins, clinicians would be able to tell what type of Arthritis an individual has. Healthy individuals have a large glycan group in this region, yet as the arthritis severity increases, this glycan group loses many of its terminal sugars.<sup>29</sup> When an increased count of the IgG proteins contain only the base glycan structure ((GlcNAc2Man6)-Man4GlcNAc-4GlcNAc also known as IgG-G0), the individual is likely suffering from Rheumatoid Arthritis (RA).<sup>30</sup> In a clinical setting it would be advantageous to create a rapid screening tool for this prescient glycan group from all the other possible IgG variants.

This second case study represents a more difficult situation of screening than the proceeding example (differentiation of one “desired” profile from 49 other “undesired”), yet the basic method outlined in the previous case study holds. The ensemble average of occupied sites is generated for each of the glycoproteins to each lectin type (50 glycoproteins with the 75-lectin database yields 3750 simulations). The response of the “desired” protein for screening is then subtracted from all the “undesired” for each lectin type. These contrast values are then normalized by dividing by the total number of lectin transducers in the simulation. An algorithm was then written that sorted the lectin database types by contrast response (best contrasting lectin to worst) for each of the “undesired” glycan types. Finally, the algorithm searched through the best contrasting lectin types for each of the “undesired” glycans to find which minimal subset of lectins could achieve clear profiling between the “desired” glycan and all the other “undesired” groups. This expanded algorithm is demonstrated for the case of IgG-G0 screening using the approximate 75-lectin database.

Our stochastic model predicted that two types of lectins would be sufficient to differentiate the RA glycan group (IgG-G0) from other similar glycans. The glycan group was identified in the database as glycan #301 and the database was then searched for other glycans with similar base structures; this yielded 49 similar glycans (Appendix 2.2). The contrast between glycan #301 and the other 49 glycans were generated for each lectin type and normalized by the total number of lectin transducers (Fig 2.5). The algorithm described above then sorted the lectin database numbers by contrast response (best contrasting lectin to worst) and found which subset of lectin types were among the top-ten contrasting lectin types for each of the similar glycans (normalized contrast response data found in Appendix 2.3). It was found that by using both Lectin #1 (LSEctin:Mouse LSection 0.5) and #52 (Nictaba:Nictaba, wild

type) high contrast could be shown between Glycan #301 and each of the other similar glycans. Used together, the two lectins exhibit high contrast over the entire range of glycans. However, these predictions are based on approximate kinetic data. As better kinetic parameters between glycans and lectins are obtained, these predictions can be further refined by the described model and method.



**Figure 2.5** – Normalized contrast of Rheumatoid Arthritis glycan (#301) from other similar glycan groups (1-49) for all lectin types in current CFG database. Contrast is given in average number of occupied lectin transducer sites divided by total number of sites. Simulation run at 200 LT and 0.2 mM concentration of glycoproteins. Arrows denote lectins with exhibit a wide range of unique contrasts, potential candidates for screening the RA glycan group from other similar groups.

### 2.4.3 Complete Profiling of Unknown Glycoprotein

The most difficult differentiation scheme to imagine for this platform would be full profiling of an unknown glycoprotein. This would be advantageous in many Glycobiology applications, such as the detection of known biomarkers and discovery of new. Glycoprotein biomarkers have been found for general inflammatory problems,<sup>31</sup> prostate cancer,<sup>32</sup> ovarian cancer,<sup>33</sup> and a host of other diseases.<sup>34</sup> The stochastic model provides an estimate of the number of lectin types would be needed for clear profiling.

Profiling is a monumentally harder problem than the previous screening examples, as there are more unknowns to differentiate. For a model protein with a single glycosylation site, there are 442 unique possibilities in the current database (442 different glycans on the current CFG array, many more groups naturally exist<sup>23</sup>). For a protein with two glycans there are 97,903 unique combinations (442 choose 2), for three groups there are 14,294,280 combinations (442 choose 3), and upward to an average biological limit of 5 glycosylation sites with an astounding number of 137 billion unique combinations (442 choose 5). The possibilities are further compounded when more glycans are added to the model and when spatial considerations are taken into account (as seen in the next example). Finding a unique, strong binding ligand for each of these glycoproteins is an impossible task, but with dynamic observation of lectins a much smaller number of lectin types could be used to make an effectual bar-code for reading each glycoprofile.

Due to computational limits, only the cases of proteins with one and two glycosylation sites have been currently simulated, but the process holds for glycoproteins with more glycans. A response matrix (average number of occupied transducers) for each glycan lectin pair was generated with the conditions of 200 lectin transducers and a glycoprotein concentration of 300  $\mu\text{M}$  (Appendix 2.4). An algorithm then sorted the lectin response vectors in order of uniqueness (most unique to least). Uniqueness was measured by scoring how many glycan responses on the given lectin were outside a tolerance of four occupied sites from the other glycan responses on the same lectin (this indicates a glycan group that could be readily differentiated from the others on the lectin). A second algorithm then searched down the resorted matrix to determine how many lectins it would take to create a unique response “bar code” for each possible glycan profile (again using the tolerance of at least four occupied site difference). For a protein with a single glycosylation site and the 75 lectins now available in the database it was found that 6 lectin types were needed for full profiling, and for a two glycosylation site protein 37 lectin types are needed (see Appendix 2.5 for expected response data of single glycosylated protein).

#### **2.4.4 Adding Accessibility to Profiling Algorithm**

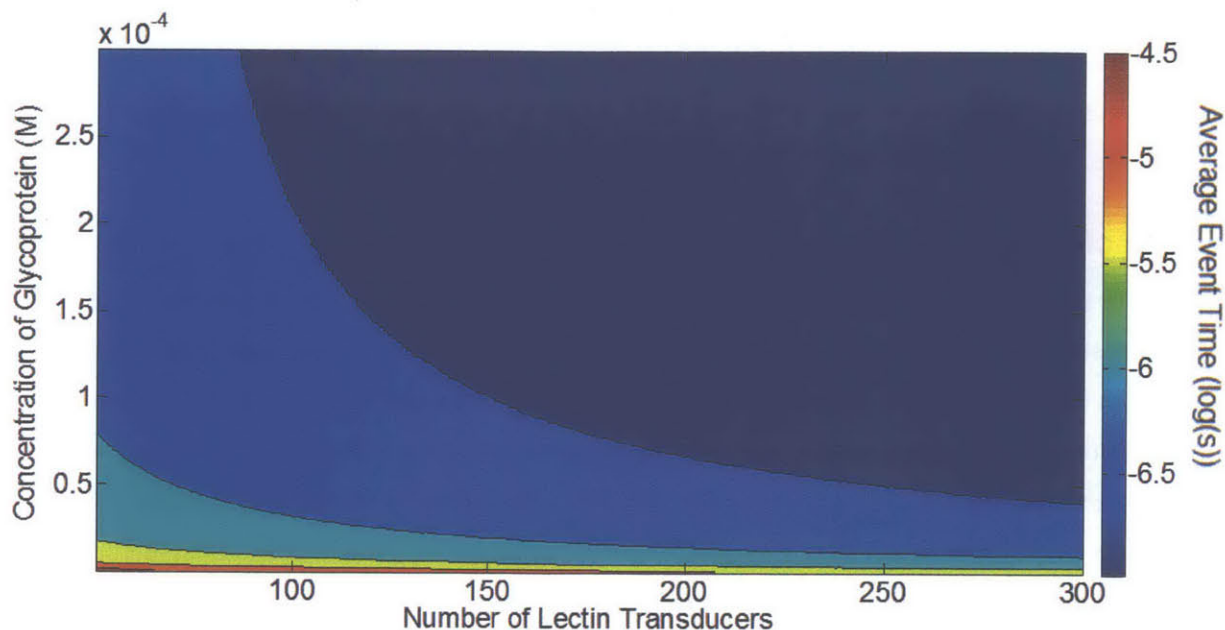
In the preceding profiling example, spatial arrangement of the glycans was neglected. If considerations are made for the accessibility of the glycans to the lectin transducer, the number of possible profiles is greatly increased as is the required number of lectin types for clear profiling. Consider a simple case where the two-site glycoprotein model has been modified to include three different spatial arrangements: 25% accessible in glycan A - 75% accessible in glycan B, 50% in both, and 75% in glycan A - 25% in glycan B. If the same profiling algorithm as above is used, the number of unique profiles increases three-fold (293,709 response profiles) and the number of lectin types required for clear profiling

is not met with the current database of lectins. A much larger set of unique lectins is needed to both profile and determine the spatial arrangement of glycans on a homogenous solution of glycoproteins. The number depends on the uniqueness of the lectin, and as more lectins and other sugar-binding proteins are explored this required number will decrease.

## 2.5 Conclusion

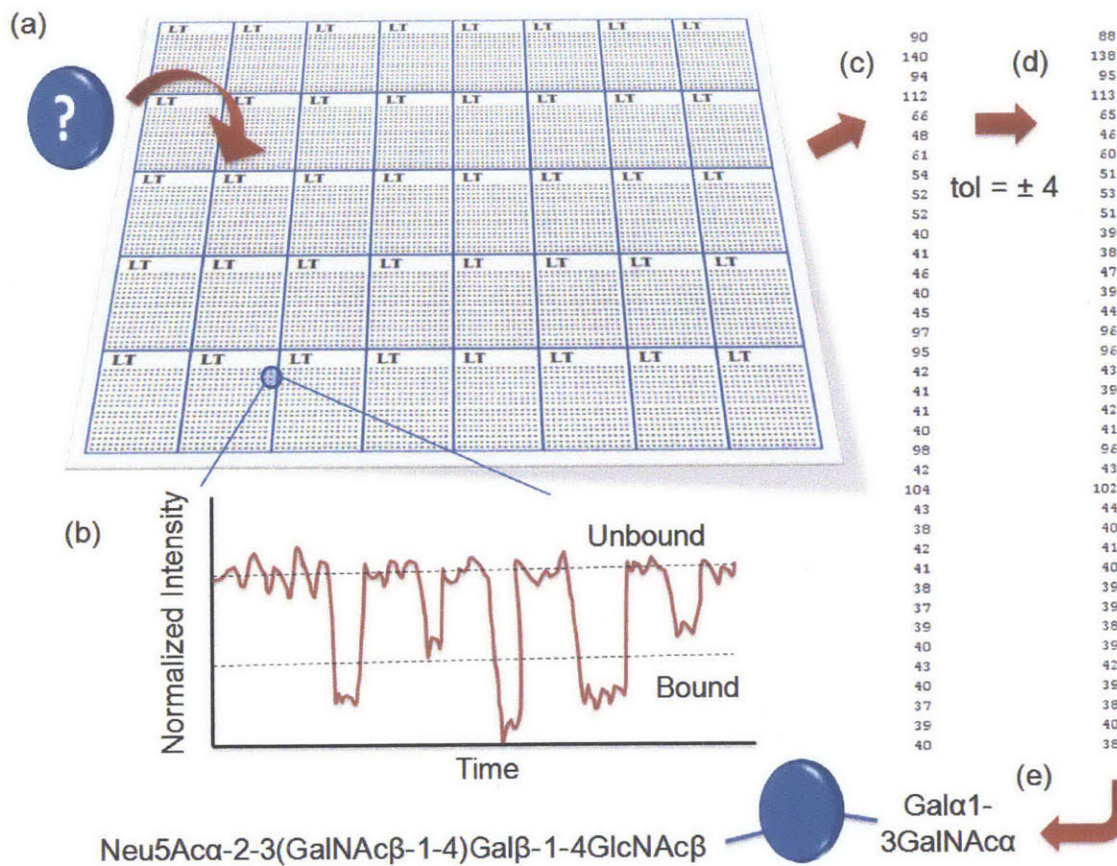
The predictions of this stochastic model are promising but they are only as good as the affinity parameters supplied to the model. A comprehensive, public database of lectin-glycan binding affinities does not yet exist. In order to make better predictions of which lectin types will be best for specific screening situations such a database should first be obtained. The WADM can be used to obtain more accurate kinetic parameters between glycan-lectin pairs. These in turn could be cycled back to the model for clearer predictions of which lectins to use for specific applications. Also, by expanding the library of lectins there is a greater possibility of finding lectin types with more unique glycan affinity signatures. This will reduce the total number of lectin types needed for profiling.

Although the affinity parameters presented in this work are approximated, they are on the correct experimental order and the operating conditions found herein will hold for any lectin dynamic array system. Such an experimental system would require the ability to monitor single protein binding events via electrical or fluorescent transduction at a very high frame rate. In order to capture the correct statistics, or ensemble average, of occupied transducers, the system would need to be sampled on the order of the binding event time scale. This time scale changes depending on the concentration of glycoproteins and number of lectin transducers used. Consider the hypothetical therapeutic screening of the first case study. The average event time scale (for binding or release) is on the order of 100 nanoseconds (Fig 2.6). This corresponds to a frame capture rate of 10 MHz. This is on the cusp of current single-molecule detection schemes using highly sensitive nano-mechanical cantilevers,<sup>35</sup> nanowire electronic sensors,<sup>36</sup> or carbon nanotube fluorescence.<sup>18</sup> It would be possible to operate at a lower frame rate for situations where the contrast is greater or with lectins that have higher affinities. Because of the extremely high frame rate, enough data could be gathered in a fraction of a second to determine what glycoprotein is present in solution, hence making this an extremely rapid profiling platform.



**Figure 2.6** – Average time scale of binding or release events in glycoprotein therapeutic case study. The time scale changes for different operating conditions (number of transducers and glycoprotein concentrations). This reveals that the experimental frame capture rate would need to be on the order of 100 nanoseconds.

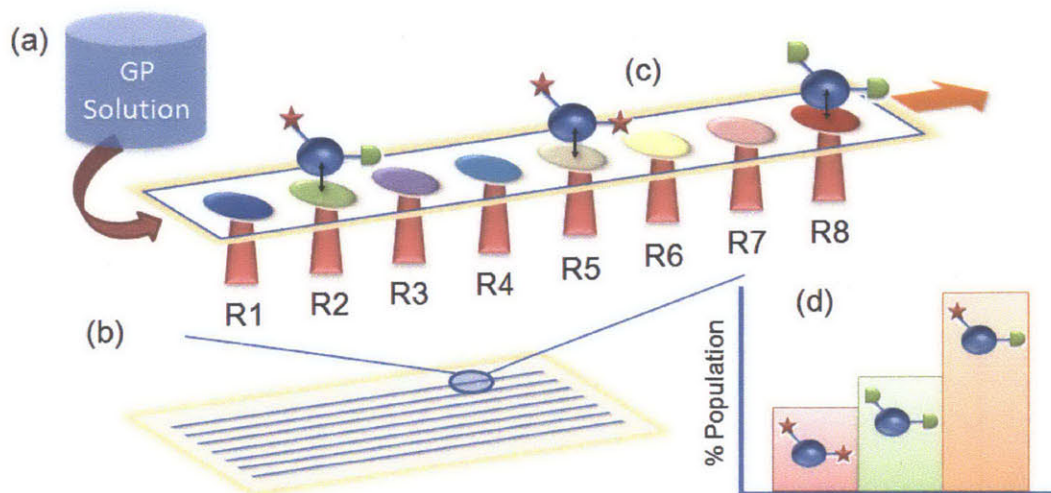
The experimental setup for profiling homogenous solutions of unknown glycoproteins would consist of a greater array of WADMs (Fig 2.7). The stochastic model currently predicts the number of lectin types that are required for clear profiling (37 lectin types for 2 site glycosylation). If the lectin transducer response is a fluorescent signal, these arrays could be micro-printed onto a single glass slide. The glycoprotein solution would then be added on top of the large array assembly. Each lectin type would be monitored for a fraction of a second to determine the average site occupancy numbers. Expected response traces (Fig 2.7b) will be similar to those found in an emerging class of stochastic sensors based on nanoscale transducers that allow for single molecule dynamics to be resolved.<sup>18</sup> The stepwise signal can be translated to both kinetic parameters (forward and reverse rates) as well as an ensemble average of bound lectin sites (Fig 2.7c). The ensemble averages could then be fed to the response database which is searched (within the given tolerance) for a profile match Fig 2.7d). If a successful match is found, the profile is supplied to the user (Fig 2.7e). This simple experimental setup could be expanded for a larger number of glycosylation sites. Again, the number of required lectin types will be reduced as more lectins are screened for unique glycan binding characteristics.



**Figure 2.7** – Proposed experimental set up for profiling unknown glycoproteins with two glycosylation sites. a) The unknown glycoprotein is added on the platform (a greater array of dynamic lectin transducers) at optimal concentration. b) If the transducers exhibit a fluorescent response (like those of Figure 1C), each lectin spot on the array could be dynamically monitored as a pixel trace via a CCD camera. c) The ensemble average of bound sites over time could be determined from each trace. The ensemble average for each lectin type would then be fed to the response database for profiling. d) The known glycoprofile database is searched for a match to the experimental data (within the given tolerance). e) Finally, the matched glycoprofile is outputted to the user. The entire procedure could be completed within seconds and can be optimized as more lectins are screened for uniqueness.

Rapid profiling of a heterogeneous solution of glycoproteins (the estimable goal of the field) would require a more elaborate setup for the dynamic array (Fig 2.8a). It could be realized with emerging nanofluidic platforms.<sup>37-39</sup> The setup would consist of a series of nanochannels with patterned lectin transducers along the bottoms, in which glycoproteins are induced to flow in single-file (Fig 2.8b). The glycoproteins would flow at a rate that would allow dynamic interaction with the lectins. The response of the transducers would be recorded for each protein as it drifts over the different types of lectins (Fig 2.8c). After a sufficient sample of proteins has been profiled, a histogram of the glycoprotein types could be generated (Fig 2.8d). The number of unique lectin types would depend greatly on the desired fidelity of

the profile and on the application. However hundreds of lectin types could easily be patterned on a single chip with existing micro and nanofabrication techniques.



**Figure 2.8** – Proposed experimental setup for profiling heterogeneous solutions of glycoproteins with lectin WADM in a nanofluidic array. A small sample of the heterogeneous solution (a) is flowed through nanofluidic trenches which induce single-file passage of the glycoforms. The flow is controlled to allow interaction over the many types of lectins deposited in the trenches (c). The transducer responses (R1-R8...) are tracked for each glycoprotein and computationally interpreted to provide a population histogram of glycoform profiles present in solution (d).

The stochastic model herein promotes the WADM as a valuable potential avenue for glycan screening and profiling. The platform would require less experimental time and does not require a unique, strong-binding site for every glycan combination. With carefully selected lectins the dynamic array could monitor the average occupancy via nanoscale transducers and accurately translate this into a glycan profile. Such transducers could be made with emerging mechanical, electronic, and fluorescent nanotechnology. It is hoped that this work stimulates further interest in the area of glycoprotein profiling and encourages an open, growing database of accurate, kinetic data for lectin-glycan pairs. The concept of a WADM might also be useful in other fields of molecular screening and profiling where finding strong-binding partners for each analyte is an insurmountable task. In the next chapter we will detail the first experimental ‘proof-of-concept’ SWNT-based glycoprotein sensor that could be implemented in a WADM.



## 2.6 References

- (1) Ohtsubo, K.; Marth, J.D. *Cell*. **2006**, *126*, 855-867.
- (2) Gambelin, D.P.; Scanlan, E.M.; Davis, B.G. *Chem Rev*. **2009**, *109*, 131-163.
- (3) Li, H; d'Anjou, M. *Curr Opin Biotechnol*. **2009**, *20*, 678-684.
- (4) Hamilton, S.R.; Bobrowicz, P.; Bobrowicz, B.; Davidson, R.C.; Li, H.; Mitchell, T; et al. *Science*. **2003**, *301*, 1244-1246.
- (5) Rich, J.R.; Withers, S.G. *Nat Chem Biol*. **2009**, *5*, 206-215.
- (6) Schwarz, F.; Huang, W.; Li, C.; Schulz, B.L.; Lizak, C.; Palumbo, A.; et al. *Nat Chem Biol*. **2010**, *6*, 264-266.
- (7) Zhou, W.; Yao, N.; Yao, G.; Deng, C.; Zhang, X.; Yang, P. *Chem Commun*. **2008**, *43*, 5577-5579.
- (8) Prater, B.D.; Connelly, H.M.; Qin, Q.; Cockrill, S.L. *Anal Biochem*. **2009**, *1*, 69-79.
- (9) Bereman, M.S.; Young, D.D.; Deiters, A.; Muddiman, D.C. *J Proteome Res*. **2009**, *8*, 3764-70.
- (10) El-Boubbou, K.; Zhu, D.C.; Vasileiou, C.; Borhan, B.; Prospero, D.; Li, W.; et al. *J Am Chem Soc*. **2010**, *132*, 4490-4499.
- (11) Pilobello, K.T.; Krishnamoorthy, L.; Slawek, D.; Mahal, L.K. *ChemBioChem*. **2005**, *6*, 985-989.
- (12) Zheng, T.; Peelen, D; Smith, L.M. *J Am Chem Soc*. **2005**, *127*, 9982-9983.
- (13) Rosenfeld, R.; Bangio, H.; Gerwig, G.J.; Rosenber, R.; Aloni, R.; Cohen, Y.; et al. *J Biochem Biophys Methods*. **2007**, *70*, 415-426.
- (14) Tao, S.C.; Li, Y.; Zhou, J.; Qian, J.; Schnaar, R.L.; Zhang, Y.; et al. *Glycobiology*. **2008**, *18*, 761-769.
- (15) Hirabayashi, J. *J Biochem*. **2008**, *144*, 139-147.
- (16) Tateno, H.; Kuno, A.; Hirabayashi, J. In: *The Sugar Code: Fundamentals of Glycosciences*; Gabius, H.J., Ed.; Wiley-VCH :Weinheim, 2009; pp 247-260.
- (17) Cognet, L; Tsyboulski, D.A.; Rocha, J.D.R.; Doyle, C.D.; Tour, J.M.; Weisman, R.B; *Science*. **2007**, *316*, 1465-1468.
- (18) Jin, H.; Heller, D.A.; Kim, J.H.; Strano, M.S. *Nano Lett*. **2008**, *8*, 4299-4304.
- (19) Jin, H.; Heller, D.A.; Kalbacova, M.; Kim, J.H.; Zhang, J.Q.; Boghossian, A.A.; Maheshri, N.; Strano, M.S. *Nat Nanotech*. **2010**, *5*, 302-309.

- (20) Goldsmith, B.R.; Coroneus, J.G.; Kane, A.A.; Weiss, G.A.; Collins, P.G. *Nano Lett.* **2008**, *8*, 189-194.
- (21) Goldsmith, B.R.; Coroneus, J.G.; Khalap, V.R.; Kane, A.A.; Weiss, G.A.; Collins, P.G. *Science.* **2007**, *315*, 77-81.
- (22) Consortium for Functional Glycomics. **2010**. *Primary Screen*. Retrieved from CFG Gateway: <http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp>
- (23) Cummings, R.D. *Mol BioSyst.* **2009**, *5*, 1087-1104.
- (24) Gillespie, D.T. *J Phys Chem.* **1977**, *81*, 2340-2360.
- (25) Creighton T.E. *Proteins*. 2nd ed. W. H. Freeman and Company: New York, 1993; p 345.
- (26) Carlsson, S.; Oberg, C.T.; Carlsson, M.C.; Sundin, A.; Nilsson, U.J.; Smith, D.; et al. *Glycobiology.* **2007**, *17*, 663-676.
- (27) Chung, C.H.; Mirakhur, B.; Chan, E.; Le, Q.T.; Berlin, J.; Morse, M.; et al. *N Engl J Med.* **2008**, *358*, 1109-17.
- (28) Qian, J.; Liu, T.; Yang, L.; Daus, A.; Crowley, R.; Zhou, Q. *Anal Biochem.* **2007**, *364*, 8-18.
- (29) Alavi, A.; Axford, J.S. *Dis Markers*, **2008**, *25*, 193-205.
- (30) Watson, M.; Rudd, P.M.; Bland, M.; Dwek, R.A.; Axford, J.S. *Arthritis Rheum.* **1999**, *42*, 1682-1690.
- (31) Gornik, O.; Lauc, G. *Dis Markers.* **2008**, *25*, 267-278.
- (32) Leoz, M.L.; An, H.J.; Kronewitter, S.; Kim, J.; Beecroft, S.; Vinall, R.; et al. *Dis Markers.* **2008**, *25*, 243-258.
- (33) Saldovaa, R.; Wormaldb, M.R.; Dwekb, R.A.; Rudda, P.M. *Dis Markers.* **2008**, *25*, 219-232.
- (34) Peracaulaa, R.; Barrab´esa, S.; Sarratsa, A.; Ruddb, P.M.; Llorensa, R.D. *Dis Markers.* **2008**, *25*, 207-218.
- (35) Waggoner, P.S.; Craighead, H.G. *Lab Chip.* **2007**, *7*, 1238-1255.
- (36) Zheng, G.; Patolsky, F.; Cui, Y.; Wang, W.U.; Lieber, C.M. *Nat Biotechnol.* **2005**, *23*, 1294-1301.
- (37) Daiguji, H. *Chem Soc Rev.* **2010**, *39*, 901-911.
- (38) Kim, S.J.; Song, Y.A.; Han, J. *Chem Soc Rev.* **2010**, *39*, 912-922.
- (39) Napoli, M.; Eijkel, J.C.T.; Pennathur, S. *Lab Chip.* **2010**, *10*, 957-985.

## Appendix 2.1

Lectin #	Lectin Name
1	LSEctin:Mouse LSEctin 0.5
2	Rat mannose-binding lectin 1:Rat mannose-binding lectin 250
3	CD15s anti-human antibody CSLEX1:CD15s anti-human antibody CSLEX1
4	Anti-human PSGL-1 antibody CHO131:Anti-human PSGL-1 antibody CHO131
5	LSEctin:Mouse LSEctin 200
6	Normal mouse sera C57/BL6 strain:Normal mouse sera C57/BL6 strain
7	3P2 antibody to bovine submaxillary mucin:3P2 antibody to bovine submaxillary mucin
8	3P3 antibody to bovine submaxillary mucin:3P3 antibody to bovine submaxillary mucin
9	3P6 antibody to bovine submaxillary mucin:3P6 antibody to bovine submaxillary mucin
10	3P4 antibody to bovine submaxillary mucin:3P4 antibody to bovine submaxillary mucin
11	3P5 antibody to bovine submaxillary mucin:3P5 antibody to bovine submaxillary mucin
12	Aspergillus fumigatus lectin (AFL):Aspergillus fumigatus lectin (AFL) 0.01 ug/ml
13	Aleuria aurantia lectin (AAL):Aleuria aurantia lectin (AAL) 0.01 ug/ml
14	Clostridium difficile toxin A:TcdA (amino acids 1832-2710)
15	Ralstonia solanacearum lectin (RSL):Ralstonia solanacearum lectin (RSL) 0.005 ug/ml
16	Tap1:Tap1
17	SpGH101 putative glycoside hydrolase:SpGH101 100 ug/ml
18	Vacuolating cytotoxin VacA:Anti-VacA antibody
19	Vacuolating cytotoxin VacA:Vacuolating cytotoxin (VacA) p88, pH 3.0
20	Vacuolating cytotoxin VacA:Vacuolating cytotoxin (VacA) p88, pH 8.0
21	Hemagglutinin serotype 2 (H2N8):Hemagglutinin serotype 2 (HA2[8]) from Influenza A grown in HEK293T cells
22	Hemagglutinin serotype 7 (H7N2):Hemagglutinin serotype 7 (HA7[2]), from Influenza A grown in HEK293T cells
23	Galectin-1:Anti-human Fc antibody
24	Galectin-9:Galectin-9 2ug/ml
25	Galectin-1:Galectin-9 N-term CRD-Galectin-9 linker-Galectin-1 C-term CRD (9-9-1) 200ug/ml
26	Galectin-1:Galectin-1 N-term CRD-Galectin-9 linker-Galectin-1 C-term CRD (1-9-1) 200ug/ml
27	Galectin-1:Galectin-1 N-term CRD-Galectin-9 linker-Galectin-9 C-term CRD (1-9-9) 200ug/ml
28	Galectin-1:Galectin-1 dimer joined by rigid helical linker L9 2ug/ml
29	Galectin-1:Galectin-1 dimer joined by Gly-Gly linker 2ug/ml
30	Galectin-1:Galectin-1 2ug/ml
31	Galectin-1:Galectin-1 N-term CRD-Galectin-9 linker-Galectin-1 C-term CRD (1-9-1) 0.2ug/ml
32	Tap1:Tap1 5ug/ml
33	Sulfo-Sialyl-Lewis x antibody 1 0.5ug/ml:Sulfo-Sialyl-Lewis x antibody 1 0.5ug/ml
34	Sulfo-Sialyl-Lewis x antibody 2 0.1ug/ml:Sulfo-Sialyl-Lewis x antibody 2 0.1ug/ml

35 Antibody BD6 made to rat basophilic leukemia cells (RBL-2H3) 10 µg/ml:Antibody BD6 made  
 to rat basophilic leukemia cells (RBL-2H3) 10 µg/ml  
 36 Galectin-1:Galectin-1 (W69 mutant) fused with human Ig Fc + anti-goat IgG  
 37 Galectin-1:Galectin-1 (WT) fused with human Ig Fc + anti-goat IgG  
 38 Galectin-1:Galectin-1 (W69 mutant) fused with human Ig Fc  
 39 Galectin-1:Galectin-1 (WT) fused with human Ig Fc  
 40 56R+Vlx -2AA:56R+Vlx -2AA  
 41 56R+Vlx +2AA:56R+Vlx +2AA  
 42 56R+Vlx:56R+Vlx  
 56R+Vk38c antibody single chain variable fragment:56R+Vk38c antibody single chain variable  
 43 fragment  
 44 56R/76R + Vk38c antibody :56R/76R + Vk38c antibody  
 MW1 antibody made to DRPLA-19Q (Huntingtin):MW1 antibody made to DRPLA-19Q  
 45 (Huntingtin)  
 JAA-F11 antibody to Thomsen-Friedenreich disaccharide-0.1:JAA-F11 antibody to Thomsen-  
 46 Friedenreich disaccharide-0.1  
 47 Influenza A hemagglutinin H5N1:Influenza A Chicken H5N1 hemagglutinin  
 48 Influenza A hemagglutinin H1N1:Influenza A California H1N1 hemagglutinin  
 49 HA33- HA component of B. neurotoxin:HA33- HA component of B. neurotoxin  
 50 Flo1:Flocculin-1 (FLO1), from *S. cerevisiae*-GST  
 51 Nictaba:Nictaba, recombinant wild type  
 52 Nictaba:Nictaba, wild type  
 53 Nictaba:Nictaba, recombinant mutant 2  
 Influenza A/Oklahoma/3052/2009 (H1N1), Alexa488 at 3.12 µg:Influenza  
 54 A/Oklahoma/3052/2009 (H1N1), Alexa488 at 3.12 µg  
 A/Oklahoma/483 Alexa488 labeled, pH5, 37°C+4°C:A/Oklahoma/483 Alexa488 labeled, pH5,  
 55 37°C+4°C  
 56 hPIV1 Alexa488 labeled:hPIV1 Alexa488 labeled  
 A/Oklahoma/483 Alexa488 labeled, pH5, 37°C+4°C with inhibitor:A/Oklahoma/483 Alexa488  
 57 labeled, pH5, 37°C+4°C with inhibitor  
 hPIV1 Alexa488 labeled after H3N2 unlabeled at pH 5:hPIV1 Alexa488 labeled after H3N2  
 58 unlabeled at pH 5  
 hPIV2 Alexa488 labeled after H3N2 unlabeled at pH5:hPIV2 Alexa488 labeled after H3N2  
 59 unlabeled at pH5  
 Influenza A/Oklahoma/447/08 (H1N1) Alexa 488 labeled after unlabeled H1N1 at pH  
 60 5:Influenza A/Oklahoma/447/08 (H1N1) Alexa 488 labeled after unlabeled H1N1 at pH 5  
 Influenza A/Oklahoma/1138/08 (H1N1) oseltamivir-resistant Alexa488 labeled, pH7:Influenza  
 61 A/Oklahoma/1138/08 (H1N1) oseltamivir-resistant Alexa488 labeled, pH7  
 Influenza A/Oklahoma/447/08 (H1N1), Alexa488 labeled, pH7:Influenza A/Oklahoma/447/08  
 62 (H1N1), Alexa488 labeled, pH7  
 Influenza A/Oklahoma 483/08 (H3N2) Alexa488 labeled after H3N2 unlabeled at  
 63 pH5:Influenza A/Oklahoma 483/08 (H3N2) Alexa488 labeled after H3N2 unlabeled at pH5

- 64 Influenza A/Oklahoma/483/08 (H3N2) Alexa488 labeled, pH 7:Influenza A/Oklahoma/483/08 (H3N2) Alexa488 labeled, pH 7
- 65 hPIV2 Alexa488 labeled after H1N1 unlabeled at pH 5:hPIV2 Alexa488 labeled after H1N1 unlabeled at pH 5
- 66 hPIV1 Alexa488 labeled after H1N1 unlabeled at pH 5:hPIV1 Alexa488 labeled after H1N1 unlabeled at pH 5
- 67 hPIV2 Alexa488 labeled:hPIV2 Alexa488 labeled
- 68 Influenza A/ Oklahoma/3052/2009 (H1N1), Alexa488 at 50 µg, constant volume:Influenza A/ Oklahoma/3052/2009 (H1N1), Alexa488 at 50 µg, constant volume
- 69 Influenza A/Oklahoma/3052/2009 (H1N1), Alexa488 at 6.25 µg:Influenza A/Oklahoma/3052/2009 (H1N1), Alexa488 at 6.25 µg
- 70 Dog Galectin-3:Dog Galectin-3 50ug
- 71 Dog Galectin-7:Dog Galectin-7 50ug
- 72 Dog Galectin-4:Dog Galectin-4 0.5ug
- 73 Dog Galectin-2:Dog Galectin-2 0.5ug
- 74 SLT lectin:SLT lectin
- 75 Aurantioporus croceus lectin (ACL) purified from fungus- 0.1:Aurantioporus croceus lectin (ACL) purified from fungus- 0.1

**Glycan**

#

**Glycan Name**

- 1 Gala-Sp8
- 2 Glca-Sp8
- 3 Mana-Sp8
- 4 GalNAca-Sp8
- 5 Fuca-Sp8
- 6 Fuca-Sp9
- 7 Rha-Sp8
- 8 Neu5Aca-Sp8
- 9 Neu5Aca-Sp11
- 10 Neu5Acb-Sp8
- 11 Galb-Sp8
- 12 Glcb-Sp8
- 13 Manb-Sp8
- 14 GalNAcb-Sp8
- 15 GlcNAcb-Sp0
- 16 GlcNAcb-Sp8
- 17 GlcN(Gc)b-Sp8
- 18 Galb1-4GlcNAcb1-3(Galb1-4GlcNAcb1-6)GalNAca-Sp8
- 19 GlcNAcb1-3(GlcNAcb1-4)(GlcNAcb1-6)GlcNAc-Sp8
- 20 [3OSO3][6OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0
- 21 [3OSO3][6OSO3]Galb1-4GlcNAcb-Sp0
- 22 [3OSO3]Galb1-4(Fuca1-3)[6OSO3]Glc-Sp0

- 23 [3OSO3]Galb1-4Glc-*Sp8*
- 24 [3OSO3]Galb1-4[6OSO3]Glc-*Sp0*
- 25 [3OSO3]Galb1-4[6OSO3]Glc-*Sp8*
- 26 [3OSO3]Galb1-3(Fuca1-4)GlcNAc-*Sp8*
- 27 [3OSO3]Galb1-3GalNAc-*Sp8*
- 28 [3OSO3]Galb1-3GlcNAc-*Sp8*
- 29 [3OSO3]Galb1-4(Fuca1-3)GlcNAc-*Sp0*
- 30 [3OSO3]Galb1-4(Fuca1-3)GlcNAc-*Sp8*
- 31 [3OSO3]Galb1-4[6OSO3]GlcNAc-*Sp0*
- 32 [3OSO3]Galb1-4[6OSO3]GlcNAc-*Sp8*
- 33 [3OSO3]Galb1-4GlcNAc-*Sp0*
- 34 [3OSO3]Galb1-4GlcNAc-*Sp8*
- 35 [3OSO3]Galb-*Sp8*
- 36 [4OSO3][6OSO3]Galb1-4GlcNAc-*Sp0*
- 37 [4OSO3]Galb1-4GlcNAc-*Sp8*
- 38 6-H<sub>2</sub>PO<sub>3</sub>Mana-*Sp8*
- 39 [6OSO3]Galb1-4Glc-*Sp0*
- 40 [6OSO3]Galb1-4Glc-*Sp8*
- 41 [6OSO3]Galb1-4GlcNAc-*Sp8*
- 42 [6OSO3]Galb1-4[6OSO3]Glc-*Sp8*
- 43 Neu5Aca2-3[6OSO3]Galb1-4GlcNAc-*Sp8*
- 44 [6OSO3]GlcNAc-*Sp8*
- 45 Neu5Ac(9Ac)a-*Sp8*
- 46 Neu5Ac(9Ac)a2-6Galb1-4GlcNAc-*Sp8*
- 47 Mana1-3(Mana1-6)Manb1-4GlcNAc1-4GlcNAc-*Sp12*
- 48 Mana1-3(Mana1-6)Manb1-4GlcNAc1-4GlcNAc-*Sp13*
- 49 GlcNAc1-2Mana1-3(GlcNAc1-2Mana1-6)Manb1-4GlcNAc1-4GlcNAc-*Sp13*  
Galb1-4GlcNAc1-2Mana1-3(Galb1-4GlcNAc1-2Mana1-6)Manb1-4GlcNAc1-4GlcNAc-*Sp12*
- 50 Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-6)Manb1-4GlcNAc1-4GlcNAc-N(LT)AVL
- 51 Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-6)Manb1-4GlcNAc1-4GlcNAc-*Sp8*
- 52 Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-6)Manb1-4GlcNAc1-4GlcNAc-*Sp12*
- 53 Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-6)Manb1-4GlcNAc1-4GlcNAc-*Sp13*
- 54 Fuca1-2Galb1-3GalNAc1-3Gala-*Sp9*
- 55 Fuca1-2Galb1-3GalNAc1-3Gala1-4Galb1-4Glc-*Sp9*
- 56 Fuca1-2Galb1-3(Fuca1-4)GlcNAc-*Sp8*
- 57 Fuca1-2Galb1-3GalNAc-*Sp8*
- 58 Fuca1-2Galb1-3GalNAc-*Sp14*
- 59 Fuca1-2Galb1-3GalNAc1-4(Neu5Aca2-3)Galb1-4Glc-*Sp0*
- 60

61 Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcB-Sp9  
62 Fuca1-2Galb1-3GlcNAcb1-3Galb1-4GlcB-Sp8  
63 Fuca1-2Galb1-3GlcNAcb1-3Galb1-4GlcB-Sp10  
64 Fuca1-2Galb1-3GlcNAcb-Sp0  
65 Fuca1-2Galb1-3GlcNAcb-Sp8  
66 Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0  
Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-  
67 3)GlcNAcb-Sp0  
68 Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0  
69 Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8  
70 Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
71 Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
72 Fuca1-2Galb1-4GlcNAcb-Sp0  
73 Fuca1-2Galb1-4GlcNAcb-Sp8  
74 Fuca1-2Galb1-4GlcB-Sp0  
75 Fuca1-2Galb-Sp8  
76 Fuca1-3GlcNAcb-Sp8  
77 Fuca1-4GlcNAcb-Sp8  
78 Fucb1-3GlcNAcb-Sp8  
79 GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0  
80 GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0  
81 [3OSO3]Galb1-4(Fuca1-3)Glc-Sp0  
82 GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0  
83 GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8  
84 GalNAca1-3(Fuca1-2)Galb1-4GlcB-Sp0  
85 GlcNAcb1-3Galb1-3GalNAca-Sp8  
86 GalNAca1-3(Fuca1-2)Galb-Sp8  
87 GalNAca1-3(Fuca1-2)Galb-Sp18  
88 GalNAca1-3GalNAcb-Sp8  
89 GalNAca1-3Galb-Sp8  
90 GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8  
91 GalNAcb1-3GalNAca-Sp8  
92 GalNAcb1-3(Fuca1-2)Galb-Sp8  
93 GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0  
94 GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0  
95 GalNAcb1-4GlcNAcb-Sp0  
96 GalNAcb1-4GlcNAcb-Sp8  
97 Gala1-2Galb-Sp8  
98 Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0  
99 Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0  
100 Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0  
101 Gala1-3(Fuca1-2)Galb1-4GlcB-Sp0  
102 Gala1-3(Fuca1-2)Galb-Sp8

103 Gala1-3(Gala1-4)Galb1-4GlcNAcb-Sp8  
104 Gala1-3GalNAca-Sp8  
105 Gala1-3GalNAca-Sp16  
106 Gala1-3GalNAcb-Sp8  
107 Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8  
108 Gala1-3Galb1-3GlcNAcb-Sp0  
109 Gala1-3Galb1-4GlcNAcb-Sp8  
110 Gala1-3Galb1-4Glc-Sp0  
111 Gala1-3Galb-Sp8  
112 Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8  
113 Gala1-4Galb1-4GlcNAcb-Sp0  
114 Gala1-4Galb1-4GlcNAcb-Sp8  
115 Gala1-4Galb1-4Glc-Sp0  
116 Gala1-4GlcNAcb-Sp8  
117 Gala1-6Glc-Sp8  
118 Galb1-2Galb-Sp8  
119 Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0  
120 Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
121 Galb1-3(Fuca1-4)GlcNAcb-Sp0  
122 Galb1-3(Fuca1-4)GlcNAc-Sp8  
123 Galb1-3(Fuca1-4)GlcNAcb-Sp8  
124 Galb1-4GlcNAcb1-6GalNAca-Sp8  
125 Galb1-3(GlcNAcb1-6)GalNAca-Sp8  
126 Galb1-3(GlcNAcb1-6)GalNAc-Sp14  
127 Galb1-3(Neu5Aca2-6)GalNAca-Sp8  
128 Galb1-3(Neu5Aca2-6)GalNAca-Sp14  
129 Galb1-3(Neu5Acb2-6)GalNAca-Sp8  
130 Galb1-3(Neu5Aca2-6)GlcNAcb1-4Galb1-4Glc-Sp10  
131 Galb1-3GalNAca-Sp8  
132 Galb1-3GalNAca-Sp16  
133 Galb1-3GalNAcb-Sp8  
134 Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0  
135 Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0  
136 Galb1-3GalNAcb1-4Galb1-4Glc-Sp8  
137 Galb1-3Galb-Sp8  
138 Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
139 Galb1-3GlcNAcb1-3Galb1-4Glc-Sp10  
140 Galb1-3GlcNAcb-Sp0  
141 Galb1-3GlcNAcb-Sp8  
142 Galb1-4(Fuca1-3)GlcNAcb-Sp0  
143 Galb1-4(Fuca1-3)GlcNAcb-Sp8  
144 Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0  
145 Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb1-4Galb1-4(Fuca1-3)GlcNAcb-Sp0



146 Galb1-4[6OSO3]GlcB-Sp0  
147 Galb1-4[6OSO3]GlcB-Sp8  
148 Galb1-4GalNAc1-3(Fuca1-2)Galb1-4GlcNAc-Sp8  
149 Galb1-4GalNAc1-3(Fuca1-2)Galb1-4GlcNAc-Sp8  
150 Galb1-4GlcNAc1-3GalNAc-Sp8  
151 Galb1-4GlcNAc1-3GalNAc-Sp14  
152 Galb1-4GlcNAc1-3Galb1-4(Fuca1-3)GlcNAc1-3Galb1-4(Fuca1-3)GlcNAc-Sp0  
153 Galb1-4GlcNAc1-3Galb1-4GlcNAc1-3Galb1-4GlcNAc-Sp0  
154 Galb1-4GlcNAc1-3Galb1-4GlcNAc-Sp0  
155 Galb1-4GlcNAc1-3Galb1-4GlcB-Sp0  
156 Galb1-4GlcNAc1-3Galb1-4GlcB-Sp8  
157 Galb1-4GlcNAc1-6(Galb1-3)GalNAc-Sp8  
158 Galb1-3(Galb1-4GlcNAc1-6)GalNAc-Sp8  
159 Galb1-3(Galb1-4GlcNAc1-6)GalNAc-Sp14  
160 Galb1-4GlcNAc-Sp0  
161 Galb1-4GlcNAc-Sp8  
162 Galb1-4GlcB-Sp0  
163 Galb1-4GlcB-Sp8  
164 GlcNAc1-3Galb1-4GlcNAc-Sp8  
165 GlcNAc1-6Galb1-4GlcNAc-Sp8  
166 GlcNAc1-2Galb1-3GalNAc-Sp8  
167 GlcNAc1-3(GlcNAc1-6)GalNAc-Sp8  
168 GlcNAc1-3(GlcNAc1-6)Galb1-4GlcNAc-Sp8  
169 GlcNAc1-3GalNAc-Sp8  
170 GlcNAc1-3GalNAc-Sp14  
171 GlcNAc1-3Galb-Sp8  
172 GlcNAc1-3Galb1-4GlcNAc-Sp0  
173 GlcNAc1-3Galb1-4GlcNAc-Sp8  
174 GlcNAc1-3Galb1-4GlcNAc1-3Galb1-4GlcNAc-Sp0  
175 GlcNAc1-3Galb1-4GlcB-Sp0  
176 GlcNAc1-4-MDPLys  
177 GlcNAc1-4(GlcNAc1-6)GalNAc-Sp8  
178 GlcNAc1-4Galb1-4GlcNAc-Sp8  
179 (GlcNAc1-4)6b-Sp8  
180 (GlcNAc1-4)5b-Sp8  
181 GlcNAc1-4GlcNAc1-4GlcNAc-Sp8  
182 GlcNAc1-6(Galb1-3)GalNAc-Sp8  
183 GlcNAc1-6GalNAc-Sp8  
184 GlcNAc1-6GalNAc-Sp14  
185 GlcNAc1-6Galb1-4GlcNAc-Sp8  
186 GlcA1-4GlcB-Sp8  
187 GlcA1-4GlcA-Sp8  
188 GlcA1-6GlcA1-6GlcB-Sp8

189 Glcb1-4GlcB-Sp8  
190 Glcb1-6GlcB-Sp8  
191 G-ol-Sp8  
192 GlcAa-Sp8  
193 GlcAb-Sp8  
194 GlcAb1-3Galb-Sp8  
195 GlcAb1-6Galb-Sp8  
196 KDNa2-3Galb1-3GlcNAcb-Sp0  
197 KDNa2-3Galb1-4GlcNAcb-Sp0  
198 Mana1-2Mana1-2Mana1-3Mana-Sp9  
199 Mana1-2Mana1-3(Mana1-2Mana1-6)Mana-Sp9  
200 Mana1-2Mana1-3Mana-Sp9  
201 Mana1-6[Mana1-2Mana1-3]Mana1-6[Mana1-2Mana1-3]Manb1-4GlcNAcb1-4GlcNAcb-Sp12  
202 Mana1-2Mana1-6(Mana1-3)Mana1-6[Mana1-2Mana1-2Mana1-3]Manb1-4GlcNAcb1-4GlcNAcb-Sp12  
203 Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-3(Mana1-2Mana1-6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12  
204 Mana1-3(Mana1-6)Mana-Sp9  
205 Mana1-3(Mana1-2Mana1-2Mana1-6)Mana-Sp9  
206 Mana1-6(Mana1-3)Mana1-6[Mana1-2Mana1-3]Manb1-4GlcNAcb1-4GlcNAcb-Sp12  
207 Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12  
208 Manb1-4GlcNAcb-Sp0  
209 Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAc-Sp0  
210 [3OSO3]Galb1-4(Fuca1-3)[6OSO3]GlcNAc-Sp8  
211 Fuca1-2[6OSO3]Galb1-4GlcNAc-Sp0  
212 Fuca1-2Galb1-4[6OSO3]GlcNAc-Sp8  
213 Fuca1-2[6OSO3]Galb1-4[6OSO3]Glc-Sp0  
214 Neu5Aca2-3Galb1-3GalNAca-Sp8  
215 Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0  
216 Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0  
217 Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4GlcB-Sp0  
218 Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0  
219 Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8  
220 Neu5Aca2-3(6-O-Su)Galb1-4(Fuca1-3)GlcNAcb-Sp8  
221 Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp0  
222 Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp8  
223 Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0  
224 Neu5Aca2-3(Neu5Aca2-3Galb1-3GalNAcb1-4)Galb1-4GlcB-Sp0  
225 Neu5Aca2-3(Neu5Aca2-6)GalNAca-Sp8  
226 Neu5Aca2-3GalNAca-Sp8  
227 Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0  
228 Neu5Aca2-3Galb1-3[6OSO3]GlcNAc-Sp8

229 Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8  
 230 Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0  
 231 Neu5Aca2-3Galb1-3(Neu5Aca2-3Galb1-4)GlcNAcb-Sp8  
 232 Neu5Aca2-3Galb1-3[6OSO3]GalNAca-Sp8  
 233 Neu5Aca2-3Galb1-3(Neu5Aca2-6)GalNAca-Sp8  
 234 Neu5Aca2-3Galb-Sp8  
 235 Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0  
 236 Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 237 Fuca1-2[6OSO3]Galb1-4Glc-Sp0  
 238 Neu5Aca2-3Galb1-3GlcNAcb-Sp0  
 239 Neu5Aca2-3Galb1-3GlcNAcb-Sp8  
 240 Neu5Aca2-3Galb1-4[6OSO3]GlcNAcb-Sp8  
 241 Neu5Aca2-3Galb1-4(Fuca1-3)[6OSO3]GlcNAcb-Sp8  
 Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0  
 242 Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0  
 243 Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8  
 244 Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8  
 245 Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8  
 246 Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 247 Neu5Aca2-3Galb1-4GlcNAcb-Sp0  
 248 Neu5Aca2-3Galb1-4GlcNAcb-Sp8  
 249 Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 250 Fuca1-2Galb1-4[6OSO3]Glc-Sp0  
 251 Neu5Aca2-3Galb1-4Glc-Sp0  
 252 Neu5Aca2-3Galb1-4Glc-Sp8  
 253 Neu5Aca2-6GalNAca-Sp8  
 254 Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0  
 255 Neu5Aca2-6Galb1-4[6OSO3]GlcNAcb-Sp8  
 256 Neu5Aca2-6Galb1-4GlcNAcb-Sp0  
 257 Neu5Aca2-6Galb1-4GlcNAcb-Sp8  
 258 Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0  
 259 Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 260 Neu5Aca2-6Galb1-4Glc-Sp0  
 261 Neu5Aca2-6Galb1-4Glc-Sp8  
 262 Neu5Aca2-6Galb-Sp8  
 263 Neu5Aca2-8Neu5Aca-Sp8  
 264 Neu5Aca2-8Neu5Aca2-3Galb1-4Glc-Sp0  
 265 Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0  
 266 Neu5Acb2-6GalNAca-Sp8  
 267 Neu5Acb2-6Galb1-4GlcNAcb-Sp8  
 268 Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0

270 Neu5Gca2-3Galb1-3GlcNAcb-Sp0  
 271 Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0  
 272 Neu5Gca2-3Galb1-4GlcNAcb-Sp0  
 273 Neu5Gca2-3Galb1-4Glc-Sp0  
 274 Neu5Gca2-6GalNAca-Sp0  
 275 Neu5Gca2-6Galb1-4GlcNAcb-Sp0  
 276 Neu5Gca-Sp8  
 277 Galb1-3(Neu5Aca2-3Galb1-4GlcNAcb1-6)GalNAca-Sp14  
 278 Galb1-3GalNAca-Sp14  
 279 Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0  
 280 Galb1-4(Fuca1-3)[6OSO3]GlcNAc-Sp0  
 281 Galb1-4(Fuca1-3)[6OSO3]Glc-Sp0  
 282 Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0  
 283 Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0  
 284 Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0  
 285 Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0  
 286 [3OSO3][4OSO3]Galb1-4GlcNAcb-Sp0  
 287 [6OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0  
 288 6-H2PO3Glc-Sp10  
 289 Gala1-3(Fuca1-2)Galb-Sp18  
 290 Galb1-3(Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6)GalNAca-Sp14  
 291 Galb1-3Galb1-4GlcNAcb-Sp8  
 Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-  
 292 4GlcNAcb-Sp12  
 293 Galb1-4GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4GlcNAc-Sp0  
 294 Galb1-4GlcNAcb1-3(GlcNAcb1-6)Galb1-4GlcNAc-Sp0  
 295 Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0  
 296 Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0  
 297 GalNAca-Sp15  
 298 GalNAcb1-3Galb-Sp8  
 299 GlcAb1-3GlcNAcb-Sp8  
 GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-  
 300 4GlcNAcb-Sp12  
 301 GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12  
 302 GlcNAcb1-3Man-Sp10  
 303 GlcNAcb1-4GlcNAcb-Sp10  
 304 GlcNAcb1-4GlcNAcb-Sp12  
 305 HOOC(CH3)CH-3-O-GlcNAcb1-4GlcNAcb-Sp10  
 306 Mana1-6Manb-Sp10  
 307 Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10  
 308 Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-6(Mana1-3)Mana1-6)Mana-Sp9  
 309 Mana1-2Mana1-2Mana1-3(Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6)Mana-Sp9  
 310 Neu5Aca2-3Galb1-3(Neu5Aca2-3Galb1-4GlcNAcb1-6)GalNAca-Sp14

311 Neu5Aca2-3Galb1-3(Neu5Aca2-6)GalNAca-Sp14  
 312 Neu5Aca2-3Galb1-3GalNAca-Sp14  
 Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-  
 313 4GlcNAcb1-4GlcNAcb-Sp12  
 Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-  
 314 4GlcNAcb-Sp12  
 Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-  
 315 4GlcNAcb-Sp12  
 316 Neu5Aca2-8Neu5Acb-Sp17  
 317 Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8  
 318 Neu5Gcb2-6Galb1-4GlcNAc-Sp8  
 Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-  
 319 Sp19  
 Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-  
 320 4GlcNAcb1-4GlcNAcb-Sp12  
 Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-  
 321 4GlcNAcb1-4GlcNAcb-Sp12  
 Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-3(Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-6)Manb1-  
 322 4GlcNAcb1-4GlcNAcb-Sp20  
 323 Neu5Ac(9Ac)a2-3Galb1-4GlcNAcb-Sp0  
 324 Neu5Ac(9Ac)a2-3Galb1-3GlcNAcb-Sp0  
 325 Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0  
 326 Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0  
 327 Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 328 Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0  
 329 GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0  
 330 GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 331 GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 332 (Neu5Aca2-3-Galb1-3)((Neu5Aca2-3-Galb1-4(Fuca1-3))GlcNAcb1-6)GalNAc-Sp14  
 333 GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 334 GlcNAca1-4Galb1-4GlcNAcb-Sp0  
 335 GlcNAca1-4Galb1-3GlcNAcb-Sp0  
 336 GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0  
  
 337 GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0  
 338 GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0  
 339 GlcNAca1-4Galb1-3GalNAc-Sp14  
 340 Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 341 Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 342 Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 343 Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 344 Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 345 Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12

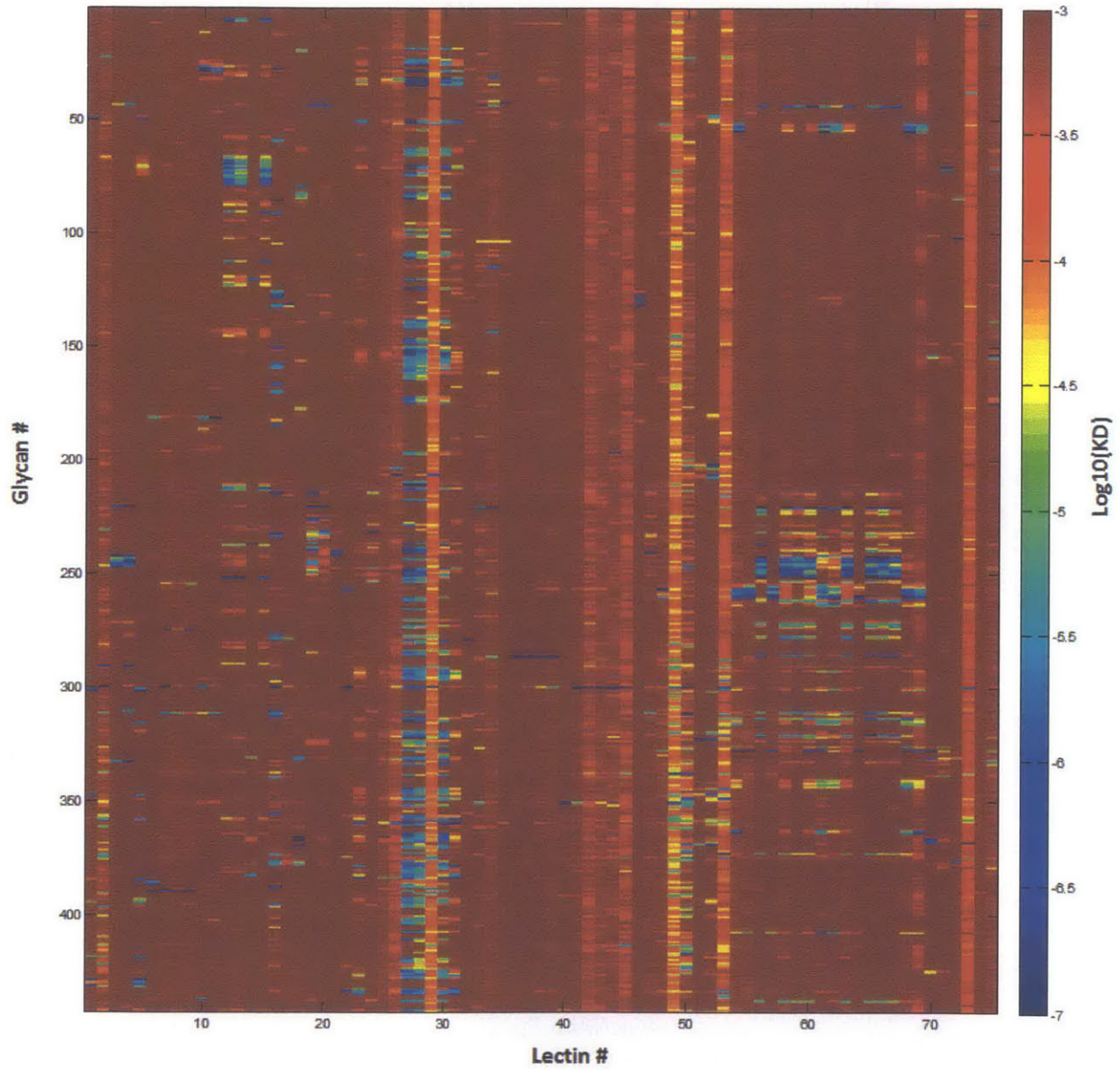
346 Galb1-4GlcNAcb1-2Mana1-3(Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12  
 347 GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 348 Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-  
 6)GlcNAcb-Sp22  
 349 Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-  
 6)GlcNAcb-Sp22  
 350 Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-  
 4GlcNAcb1-4GlcNAcb-Sp19  
 351 [6OSO3]GlcNAcb1-3Gal b1-4GlcNAc-b-Sp0  
 352 KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0  
 353 KDNa2-6Galb1-4GlcNAc-Sp0  
 354 KDNa2-3Galb1-4Glc-Sp0  
 355 KDNa2-3Galb1-3GalNAca-Sp14  
 356 Fuca1-2Galb1-3GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3GlcNAcb1-2Mana1-6)Manb1-  
 4GlcNAcb1-4GlcNAcb-Sp20  
 357 Fuca1-2Galb1-4GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-  
 4GlcNAcb1-4GlcNAcb-Sp20  
 358 Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-  
 2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp20  
 359 Gala1-3Galb1-4GlcNAcb1-2Mana1-3(Gala1-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-  
 4GlcNAcb1-4GlcNAcb-Sp20  
 360 Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12  
 361 Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-  
 4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 362 Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21  
 363 Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21  
 364 Fuca1-2Galb1-3GlcNAcb1-3(Galb1-4(Fuca1-3)GlcNAcb1-6)Galb1-4Glc-Sp21  
 365 Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-  
 4GlcNAcb1-4GlcNAc-Sp21  
 366 GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-  
 2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 367 Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-  
 6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 368 Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-  
 6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 369 GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-  
 2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 370 Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-  
 6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 371 Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-  
 2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19  
 372 Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14  
 373 Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14

374 Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14  
 (GalNAcb1-4GlcNAcb1-2Mana1-6)GalNAcb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-  
 375 4GlcNAc-Sp12  
 376 Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0  
 377 Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp0  
 378 Galb1-3GlcNAcb1-3(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6)Galb1-4Glc-Sp0  
 379 Galb1-3GlcNAcb1-3(Galb1-4(Fuca1-3)GlcNAcb1-6)Galb1-4Glc-Sp21  
 380 Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4Glc-Sp21  
 381 Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-3(Galb1-4(Fuca1-3)GlcNAcb1-6)Galb1-4Glc-Sp21  
 382 Galb1-3GlcNAcb1-3(Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6)Galb1-4Glc-Sp21  
 Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3(Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-  
 383 6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21  
 384 GlcNAcb1-2(GlcNAcb1-4)Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21  
 385 Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0  
 386 Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0  
 387 Galb1-3GlcNAcb1-3GalNAca-Sp14  
 388 Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb1-3GalNAca-Sp14  
 389 GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0  
 Gala1-3Galb1-3GlcNAcb1-2Mana1-3(Gala1-3Galb1-3GlcNAcb1-2Mana1-6)Manb1-  
 390 4GlcNAcb1-4GlcNAc-Sp19  
 Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-  
 391 6)Manb1-4GlcNAcb1-4GlcNAc-Sp19  
 Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6)Manb1-  
 392 4GlcNAcb1-4GlcNAc-Sp19  
 393 Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 394 GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 395 Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14  
 396 Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14  
 397 Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14  
 398 GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0  
 Gala1-4Galb1-3GlcNAcb1-2Mana1-3(Gala1-4Galb1-3GlcNAcb1-2Mana1-6)Manb1-  
 399 4GlcNAcb1-4GlcNAc-Sp19  
 Gala1-4Galb1-4GlcNAcb1-2Mana1-3(Gala1-4Galb1-4GlcNAcb1-2Mana1-6)Manb1-  
 400 4GlcNAcb1-4GlcNAc-LVaNKT  
 401 Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14  
 402 Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0  
 403 Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0  
 404 GalNAcb1-3Gala1-6Galb1-4Glc-Sp8  
 405 GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14  
 406 Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glc-Sp21  
 407 Neu5Aca2-6Galb1-3GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4Glc-Sp21  
 408 Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp0  
 409 Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp0

410 Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14  
 411 GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14  
 412 GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0  
 413 Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14  
 414 Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14  
 415 GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14  
 Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-  
 416 4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 Fuca1-2Galb1-4GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-  
 417 4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 GlcNAcb1-2Mana1-3(GlcNAcb1-2(GlcNAcb1-6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-  
 418 Sp19  
 419 Fuca1-2Galb1-3GlcNAcb1-3GalNAc-Sp14  
 420 Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14  
 421 GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14  
 422 Gala1-3Galb1-3GlcNAcb1-3GalNAc-Sp14  
 Fuca1-2Galb1-3GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3GlcNAcb1-2Mana1-6)Manb1-  
 423 4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-  
 424 6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2(Galb1-3GlcNAcb1-6)Mana1-6)Manb1-  
 425 4GlcNAcb1-4GlcNAcb-Sp19  
 426 Fuca1-2Galb1-3GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4Glc-Sp21  
 427 Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3GlcNAcb1-6)Galb1-4Glc-Sp21  
  
 GlcNAcb1-2Mana1-3(GlcNAcb1-4)(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21  
 428  
 GlcNAcb1-4(GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-  
 429 4GlcNAc-Sp21  
 GlcNAcb1-2Mana1-3(GlcNAcb1-4)(GlcNAcb1-6(GlcNAcb1-2)Mana1-6)Manb1-4GlcNAcb1-  
 430 4GlcNAc-Sp21  
 GlcNAcb1-4(GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(GlcNAcb1-6(GlcNAcb1-2)Mana1-6)Manb1-  
 431 4GlcNAcb1-4GlcNAc-Sp21  
 Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-2)Manb1-4GlcNAcb1-  
 432 4GlcNAc-Sp21  
 Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-  
 433 6)Manb1-4GlcNAcb1-4GlcNAc-Sp21  
 Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-  
 434 6)Manb1-4GlcNAcb1-4GlcNAc-Sp21  
 435 Galb1-4Galb-Sp10  
 436 Galb1-6Galb-Sp10  
 437 Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8  
 438 GalNAcb1-6GalNAcb-Sp8  
 439 [6OSO3]Galb1-3GlcNAcb-Sp0



440 [6OSO3]Galb1-3[6OSO3]GlcNAc-Sp0  
 Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3(Fuca1-2Galb1-4  
 441 GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-N  
 Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4)Mana1-  
 442 3(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-N



## Appendix 2.2

### Rheumatoid Arthritis Glycan Group and Similar Glycan Groups

#### RA Glycan

##### Group

Arth Glycan #	Name
301	GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12

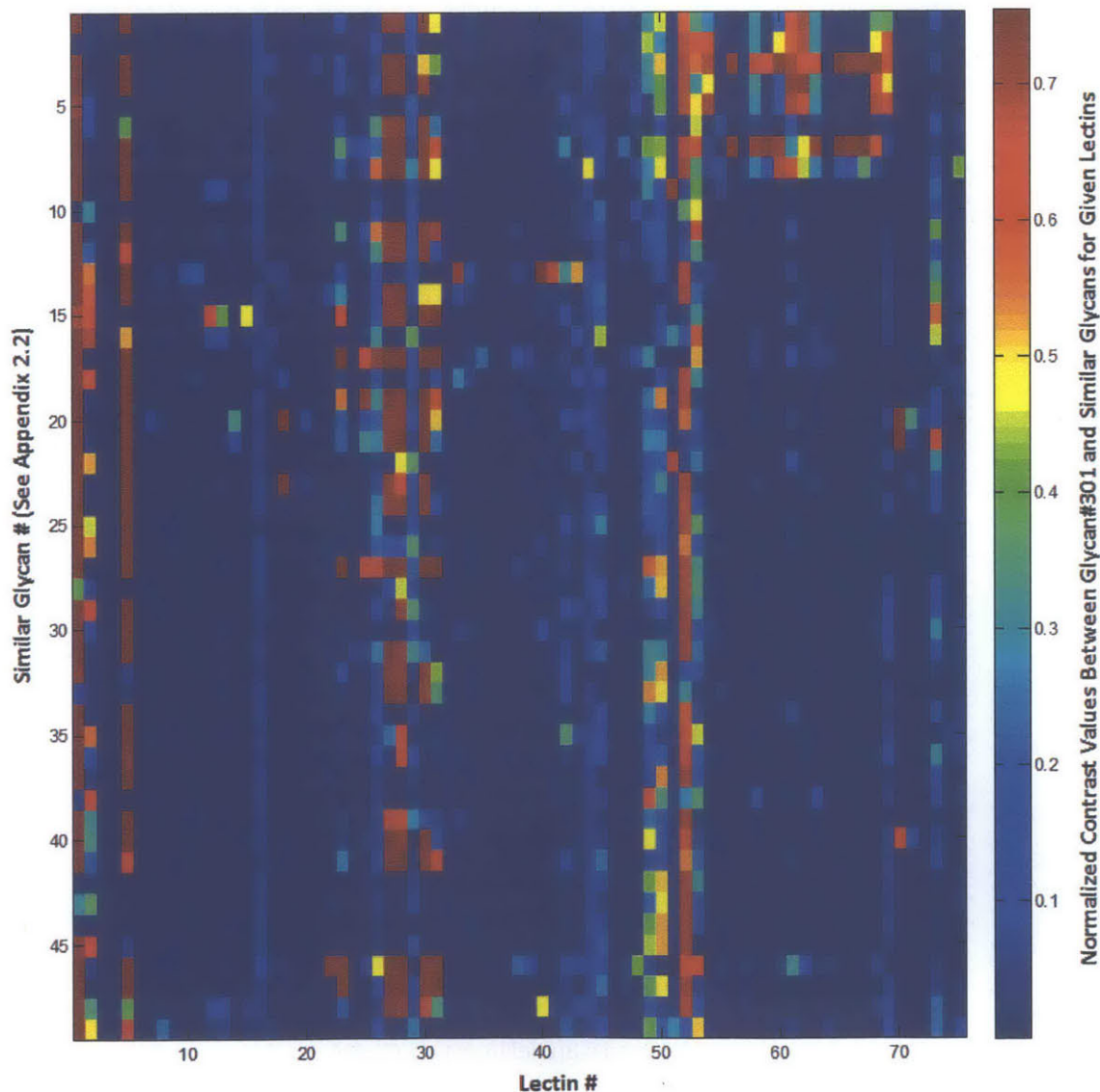
#### Similar Glycan Groups:

Sim #	DB #	Name
1	292	Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
2	300	GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
3	313	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
4	314	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
5	315	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
6	319	Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
7	320	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
8	321	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp12
9	322	Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-3(Fuca1-3(Galb1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
10	347	GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
11	348	Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
12	349	Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22
13	350	Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19
14	356	Fuca1-2Galb1-3GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
15	357	Fuca1-2Galb1-4GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
16	358	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
17	359	Gala1-3Galb1-4GlcNAcb1-2Mana1-3(Gala1-3Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20
18	361	Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22

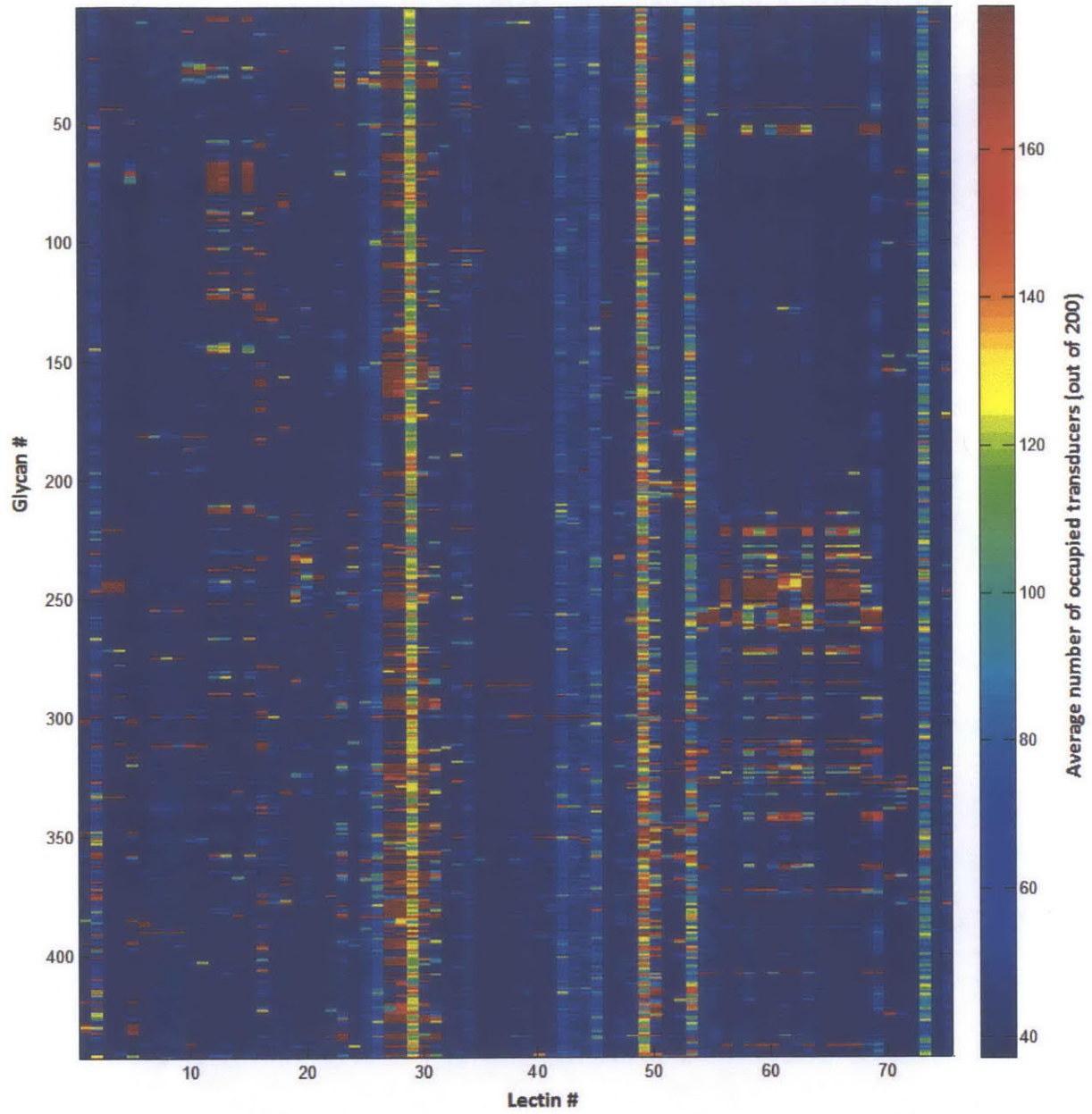
19 365 Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21  
 20 366 GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 21 367 Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 22 368 Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 23 369 GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 24 370 Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20  
 25 371 Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19  
 26 375 (GalNAcb1-4GlcNAcb1-2Mana1-6)GalNAcb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 27 383 Galb1-4GlcNacb1-2(Galb1-4GlcNacb1-4)Mana1-3(Galb1-4GlcNacb1-2(Galb1-4GlcNacb1-6)Mana1-6)Manb1-4GlcNacb1-4GlcNacb-Sp21  
 28 384 GlcNacb1-2(GlcNacb1-4)Mana1-3(GlcNacb1-2Mana1-6)Manb1-4GlcNacb1-4GlcNac-Sp21  
 29 390 Gala1-3Galb1-3GlcNAcb1-2Mana1-3(Gala1-3Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp19  
 30 391 Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp19  
 31 392 Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp19  
 32 393 Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 33 394 GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp12  
 34 399 Gala1-4Galb1-3GlcNacb1-2Mana1-3(Gala1-4Galb1-3GlcNacb1-2Mana1-6)Manb1-4GlcNacb1-4GlcNacb-Sp19  
 35 400 Gala1-4Galb1-4GlcNacb1-2Mana1-3(Gala1-4Galb1-4GlcNacb1-2Mana1-6)Manb1-4GlcNacb1-4GlcNacb-LVaNKT  
 36 416 Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 37 417 Fuca1-2Galb1-4GlcNAcb1-2Mana1-3(Fuca1-2Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 38 418 GlcNAcb1-2Mana1-3(GlcNAcb1-2(GlcNAcb1-6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19  
 39 423 Fuca1-2Galb1-3GlcNAcb1-2Mana1-3(Fuca1-2Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 40 424 Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22  
 41 425 Galb1-3GlcNAcb1-2Mana1-3(Galb1-3GlcNAcb1-2(Galb1-3GlcNAcb1-6)Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp19  
 42 428 GlcNAcb1-2Mana1-3(GlcNAcb1-4)(GlcNAcb1-2Mana1-6)Manb1-4GlcNacb1-4GlcNAc-Sp21  
 43 429 GlcNAcb1-4(GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21

- 44 430 GlcNAcb1-2Mana1-3(GlcNAcb1-4)(GlcNAcb1-6(GlcNAcb1-2)Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
- 45 431 GlcNAcb1-4(GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(GlcNAcb1-6(GlcNAcb1-2)Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
- 46 433 Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
- 47 434 Galb1-4GlcNAcb1-2Mana1-3(GlcNAcb1-4)(Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6)Manb1-4GlcNAcb1-4GlcNAc-Sp21
- 48 441 Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3(Fuca1-2Galb1-4 GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-N
- 49 442 Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4)Mana1-3(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-N

### Appendix 2.3

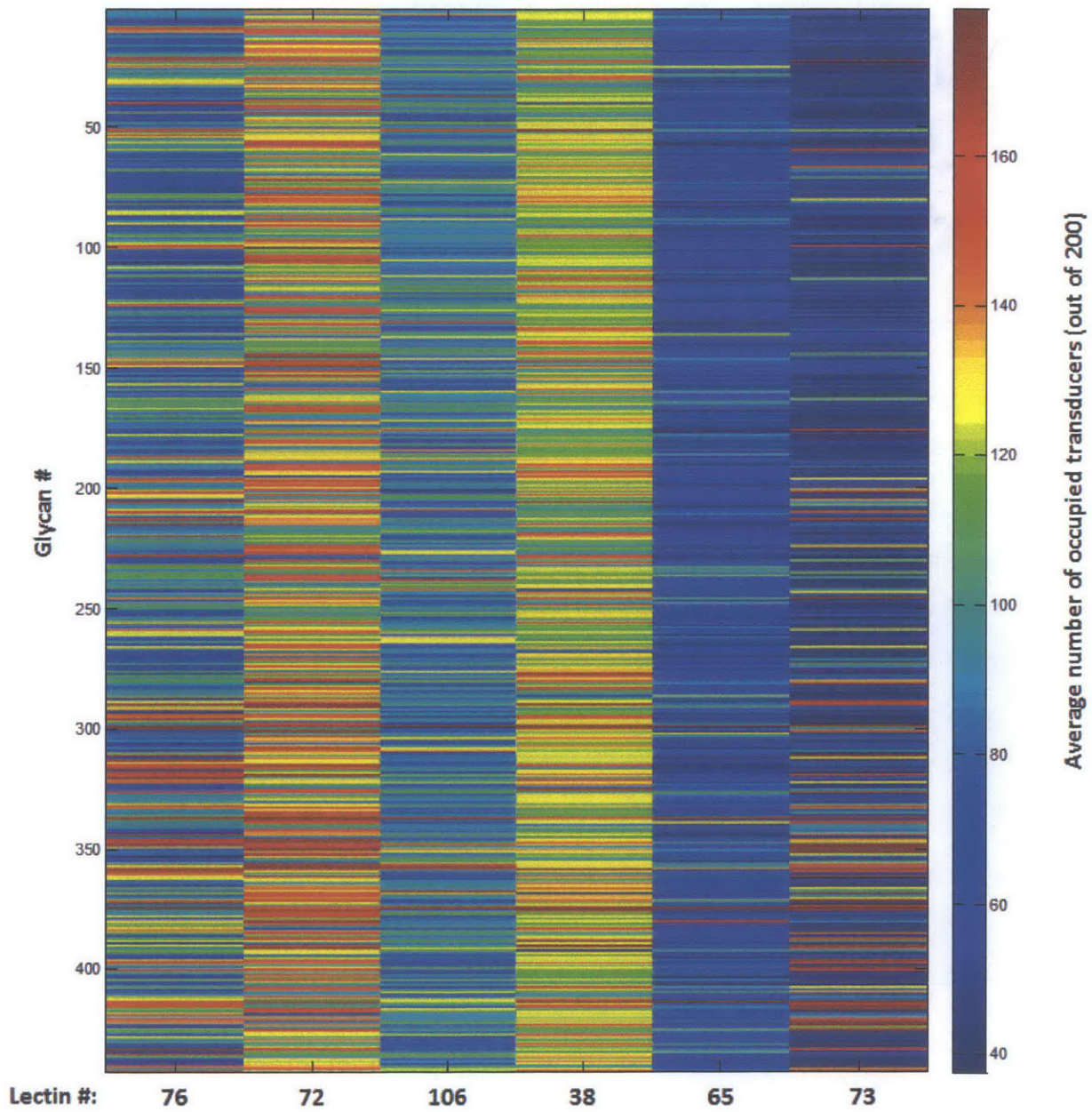


## Appendix 2.4



## Appendix 2.5

Unique 'barcode' for each of the glycans using the best 6 lectins for differentiation of single glycans:



### 3. Proof of Concept - Nanotube Sensors for Proteins and Glycans

Some of the work, text and figures presented in this chapter are reprinted or adapted from reference<sup>1</sup> (reprinted under thesis use allowance from the American Chemical Society © 2011)

#### 3.1. Introduction

Understanding the interactions between glycans and carbohydrate recognition domains (CRD) found on cellular and protein surfaces is vital to the fields of glycobiology, immunology, and pharmacology. Glycans decorating the surface of proteins substantially influence function, such as folding pathways, signaling, retention and pharmacokinetics.<sup>2,3</sup> The efficacy of protein-based therapeutics is largely dictated by their glycosylation<sup>4</sup>, and thus *de novo* design of drugs that interact with known CRD sites, such as cell-adhesion modulating galectins<sup>5</sup>, requires a greater understanding of the kinetic parameters between glycans and CRDs. Despite this importance, robust assays of protein glycosylation are under-developed, with the dominant profiling technologies falling to frontal affinity chromatography<sup>6,7</sup> and mass spectrometry.<sup>8-10</sup> Recently, the concept of the lectin microarray has emerged as a promising approach to investigating glycan-lectin interactions. These arrays take advantage of multivalent interactions to overcome the weak monovalent binding of lectin-glycan pairs (typically  $10^{-7}$  M < Kd <  $10^{-3}$  M), but are limited in their ability to transduce weak mono or multivalent interactions and also require fluorescence labeling.<sup>6,11-15</sup> An emerging concept<sup>6,11</sup> is to use multivariate responses of glycans binding to a library of lectins to discern their identity, but this requires detection methods that are necessarily sensitive enough to transduce the presence of weakly bound proteins. Label free methods, which reduce sample volume requirements, have a distinct advantage in this approach by decreasing the absolute detection limit. Herein, we develop a fluorescent single walled carbon nanotube sensor<sup>16</sup> for glycan-lectins interactions with the ultimate goal of profiling glycans.

Our approach is to couple band gap fluorescent SWNT to receptor lectins, which are a host of naturally occurring carbohydrate binding proteins.<sup>17</sup> We benchmark the sensor by evaluating the kinetic parameters between anti-His tag antibody and comparing it to literature surface plasmon resonance (SPR) parameters. We then demonstrate the detection of fucose (Fuc) to PA-IIL lectin and *N*-acetylglucosamine (GlcNAc) to GafD lectin. Kinetic parameters are obtained by first measuring the fluorescence intensity of a large spot of SWNT, and then it is shown how the same signal can be increased by probing individual SWNT sensors and determining which sensors are most responsive to glycosylated analyte addition. Our objective of glycan profiling makes this effort distinct from recent work creating glycosylated carbon nanotubes for therapeutic purposes<sup>18,19</sup> and electronic FET sensors for lectin, but not glycan, binding<sup>20-22</sup>.

In contrast, the aim of our work is to profile glycans via their selective binding to a fluorescent nanotube surface, a concept not yet demonstrated in the literature to date.

Kinetic information on glycans–CRD interactions is currently determined by two types of analytical methodology: equilibrium and non-equilibrium.<sup>23</sup> Non-equilibrium methods yield relative binding information rather than physical kinetic rates; that is, they specify which glycan–CRD combinations bind with greater or lesser affinity in reference to each other. These methods include ELISA<sup>24</sup>, glycan microarrays<sup>25</sup>, agglutination<sup>26</sup>, and electrophoresis<sup>27</sup>. A notable exception in this category of relative binding assays are some of the carbohydrate arrays from the CH Wong group that can yield quantitative fluorescent binding curves similar to SPR (below) and be translated to  $K_D$  values.<sup>15</sup> Equilibrium dialysis can determine the forward reaction rate ( $k_f$ ) of glycan–CRD binding but requires a large amount of glycan reagent, which is often a significant investment of time and money if complex sugars are used. Equilibrium titration calorimetry<sup>28,29</sup> is a delicate technique to determine kinetic parameters from thermodynamic information, but is rarely employed because of time and reagent expenses. Another equilibrium technique, frontal affinity chromatography<sup>30</sup>, can be used to determine the affinity constant ( $K_D$ ) for most glycan–CRD pairs ( $K_D > 10^7$ ) however the glycan must be labeled for detection. The current standard for obtaining kinetic information from label-free groups is surface plasmon resonance (SPR) machines, such as the Biacore © systems. In the case of glycan–CRD interactions, SPR can detect both the forward and reverse kinetic rates for a wide range of affinities ( $K_D$ : mM–pM range). However, to induce a detectable signal the analyte must have significant mass. Thus glycans are typically immobilized on the gold surface (often using neoglycoproteins<sup>31</sup>, i.e. glycans synthetically coupled to a protein backbone) and the more substantial lectins are used as the binding analytes. This can bias the analysis of single lectin–glycan interactions as presentation and density of the glycan is a critical parameter in CRD binding and the immobilization methodology can alter this.<sup>32</sup> Though more difficult, lectins can also be immobilized on the SPR surface and detect glycosylated analytes, but again these must have enough mass to transduce a change in refractive index.

We show that the SWNT-based fluorescence sensors developed in this work demonstrate loading curve signals competitive with SPR, both in shape and analysis technique, but they differ in a few significant ways. First, the detection scheme is reversed. The lectins are the tethered sensors and the glycans are the analyte in solution. This allows us to determine the kinetics of free glycans as well as glycoproteins giving us precise control over carbohydrate presentation in the interaction. Second the amount of analyte needed for each experiment (2  $\mu$ g of glycosylated protein or 100 ng of free glycan) is smaller than what is necessary for SPR experiments, which require analyte flow to overcome mass-transport effects (at its best, SPR requires 20–250  $\mu$ g of analyte protein<sup>33</sup>). Third, each SWNT-sensor spot



can be tethered to different lectins and illuminated simultaneously, creating the potential for a multiplexed, quantitative detection of analyte binding analogous to an array-reading SPR machine with distinct advantages in sample size and run time.

## 3.2 Methods

### 3.2.1 Lectin Expression

The resulting plasmid pET41GafD and pET41-PAILL were transformed into BL21(DE3)<sup>TM</sup> Star (Invitrogen) according to standard procedures.<sup>34</sup> A single colony was used to inoculate 5 mL LB medium containing kanamycin (50  $\mu\text{g mL}^{-1}$ ). Three milliliter of the overnight culture at 37 °C was used as an inoculum to a 350 mL flask of LB containing 50  $\mu\text{g mL}^{-1}$ , and this was incubated again with shaking at 37 °C. Heterologous protein was induced by the addition of isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG, final concentration 1 mM) once this culture had reached log phase ( $A_{600}$  of 0.6). Growth was continued for 6 h before the cells were harvested by centrifugation. For His-tag protein purification, harvested cell pellet was washed twice with phosphate-buffered saline, PBS (10 mM, pH 7.4) and then lysed with Complete Lysis-B (Roche Applied Science). The crude lysate was clarified by centrifugation prior to application to a 3 mL Ni-NTA agarose column (Qiagen). Non-specifically bound proteins removed from the column with wash buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , pH 8.0, 300 mM NaCl, 20 mM imidazole) and bound His-tag GafD and PAILL were eluted with elution buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , pH 8.0, 300 mM NaCl and 250 mM imidazole). Eluted samples were analyzed by 15% SDS-PAGE and protein concentration was determined with BCA assay kit according to manufacturer's instructions (Pierce). To change the buffer with PBS (10 mM, pH 7.4), the eluted solution was centrifuged through a centrifugal filter with a molecular cutoff of 10 kDa (Millipore) and the concentration of lectin was finally adjusted to 4  $\text{mg mL}^{-1}$ .

### 3.2.2 Glycan and Model Glycoprotein Probes

Biotinylated glycans were provided by the Consortium of Functional Glycomics – Scripps Institute Group. For this work, glycans B121 (GlcNAc $\beta$ -SpNH-LC-LC-Biotin) and B158 were used (Fuc $\alpha$ 1-2Gal $\beta$ 1-4Glc $\beta$ -SpNH-LC-LC-Biotin) where Sp is 2-azidoethyl and LC-LC biotin is a standard biotin reagent with an extra-long spacer group (Pierce ID# 21343). The lyophilized sugars were dissolved in 3mL of 1x PBS to create stock solutions and stored at -20°C. To construct model glycoprotein probes, the biotinylated glycans were incubated with streptavidin (Sigma Aldrich SO677) for 1 hour at 20°C in a 6:1 molar ratio to allow maximum binding to the four biotin binding sites on each streptavidin. Excess

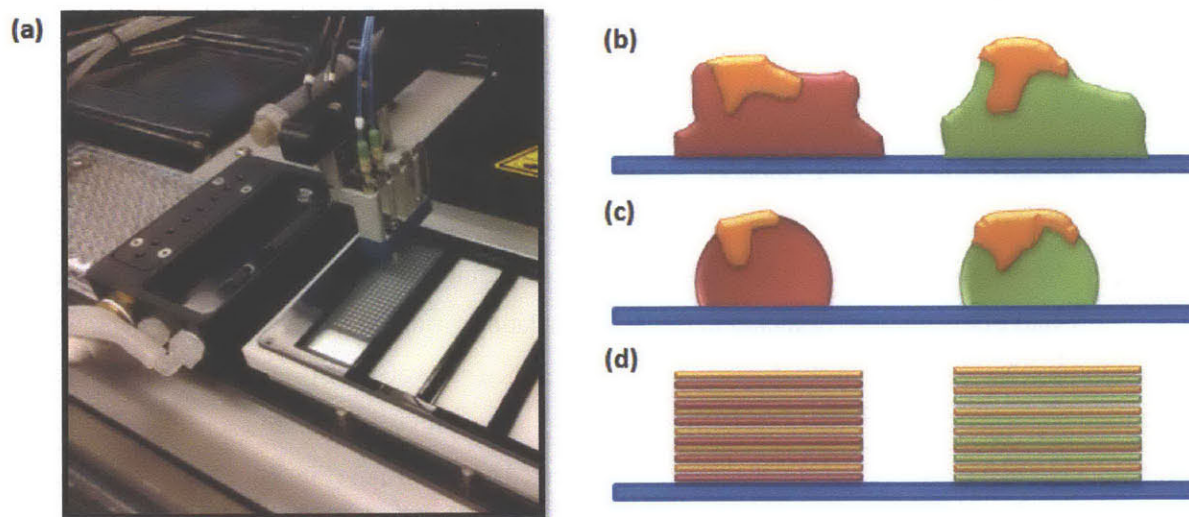
biotinylated glycans were filtered away from the glycoproteins by centrifugation through an Amicon filter (16,300 x g on Labnet Inc centrifuge, 10 min, 30,000 kDa cutoff, Milipore). The glycoproteins were washed on the filter 3x (400  $\mu$ L PBS) and then resuspended in PBS at the desired concentrations.

### **3.2.3 Native PAGE Binding Analysis**

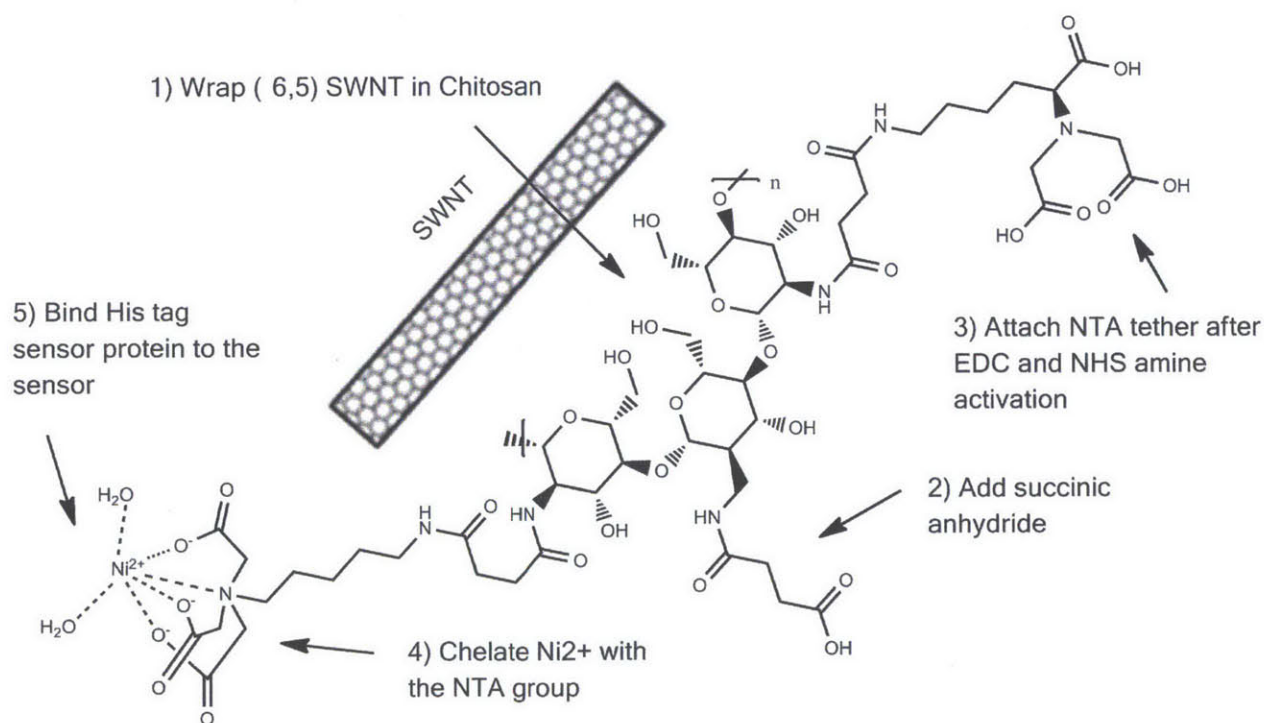
The interaction between lectin and glycan was analyzed by 15% native PAGE according to the method of Schagger and von Jagow with minor modifications.<sup>35</sup> Briefly, electrophoresis was performed using a Mini Protean Electrophoresis system (Bio-Rad, USA) under nondenaturing conditions to examine molecular interactions. Nonreduced protein/glycan samples (GafD Lectin held constant at 30 $\mu$ g per lane, GlcNAc-Strept added at 5, 10, 20, and 30 $\mu$ g) in the sample buffer (20 % glycerol, 200 mM Tris-HCl, pH 6.8, 0.05% bromophenol blue) were applied to the gel (gel buffer: 25 mM Tris-HCl, 200 mM glycine). Electrophoresis was performed at 80 V for 120 min. After electrophoresis, the protein bands were visualized by staining with Coomassie Brilliant Blue R-250.

### **3.2.4 Construction of Chitosan-SWNT Sensor Chips**

To increase the reproducibility of our sensors, we have introduced a careful automated printing method of the chitosan gel (Fig 3.1). Patterned glass microscope slides (Tekdon) were inserted in a microarray printer (Digilab MicroSys System) where the robotic head was programmed to dispense alternating layers of chitosan-SWNT (0.25wt% chitosan (CHI), 1 vol% acetic acid, 30 $\mu$ g/ml suspended (6,5) SWNT) and crosslinker (10 vol% glutaraldehyde). The suspended SWNT was made from Southwest Nanotechnologies, Inc. CoMoCAT® nanotubes sonicated in 0.25wt% chitosan (CHI) and 1 vol% acetic acid for 45 minutes at 40% amplitude with a probe-tip sonicator (Cole Parmer, Model CV18). For each sensor spot ten alternating layers of SWNT-CHI and cross-linker were printed at 100nL per layer, resulting in a highly-uniform gel of 1 $\mu$ l SWNT-CHI material. The chips were printed in a humidified enclosure (85% RH) at 25°C and allowed to cross-link overnight in the same environment. Nickel-NTA groups were introduced as previously reported<sup>16</sup>. Briefly, the chips were washed with a dilute basic buffer (0.01 M NaOH) and water three times. Carboxylic acid groups were introduced to the chitosan wrapped SWNT by bathing the chips in succinic anhydride (0.1 M) overnight (Fig 3.2 step 2).



**Figure 3.1** - An automated printer (a) is now used to print the SWNT-Chitosan gel. When the gel is printed by hand (b) the gel morphology is irregular and the crosslinking is inhomogenous leading to different spot sizes and intensities after subsequent chemistry steps. Automated printing of large spots with subsequent addition of cross-linker on top of the gel (c) also leads to inhomogenous spots. By printing many alternating layers of SWNT-Chitosan gel and crosslinker (d) we can build a spot that is homogenous in shape and signal.

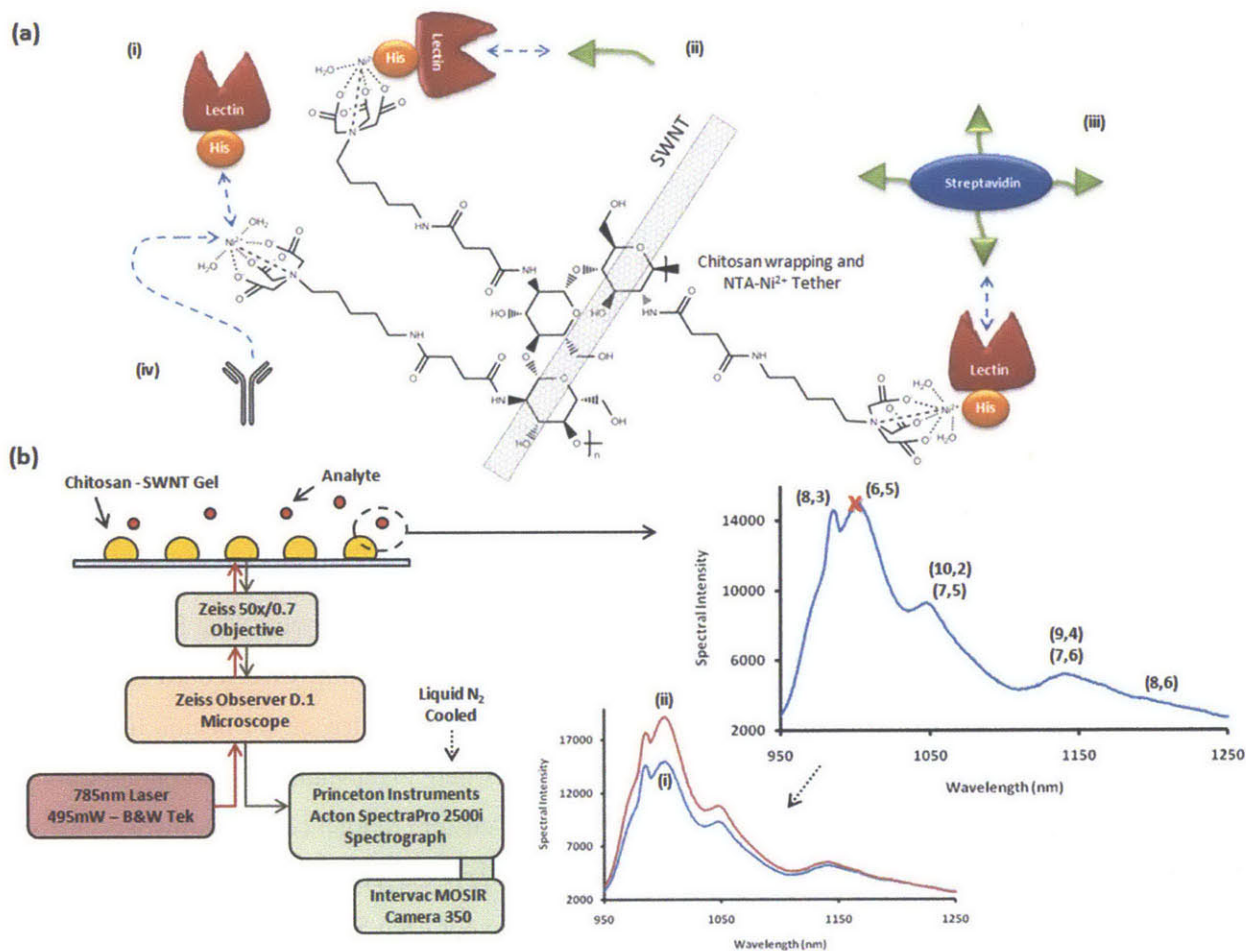


**Figure 3.2** - Processing steps to tether a His tag sensor protein to SWNT: 1) wrap SWNT in chitosan, 2) functionalize chitosan with carboxylic acids, 3) attach tethered NTA group, 4) chelate nickel with NTA, 5) add His tag protein to sensor.

The chips were then washed three times with water and the carboxylic acid groups were activated via *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC – 0.1M) and *N*-hydroxysuccinimide (NHS – 0.1M). The chips were bathed in this solution for 2 hours at 25°C and examined for the expected formation of bubbles. The chips were again washed and allowed to bathe in a solution of a linked tricarboxylic acid group ( $N\alpha, N\alpha$ -bis(carboxymethyl)-L-lysine a.k.a. NTA - 33mM) overnight (Fig 3.2 step 3). The chips were washed in water again and stored in a 100mM solution of  $NiSO_4$  to allow maximum binding of nickel to the NTA chelating groups (Fig 3.3 step 4). Thus in each experiment the SWNT sensors start in their maximally quenched state due to close proximity of the nickel. During experimentation the excess  $NiSO_4$  is washed away with water and the sensor protein, His-tagged lectin, is tethered to the SWNT sensors via the chelated nickel group (Fig 3.4 step 5). The large protein groups cause the nickel group to move away from the SWNT sensor, due to steric loading (discussed below), and part of the quenched fluorescent signal returns.

### 3.2.5 Ensemble Measurements of Sensors

A custom-made near infrared inverted microscope (Zeiss D.1 Observer) setup allows us to probe the fluorescent emissions of our SWNT sensors (Fig 3.3a). The chips are secured on the microscope stage and the objective (50x/0.7 Zeiss) is pushed into contact with a blank portion of the glass slide (no SWNT-CHI) to obtain a 5 second background spectrum, which is subsequently subtracted from the response spectra. The objective is then moved under a SWNT-CHI gel spot and again pushed in contact with the glass slide. By placing the objective in the maximum z-axis position, the microscope can image a higher plane of the SWNT-CHI gel where more analyte response is observed. The SWNT are excited by a 785nm laser (B&W Tek – 495mW) and the emission is sent to a spectrometer (Princeton Instruments Acton SpectraPro 2500i Spectrograph) and accompanying nIR camera (Intervac MOSIR Camera 350). The spectra are collected via WinSpec software (Princeton Instruments) and analyzed with custom Matlab (Mathworks) code. To maximize signal stability, the spectrometer is cooled with liquid nitrogen two hours prior to experimentation and the laser is allowed to reach peak stability for two hours. The SWNT-CHI gel has a small transient region when first exposed to the laser due to local heating and further permeation of  $Ni^{2+}$  in the gel; thus each spot is pre-exposed to the laser for 5 minutes before data is gathered. Data is gathered in the form of emission intensity spectra (950-1250nm) integrated for 5 seconds.



**Figure 3.3** – Ensemble measurements of Chitosan-SWNT sensors for glycan lectin detection. A) The chitosan wrapped SWNT are processed (see text) to include tethered NTA groups and chelated Ni<sup>2+</sup> so that His-tagged lectins (i) can attach to the sensors. An analyte (anti His-tag antibody (iv), free biotinylated glycan (ii), or glycan tethered to streptavidin (iii)) is added and the emission fluorescence is increased (Part B (i) to (ii)) due to the Ni<sup>2+</sup> group moving away from the SWNT, caused by steric loading of the sensor (see text). B) Ensemble measurement setup: the chitosan-SWNT gel is spotted onto glass chips which are excited by a 785nm laser in a custom inverted microscope setup. The resulting emission spectra are then analyzed looking at the intensity of the (6,5) nanotube peak over time.

A typical experiment was run for approximately 1000 frames (at 5 seconds each) and included a few addition and washing steps to detect lectin-glycan binding (Figure 1b). First NiSO<sub>4</sub> was again added to ensure that the SWNT sensors were responsive and that the NTA chelating groups were fully loaded with Ni<sup>2+</sup> groups. The nickel was then washed away with PBS three times, leaving ~20ul of PBS on the sensors. The His-tag lectin was then added to the sensor (20μl at 2mg/ml) and allowed to bind for 300 seconds. Excess lectin was again washed by PBS three times and 20 ± 1 μl PBS was left to bathe the sensor. The sensor was allowed to equilibrate for 100 seconds and 20μl of analyte was added. The

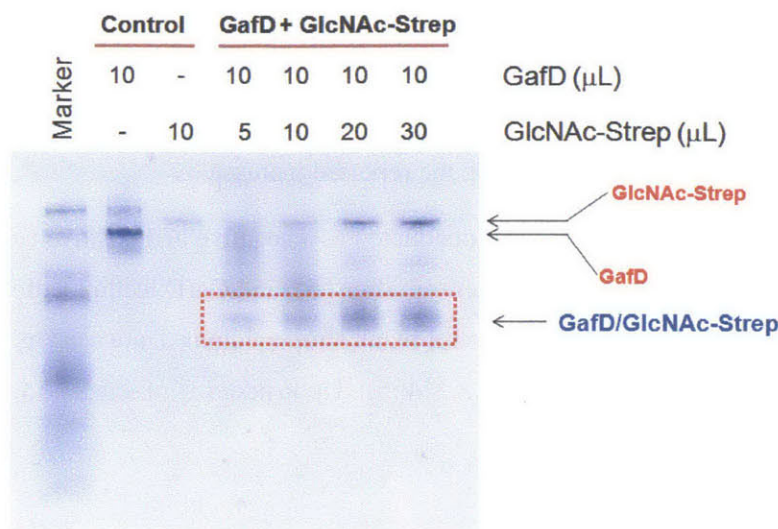
analytes tested in this paper include free biotinylated glycans, glycans tethered to streptavidin, and anti His-tag antibody. Each of these, upon binding, causes an increase in SWNT luminescence (Figure 1b). The sensor response was recorded for 500 seconds. It is essential to record stabilization frames before and after the analyte addition in order to correct for any focal drift caused by the tension of the objective in contact with the glass slide (see Results section).

### 3.2.6 Single SWNT Sensor Measurements

A second custom microscope was used to collect emission intensities of single SWNT sensors. The SWNT-CHI gel was diluted to 3 $\mu$ g/ml of SWNT and spin-coated (3000 rpm for 30s on Laurell Technologies Corporation, WS-650MZ-23NPP/LITE) on glass-bottom petridishes (MarTec Corp). The petri-dishes were then placed on the microscope platform (Zeiss D.1 Observer) and the oil-immersed objective (Zeiss 100x/1.46) was focused on the SWNT sensors on the glass surface. The SWNT were excited by a 660nm laser (Crystal Laser – 100mW) and the emission intensities were recorded by a nitrogen-cooled InGaAs array (Princeton Instruments). Again WinSpec software (Princeton Instruments) was used to collect the data in the form of a stacked Tiff image where pixel values corresponded to spectral intensity. These Tiff images were then analyzed using custom Matlab code (Mathworks) to: 1) construct an intensity versus time trace for each SWNT sensor, 2) noise reduce the intensity signal, 3) fit the signal to a step-wise curve, and 4) determine the kinetic rates for each SWNT sensor according to a previously published model.<sup>36,37</sup>

## 3.3. Results and Discussion

We first confirmed that our Streptavidin-based model glycoproteins bind to our expressed His-tag lectins using native PAGE gel analysis, which allows the protein-protein complex to remain in its native, non-denatured form. Note that separation is dictated by native complex charge and morphology, not strictly molecular weight as in SDS-PAGE<sup>38</sup>. The resulting gel (Fig 3.4) clearly indicates a bound complex that arises when GlcNAc-streptavidin probe (1mg/ml) is added in solution with GafD lectin, a lectin from *Escherichia coli*, which binds  $\beta$ -GlcNAc (3mg/ml).

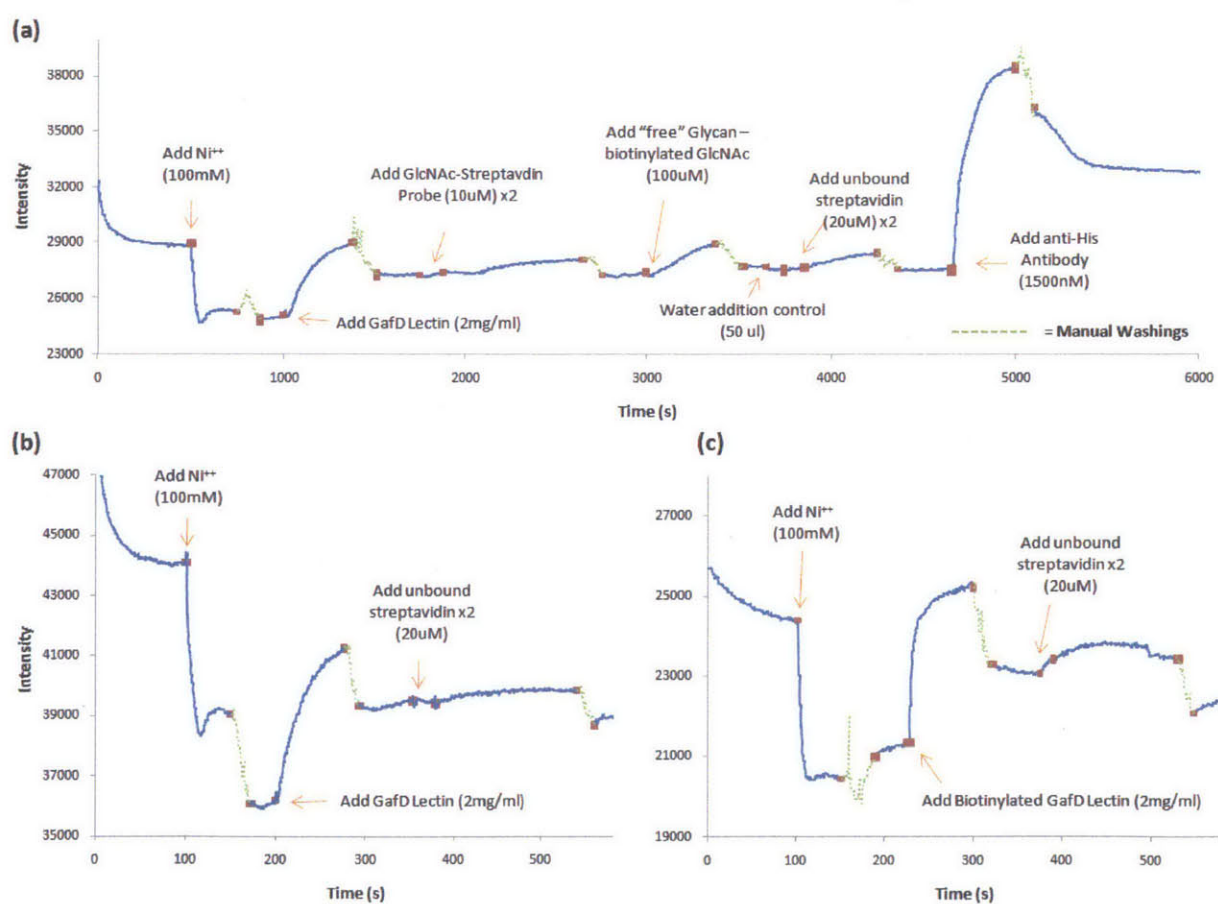


**Figure 3.4** – Native PAGE gel of GlcNAc-Streptavidin (1mg/ml) to GafD Lectin (3mg/ml) in solution. The gel reveals an increasing band of GafD-GlcNAc conjugate as more GlcNAc-Streptavidin is added to solution.

Next the time response of the sensor during construction and various analyte additions was analyzed using the ensemble measurement technique outlined in Figure 1. In the first test,  $\text{Ni}^{2+}$  was added (100mM) which caused a clear quenching response as the  $\text{Ni}^{2+}$  interacted with the exposed SWNT decreasing their fluorescence (Figure 3.5a). The effect of divalent cation quenching of nanotubes is established in the literature.<sup>39,40</sup> In brief, the SWNT exciton formed by laser excitation is affected by the electronic field of metal ion and caused to decay in non-radiative pathways; thus the fluorescent emission of the SWNT is effectively reduced as the metal ion comes in close proximity. The excess  $\text{Ni}^{2+}$  was then washed away and the His-tagged lectin GafD was added (40ug in 20ul). A loading curve (much like that observed in SPR assays) was observed as the lectin binds to the NTA- $\text{Ni}^{2+}$  complexes. The mechanism consistent with our previous analysis is that the increase in SWNT fluorescence is caused by an increase in the distance between the  $\text{Ni}^{2+}$  complex, lessening its proximity quenching effect.<sup>16</sup> Both steric loading and multivalent effects can influence this as discussed below. The excess lectin is removed and the Streptavidin-GlcNAc probe is added (40 $\mu\text{g}$  in 40 $\mu\text{l}$ ), in two steps. Again, an SPR-like loading curve is observed, except with slower kinetics and less overall response. The excess Streptavidin-GlcNAc probe is washed away and free biotinylated GlcNAc is added (20 $\mu\text{l}$  of 100 $\mu\text{M}$ ). Again the loading curve is observed. After washing away the excess free glycan, we added 40  $\mu\text{l}$  of PBS to the system to ensure that the loading responses were due to the analyte and not some focal change due to increased mass on the sensor. We then checked to see if the biotinylated glycans, now bound to the sensor, were accessible to blank streptavidin. Upon streptavidin addition we see another loading curve, with slower kinetics,

confirming binding of streptavidin to the biotin ends of glycans decorated on the SWNT-Lectin sensors. As a positive control, we used an anti His-tag antibody, as previously<sup>16</sup> to verify the integrity of His-tag binding at the end of the experiment. Here, binding to the His-tagged lectin displaces it and increases the distance between the SWNT and Ni<sup>2+</sup>, causing the reported increase.

Two other time-series analyses were conducted as a negative and positive control. In the negative control, blank streptavidin was added to the sensor, after Ni<sup>2+</sup> and GafD lectin loading (Fig 3.5b). No loading curve is observed. In the positive control blank streptavidin is again used as the analyte but this time the GafD lectin is biotinylated (Pierce Kit 21455). Upon addition of streptavidin we see the expected loading curve (Fig 3.5c).



**Figure 3.5** – Time response curves of sensor chemistry steps and analyte additions. A) Responses during construction of a GafD lectin sensor and responses to free GlcNAc, GlcNAc-Streptavidin, unconjugated streptavidin, and anti His-tag antibody. Each of the additions with known affinity exhibit a loading similar to those found in SPR experiments. B) Negative control: adding unconjugated streptavidin to GafD lectin results in negligible response. C) Positive control: adding unconjugated streptavidin to biotinylated GafD lectin.



The time-series analyses reveal activation limited kinetic responses as expected for our free glycan and streptavidin-tethered probes, since the response rates are highly dependent upon the analyte, and therefore not limited by diffusion through the chitosan matrix. To obtain forward and reverse kinetic rates of binding we analyzed the loading curves of the sensors at varying analyte concentrations. Assuming our reaction model is that of a single-site surface absorption:



where G is the glycosylated analyte, L is the lectin binding sites and GL is the bound complex, than we would expect the following kinetic model to express the rate of change of bound complex:

$$\frac{d[GL]}{dt} = k_f[G][L] - k_r[GL] \tag{3.2}$$

The concentration of the free Lectin sites at a given time can be expressed as:

$$[L]_t = [L]_{t_0} - [GL]_t \tag{3.3}$$

Thus equation 3.2 can be rewritten as:

$$\frac{d[GL]}{dt} = k_f[G]([L]_{t_0} - [GL]_t) - k_r[GL] \tag{3.4}$$

We assume that the analyte concentration, [G], is constant (as the bulk of fluid above the sensor is large in comparison to the number of lectin binding sites). The change in our fluorescent intensity (I) is a measure of the change in bound complex [GL], so an analogous form of equation 4 for our sensor system would be:

$$\frac{dI}{dt} = k_f C_g (I_{max} - I_t) - k_r(I_t) \tag{3.5}$$

This equation can be rearranged as to lump the kinetic parameters together into one kinetic variable ( $k_s$ ) as is done in fitting SPR data<sup>41</sup>:

$$\frac{dI}{dt} = k_f C_g I_{max} - k_s(I_t) \quad (3.6)$$

$$k_s = k_f C_g + k_r \quad (3.7)$$

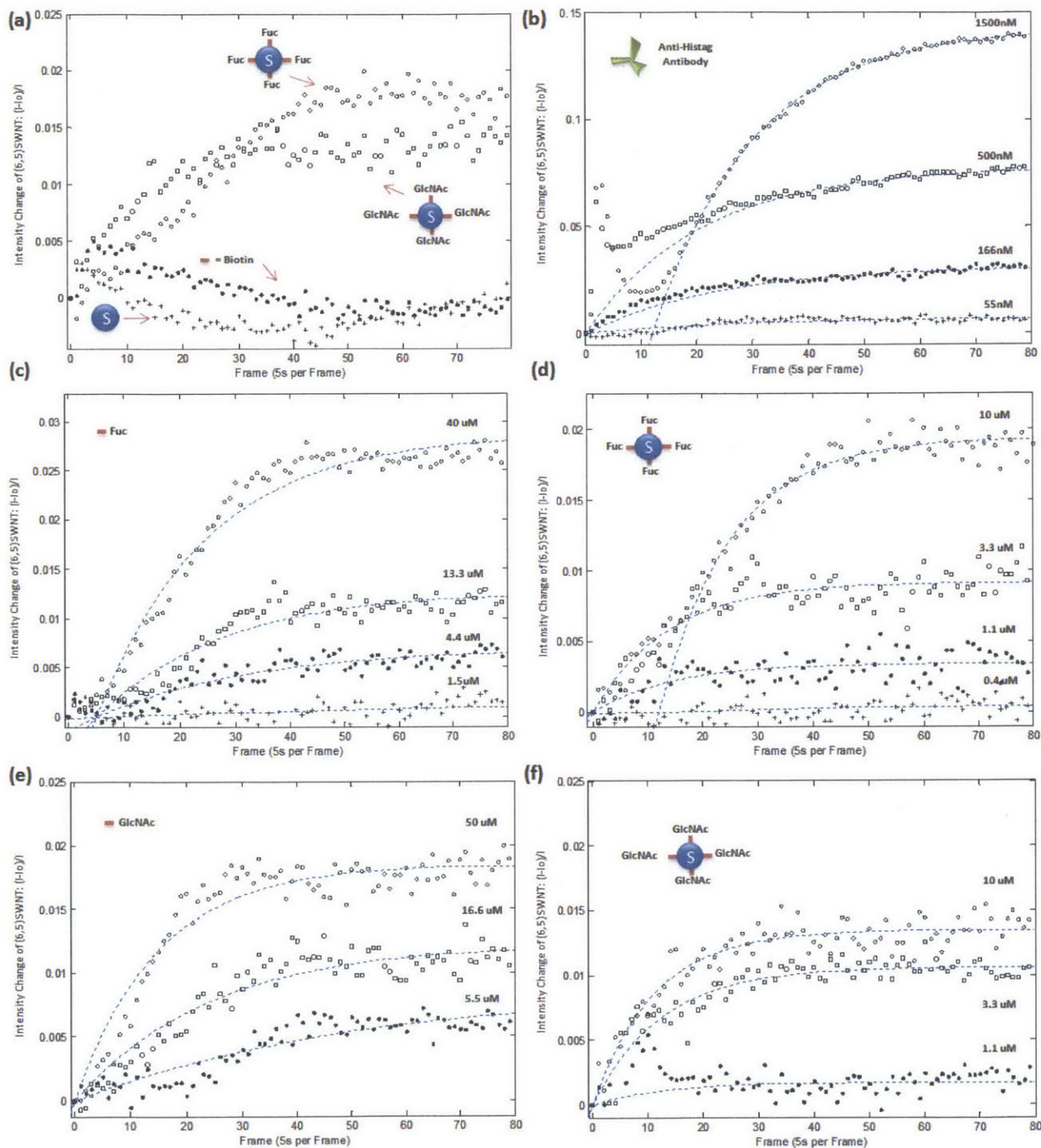
Integrating equation (3.6) yields the equation for the absorption curve:

$$I_t = M(1 - e^{-k_s t}) + I_o \quad (3.8)$$

$$M = \frac{k_f C_g I_m}{k_f C_g + k_d} \quad (3.9)$$

Thus by obtaining absorption curves at three to four different concentrations of glycosylated analytes ( $C_g$ ) and fitting them to equation 3.8, one can plot  $k_s$  versus  $C_g$ . A linear trend supports this binding mechanism and kinetics. As equation 3.7 shows, the slope and y-intercept of this line correspond to  $k_f$  and  $k_r$  respectively.

Concentration dependent absorption curves were obtained for controls, anti His-tag antibody, free biotinylated glycans (Fuc and GlcNAc), and glycans tethered to streptavidin (Fuc and GlcNAc). The controls (Figure 3.6a) revealed a positive response for Fuc-Streptavidin probe to PA-IIL lectin and GlcNAc-Streptavidin probe to GafD lectin. It is also revealed a negligible response to blank streptavidin as well as biotin. Here we must mention that the first 20 frames of the absorption curve often contain artifacts due to the manual additions of analyte (pipette tip to edge of spot). Thus in fitting the absorption curves to equation 3.8 we have set the fit parameters to optimize the fit on the curved portion of the isotherm rather than the artifacts at the beginning of the curve. Also the absorption curves have been linearly corrected for focus drift caused by tension on the z-axis focus due to direct contact with the glass. This small correction (less than 0.05% of the signal) is made by linearly fitting the end of the absorption curve when the system is again at equilibrium.



**Figure 3.6** – Concentration dependent curves loading curves of various analytes to SWNT-Chitosan sensors. A) Control curves showing positive response of GlcNAc-Strept to GafD and Fuc-Strept to PA-IIL and negative responses to biotin and blank streptavidin. Concentration dependent loading curves for (b) anti His-tag antibody to GafD lectin, (c) biotinylated Fuc to PA-IIL, (d) Fuc-Streptavidin to PA-IIL, (e) biotinylated GlcNAc to GafD, and (f) GlcNAc-Streptavidin to GafD. The dotted lines denotes the kinetic model fit and resulting parameters are reported in Table 3.1.

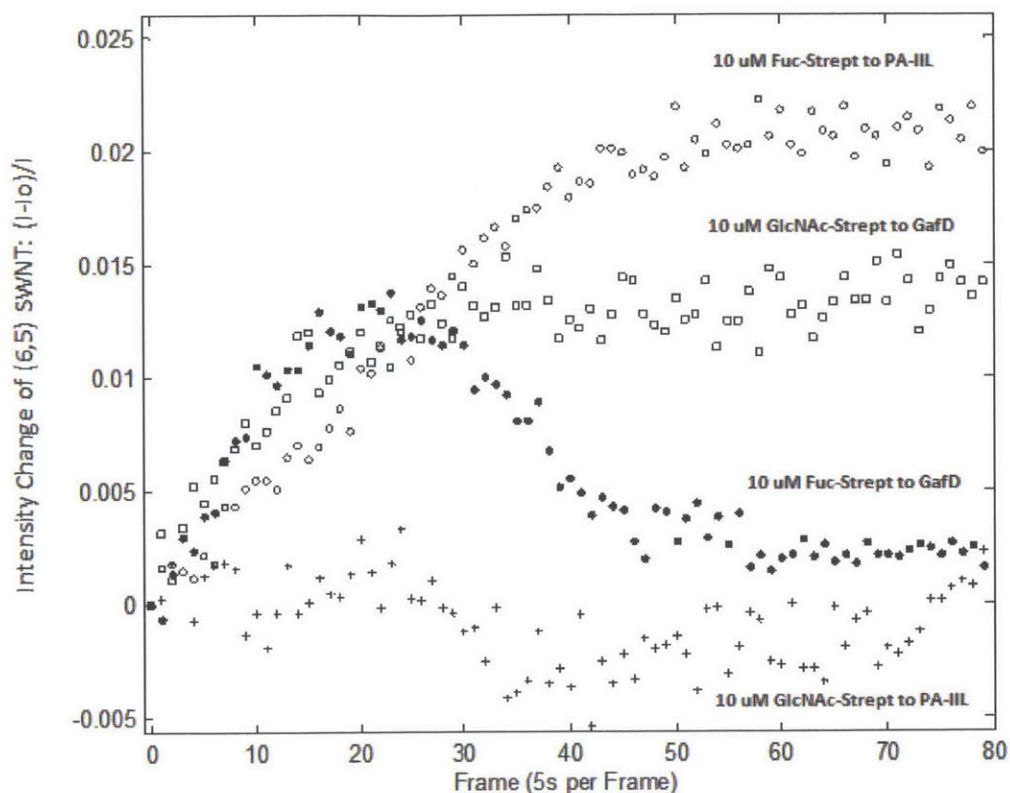
The curves for anti His-tag antibody (Fig 3.6b) were obtained for 1500nM, 500nM, 166nM, and 55nM concentrations interacting with GafD lectin tethered to the SWNT sensor. The resulting  $k_s$  fit was highly linear ( $R^2 = 0.971$ ) and resulting  $k_f$ ,  $k_r$ , and  $K_D$  values are tabulated in Table 3.1. The  $K_D$  of 4  $\mu\text{M}$  found for our murine produced anti His-tag antibody (Sigma H1029) correlates well with the 1  $\mu\text{M}$  found by Biacore SPR studies<sup>42</sup>. The concentration dependant curves for free biotinylated fucose (40, 13.3, 4.4, and 1.5  $\mu\text{M}$ ) to PA-IIL lectin (Fig 3.6c) and tethered fucose-streptavidin probes (10, 3.3, 1.1, and 0.4  $\mu\text{M}$ ) to PA-IIL lectin (Fig 3.6d) also fit the SPR-like kinetic model well and their kinetic parameters are reported in Table 1. The  $K_D$  of 106 and 142  $\mu\text{M}$  found by our sensor for free and tethered fucose to PA-IIL is weaker than the previously reported 3  $\mu\text{M}$  found by isothermal titration microcalorimetry (ITC)<sup>43</sup>. It is often observed that surface tethering of a receptor results in an increase in  $K_D$ , often by orders of magnitude<sup>44</sup>, therefore the decreased affinity of our surface tethered approach compared with ITC measurements of solution phase binding is expected. The discrepancy in  $K_D$  measurement may also arise in the chemical modification of the glycan or the tethered presentation of the lectin on the surface. The binding of free biotinylated GlcNAc (50, 16.6, and 5.5  $\mu\text{M}$ ) and GlcNAc-streptavidin probe (10, 3.3, 1.1  $\mu\text{M}$ ) to GafD lectin (Fig 3.6e-f) also followed the model. The fitted kinetic parameters are reported in Table 1. To our knowledge, this is the first measurement of GlcNAc to GafD kinetics on any platform, although there are many glycan-array studies showing a high relative affinity of GlcNAc to GafD over other glycans.<sup>45-48</sup>

**Table 1** – Kinetic parameters found from concentration dependent curves. 95% confidence ranges are not available for the less-affined glycan-streptavidin probes with the current platform.

Experiment	$R^2$ Val	$K_f (\mu\text{M}\cdot\text{s})^{-1}$	$k_r (\text{s})^{-1}$	$K_D (\mu\text{M})$	95% Confidence $K_D (\mu\text{M})$ Range
ATB to GafD	0.9707	2.00E-06	0.0082	4	2-14
Fuc-Biotin to PA-IIL	0.8649	7.00E-05	0.0074	106	58-192
Fuc-Strept to PA-IIL	0.9367	0.0001	0.0142	142	N/A
GlcNAc-Biotin to GafD	0.941	0.0002	0.0037	19	11.5-25.5
GlcNAc-Strept to GafD	0.99	0.0003	0.015	50	N/A

We also checked the selectivity of the SWNT-Lectin sensors by measuring the cross response of our glycans and lectins (Fig 3.7). The PA-IIL lectin showed negligible binding to GlcNAc-Streptavidin

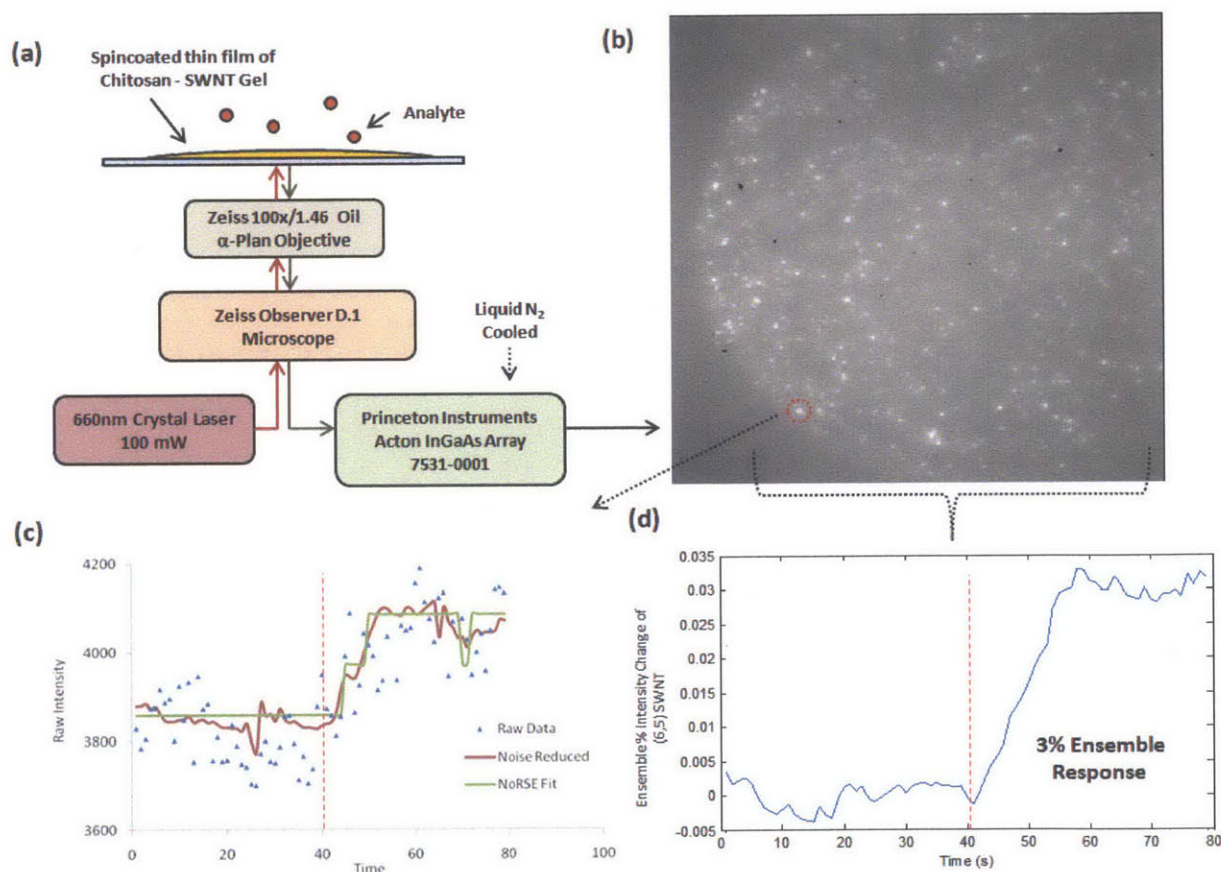
probe, whereas the GafD lectin exhibited a small affinity for Fuc-Streptavidin. However the cross affinity of Fuc-Streptavidin to GafD was much smaller than the known strong-binding combination of fucose to PA-IIL. This demonstrates that the SWNT-lectin sensors could potentially be used to distinguish between sugar groups, especially as the sensor signal is optimized.



**Figure 3.7** – Selectivity of SWNT-Lectin sensors. Greater response of known high affinity pairs (Fuc to PA-IIL and GlcNAc to GafD) than the cross reactions.

The overall change in signal intensity is small (3-5%) for ensemble measurements of glycan-lectin binding for this system (Fig 3.6). We asked if the observed response was homogeneous, meaning that each SWNT responds to this small degree, or inhomogeneous where a subset of SWNT modulate to a much greater extent. The use of single nanotube spectroscopy allows one to address this question (Figure 3.8a). Our laboratory has used this approach for other single molecule sensitive platforms using SWNT-based sensors for  $H_2O_2$ <sup>49,50</sup>,  $NO$ <sup>36,51</sup>, glucose<sup>44</sup>, and nitroaromatics.<sup>52</sup> The resulting thin film of chitosan wrapped SWNT (Figure 3.8b) was imaged at a frequency of 1 frame per second using our InGaAs array setup. Using software developed in house, we can analyze the movies of SWNT fluorescence and extract intensity time traces for each of the individual SWNT sensors (Fig 3.8c).<sup>37</sup> We have also recently developed an efficient algorithm (see thesis chapter 6) for fitting large quantities of these time traces to

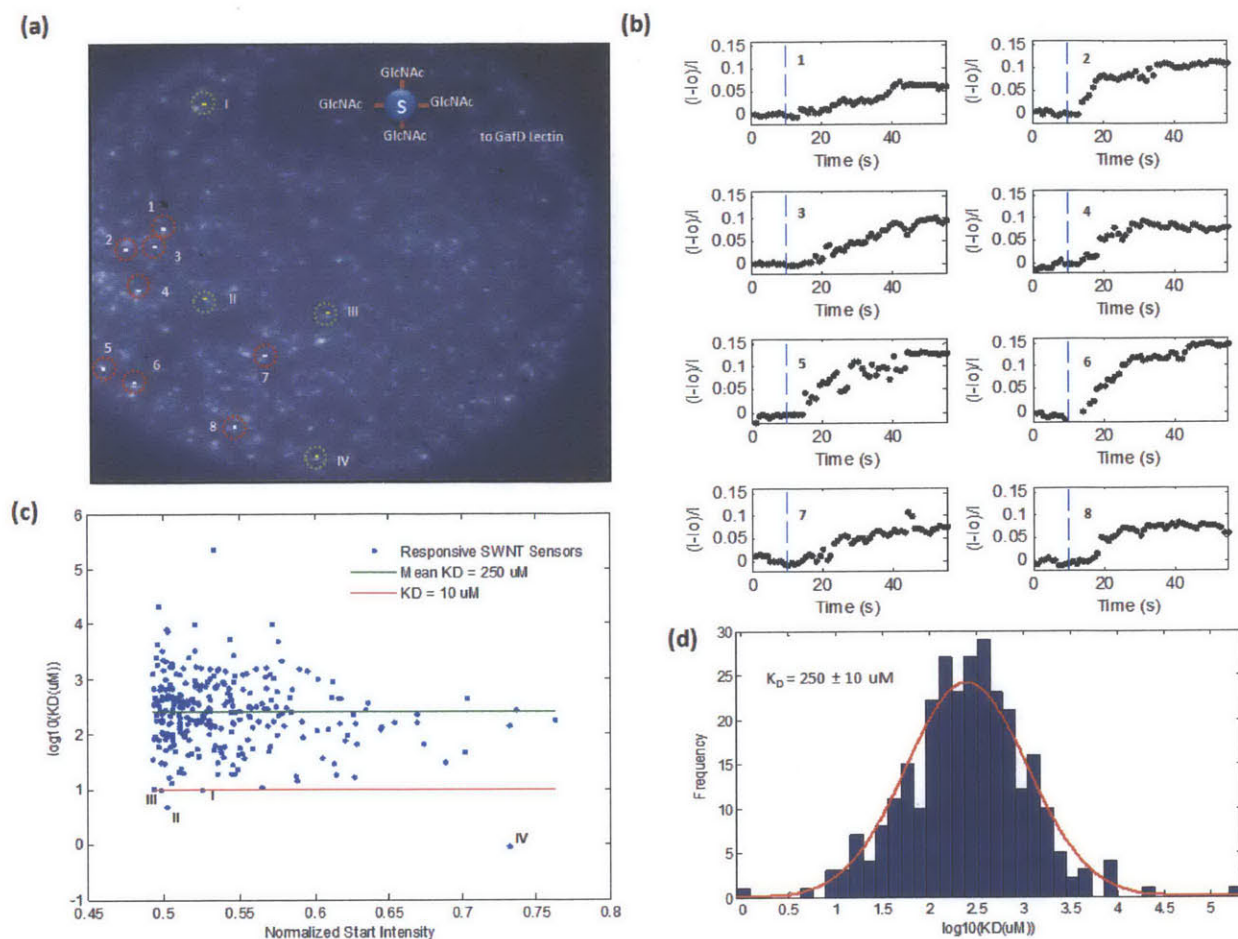
embedded fluorescent levels.<sup>53</sup> Briefly, the algorithm uses an optimized form of an established noise-reduction algorithm for biological experiments<sup>54</sup> to clean the traces (Fig3.8c). It then evaluates all-points histograms of each trace to determine the unique fluorescent states of each trace. The resulting step traces (Fig 3.8c) are then used to determine the forward and reverse kinetics of each SWNT sensor. Before analyzing the kinetics of each individual trace, the prior ensemble experiments were approximated by summing the intensity values of noise-reduced traces from 150 individual SWNT sensors (Fig 3.8d). The resulting signal modulation of GlcNAc-Streptavidin probe (10  $\mu$ M) to GafD is nearly identical to that of our ensemble measurements ( $\sim$ 3.5% response).



**Figure 3.8** – Single SWNT sensor measurements. A) Experimental setup: a thin film of chitosan-SWNT is spin-coated on the glass chip and excited by a 660nm crystal laser, the resulting emission is analyzed by an InGaAs array. B) The array produces a tiff image per time frame where each pixel value denotes the fluorescent intensity, in this manner single SWNT can be visualized. C) By binning 2x2 pixel regions for the brightest 1000 SWNT, individual traces of fluoresce intensity versus time can be created for each SWNT sensor. These traces are then noise-reduced and fitted to determine kinetic parameters (--- denotes addition of GlcNAc-Streptavidin at 10  $\mu$ M to GafD). D) An ensemble average of the individual SWNT sensors can be approximated by adding the signals of 150 SWNT sensors (--- denotes addition of GlcNAc-Streptavidin at 10  $\mu$ M to GafD).

The kinetic analysis of individual SWNT sensors helps determine the locations of “strong-transducers” based on overall signal modulation and kinetic strength. Intensity versus time traces are extracted from the brightest 1000 SWNT sensors (Fig 3.9a), and analyzed for a time period of 40 seconds before and after glycan addition. The location of the eight top SWNT sensors based on signal modulation (each greater than 5%) is easily determined (Fig 3.9b). To determine the  $K_D$  of each SWNT sensor, the traces are noise-reduced and fitted by the NoRSE algorithm<sup>55</sup> and then kinetic parameters are found with the previously reported Birth-Death kinetic model.<sup>36,37</sup> The first 40 seconds of the trace (before glycan addition) are used to determine the background  $k_f$  and  $k_r$  rates for each SWNT sensor (due to intrinsic fluctuations of the tethered group) and then subtracted from the  $k_f$  and  $k_r$  rates found after glycan addition. Of the 1000 traces analyzed, 289 traces had sufficient signal over the background fluctuations to determine the  $K_D$  upon glycan addition (Fig 3.9c). Locations of the four strongest transducers, with  $K_D$  values less than 10 $\mu$ M, were determined. It is interesting that these strong kinetic transducers were not the brightest SWNT traces but rather traces with 50 to 75% the intensity of the brightest recorded. This may reflect the insensitivity of small SWNT bundles to this sensing mechanism. Bundles appear brighter as a composite fluorescent spot, and their construction would necessarily shield the interior SWNT from modulation. Future work will explore this. The population of 289 SWNT transducers was analyzed to find a  $K_D$  of 250  $\pm$  9  $\mu$ M (Fig 3.9d). Further optimization of glycan-lectin kinetic parameters from single molecule analysis could be achieved with a system that has a faster sampling time than the current limit of 1 frame per second.

The SWNT may differ in their ability report the Glycan-Lectin binding events due to accessibility to analyte in the gel, inhomogeneous chemistry modifications (more or less NTA groups per SWNT), SWNT defects, and the influence of multivalent binding. We note that we do not see single molecule steps associated with discrete adsorption steps, as we have seen for other, small molecule quenchers<sup>36,52</sup>. In this case, the interaction of the analyte with the SWNT is indirect, through the spacer chemistry that adjusts the Ni<sup>2+</sup> distance to the SWNT. This mechanism need not be discretized as in the case of adsorption/desorption of a molecular quencher directly on the SWNT surface. In the glycan/lectin system, the quenching distance is continuous. Never the less, single SWNT do respond and contribute to the ensemble response. The fact that the response can be monitored in a single 2x2 pixel spot offers possibilities to dramatically decrease the quantity of required analyte. Even the current responses are an improvement over SPR (which can require 20-250  $\mu$ g of protein analyte at optimal run conditions<sup>33</sup>) since we utilized less than 2  $\mu$ g of glycosylated protein or 100 ng of free glycan as the analyte probe. The amount of lectin (20  $\mu$ g) used for the SWNT sensors can also be dramatically reduced by microprinting smaller volumes of protein directly on an array of optimally responding SWNT sensors.

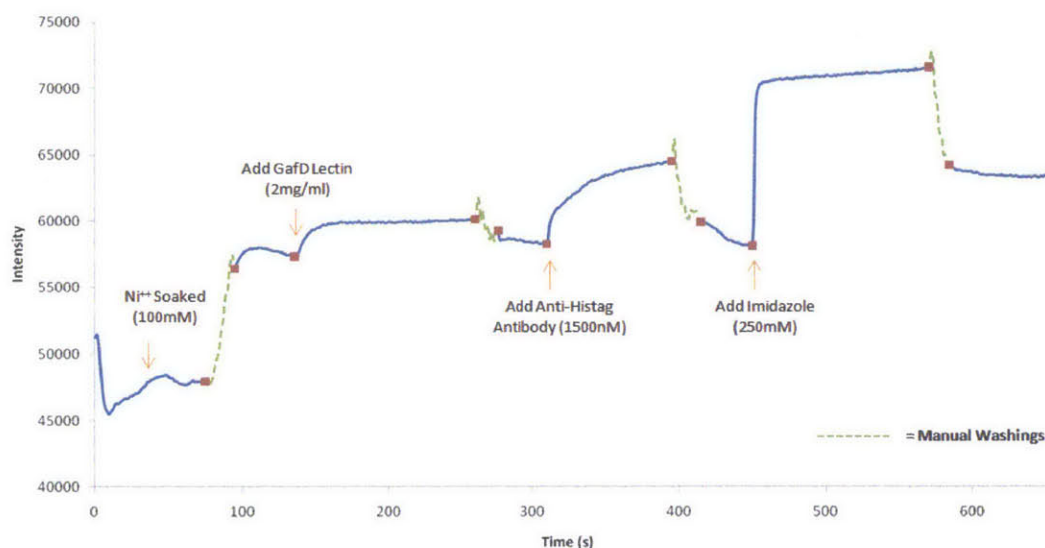


**Figure 3.9** – Single SWNT sensor measurement results. (a) thin film of CHI-SWNT is imaged by the single sensor setup showing a field of sensors on the glass surface. (b) The sensors can be evaluated individually to find which ones give maximum signal modulation. Shown are curves from eight of the most responsive with their positions marked by red circles in (a). Vertical dotted lines denote the time of analyte addition. (c) 289 SWNT sensors have signals strong enough to determine  $K_D$  values, these are plotted versus starting intensity. Four sensors are found to have strong  $K_D$  values ( $< 10\mu\text{M}$ ) and their locations are denoted by green circles on (a). (d) The population of  $K_D$  values yields a single Gaussian distribution when plotted as a histogram of  $\log_{10}(K_D)$  values with a mean  $K_D$  of 250  $\mu\text{M}$ .

Finally, these data sets reveal more about the response mechanism of our Ni-NTA tethered SWNT sensor. As demonstrated previously<sup>16</sup>, the  $\text{Ni}^{2+}$  appears to act as a proximity quencher<sup>39</sup> to the SWNT, however in this work we have now done careful time-series analyses rather than static before and after measurements. In each case of analyte binding, the fluorescent signal increases consistent with the  $\text{Ni}^{2+}$  group moving further away from the SWNT group. According to this model, the higher the affinity an analyte has for the sensor protein, the larger the observed increase. To demonstrate this, we include another time trace of anti His-tag antibody response to GafD (Fig 3.10) which includes an addition step of



adding imidazole (250mM). Imidazole exhibits a higher affinity for  $\text{Ni}^{2+}$  than a His-tag group and is often used in protein purification to elute proteins bound in a Ni-agarose column. The time trace shows that for each of the additions (His-tag lectin, antibody, and imidazole) we see an increase in SWNT intensity. This seems to suggest that the increase in tether length is due to “steric-loading” of the NTA- $\text{Ni}^{2+}$ -sensor protein complex. As more analyte binds to the complex, the required space increases and the tethered group’s fluctuations tend to be further from the nanotube. Multivalency of the analyte may also play a role (as in the case of glycan-streptavidin addition) as multiple NTA- $\text{Ni}^{2+}$ -sensor protein complexes stretch to meet a single, multivalent analyte.



**Figure 3.10** – Time response curve showing increased fluorescence response upon addition of all affinity analytes, including Imidazole. Upon addition, the analytes cause steric loading of the tether causing the  $\text{Ni}^{2+}$  to move away from the SWNT; this causes a partial return of the quenched SWNT fluorescence.

### 3.4. Conclusions

In conclusion, we demonstrate a sensor for measurement of binding kinetics of model glycans. The approach uses recombinant lectins as glycan recognition sites tethered via Histidine tags to  $\text{Ni}^{2+}$  complexes that act as fluorescent quenchers for semi-conducting single walled carbon nanotubes embedded in a chitosan hydrogel spot. As model glycans, both free and streptavidin-tethered biotinylated monosaccharides are studied with two higher-affined glycan-lectin pairs: fucose (Fuc) to PA-IIL and *N*-acetylglucosamine (GlcNAc) to GafD. We find that the dissociation constants ( $K_D$ ) for these pairs as free glycans can be measured as 106 and 19  $\mu\text{M}$  respectively and streptavidin-tethered 142 and 50  $\mu\text{M}$  respectively. The absolute detection limit for the current platform was found to be 2  $\mu\text{g}$  of glycosylated protein or 100 ng of free glycan to 20  $\mu\text{g}$  of lectin. Glycan detection (GlcNAc-streptavidin at 10  $\mu\text{M}$ ) is

demonstrated at the single nanotube level as well by monitoring the fluorescence from individual SWNT sensors tethered to GafD lectin. Over a population of 1000 nanotubes, 289 of the SWNT sensors had signals strong enough to yield kinetic information ( $K_D$  of  $250 \pm 10 \mu\text{M}$ ). This single molecule approach allows us to identify the locations of “strong-transducers” on the basis of kinetic strength (4 sensors with  $K_D < 10 \mu\text{M}$ ) or overall signal modulation (8 sensors with  $> 5\%$  quench response). The brightest SWNT are clearly not the best transducers of glycan binding. SWNT ranging in intensity between 50 and 75% of the maximum show the greatest response. The ability to pinpoint strong-binding, single sensors is promising to build a nanoarray of glycan-lectin transducers as a high throughput method to profile glycans without protein labeling or glycan liberation pretreatment steps. In the next chapter we improve upon this first-generation sensor and apply them to biomanufacturing of antibodies.

### 3.5 References

- 1 Reuel, N. F. *et al.* Transduction of Glycan-Lectin Binding Using Near-Infrared Fluorescent Single-Walled Carbon Nanotubes for Glycan Profiling. *Journal of the American Chemical Society* **133**, 17923-17933, doi:10.1021/ja2074938 (2011).
- 2 Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867, doi:10.1016/j.cell.2006.08.019 (2006).
- 3 Gamblin, D. P., Scanlan, E. M. & Davis, B. G. Glycoprotein Synthesis: An Update. *Chemical Reviews* **109**, 131-163, doi:10.1021/cr078291i (2009).
- 4 Li, H. J. & d'Anjou, M. Pharmacological significance of glycosylation in therapeutic proteins. *Current Opinion in Biotechnology* **20**, 678-684, doi:10.1016/j.copbio.2009.10.009 (2009).
- 5 Elola, M. T., Wolfenstein-Todel, C., Troncoso, M. F., Vasta, G. R. & Rabinovich, G. A. Galectins: matricellular glycan-binding proteins linking cell adhesion, migration, and survival. *Cellular and Molecular Life Sciences* **64**, 1679-1700, doi:10.1007/s00018-007-7044-8 (2007).
- 6 Hirabayashi, J. Concept, strategy and realization of lectin-based glycan profiling. *Journal of Biochemistry* **144**, 139-147, doi:10.1093/jb/mvn043 (2008).
- 7 Hirabayashi, J., Kuno, A. & Tateno, H. Lectin-based structural glycomics: A practical approach to complex glycans. *Electrophoresis* **32**, 1118-1128, doi:10.1002/elps.201000650 (2011).
- 8 Harvey, D. J. Identification of protein-bound carbohydrates by mass spectrometry. *Proteomics* **1**, 311-328, doi:10.1002/1615-9861(200102)1:2<311::aid-prot311>3.3.co;2-a (2001).
- 9 Cortes, D. F., Kabulski, J. L., Lazar, A. C. & Lazar, I. M. Recent advances in the MS analysis of glycoproteins: Capillary and microfluidic workflows. *Electrophoresis* **32**, 14-29, doi:10.1002/elps.201000394 (2011).
- 10 Lazar, I. M., Lazar, A. C., Cortes, D. F. & Kabulski, J. L. Recent advances in the MS analysis of glycoproteins: Theoretical considerations. *Electrophoresis* **32**, 3-13, doi:10.1002/elps.201000393 (2011).
- 11 Pilobello, K. T., Krishnamoorthy, L., Slawek, D. & Mahal, L. K. Development of a lectin microarray for the rapid analysis of protein glycopatterns. *Chembiochem* **6**, 985-989, doi:10.1002/cbic.200400403 (2005).
- 12 Zheng, T., Peelen, D. & Smith, L. M. Lectin arrays for profiling cell surface carbohydrate expression. *Journal of the American Chemical Society* **127**, 9982-9983, doi:10.1021/ja0505550 (2005).
- 13 Rosenfeld, R. *et al.* A lectin array-based methodology for the analysis of protein glycosylation. *J. Biochem. Biophys. Methods* **70**, 415-426, doi:10.1016/j.jbbm.2006.09.008 (2007).

- 14 Tao, S. C. *et al.* Lectin microarrays identify cell-specific and functionally significant cell surface  
glycan markers. *Glycobiology* **18**, 761-769, doi:10.1093/glycob/cwn063 (2008).
- 15 Liang, P. H., Wang, S. K. & Wong, C. H. Quantitative analysis of carbohydrate-protein  
interactions using glycan microarrays: Determination of surface and solution dissociation  
constants. *Journal of the American Chemical Society* **129**, 11177-11184, doi:10.1021/ja072931h  
(2007).
- 16 Ahn, J. H. *et al.* Label-Free, Single Protein Detection on a Near-Infrared Fluorescent Single-  
Walled Carbon Nanotube/Protein Microarray Fabricated by Cell-Free Synthesis. *Nano Lett* **11**,  
2743-2752, doi:10.1021/nl201033d (2011).
- 17 Weis, W. I. & Drickamer, K. Structural basis of lectin-carbohydrate recognition. *Annual Review*  
*of Biochemistry* **65**, 441-473 (1996).
- 18 Wijagkanalan, W., Kawakami, S. & Hashida, M. Glycosylated carriers for cell-selective and  
nuclear delivery of nucleic acids. *Frontiers in bioscience : a journal and virtual library* **17**, 2970-  
2987 (2011).
- 19 Hong, S. Y. *et al.* Filled and glycosylated carbon nanotubes for in vivo radioemitter localization  
and imaging. *Nature Materials* **9**, 485-490, doi:10.1038/nmat2766 (2010).
- 20 Vedala, H. *et al.* Nanoelectronic Detection of Lectin-Carbohydrate Interactions Using Carbon  
Nanotubes. *Nano Letters* **11**, 170-175, doi:10.1021/nl103286k (2011).
- 21 Nagaraj, V. J. *et al.* Nano Monitor: a miniature electronic biosensor for glycan biomarker  
detection. *Nanomedicine* **5**, 369-378, doi:10.2217/nnm.10.11 (2010).
- 22 Xue, Y. *et al.* Noncovalent functionalization of carbon nanotubes with lectin for label-free  
dynamic monitoring of cell-surface glycan expression. *Analytical Biochemistry* **410**, 92-97,  
doi:10.1016/j.ab.2010.11.019 (2011).
- 23 Varki, A. *Essentials of glycobiology*. 2nd edn, (Cold Spring Harbor Laboratory Press, 2009).
- 24 Larsen, K., Thygesen, M. B., Guillaumie, F., Willats, W. G. T. & Jensen, K. J. Solid-phase  
chemical tools for glycobiology. *Carbohydrate Research* **341**, 1209-1234,  
doi:10.1016/j.carres.2006.04.045 (2006).
- 25 Stevens, J., Blixt, O., Paulson, J. C. & Wilson, I. A. Glycan microarray technologies: tools to  
survey host specificity of influenza viruses. *Nature Reviews Microbiology* **4**, 857-864,  
doi:10.1038/nrmicro1530 (2006).
- 26 Hagiwara, K., Colletcassart, D., Kobayashi, K. & Vaerman, J. P. JACALIN - ISOLATION,  
CHARACTERIZATION, AND INFLUENCE OF VARIOUS FACTORS ON ITS  
INTERACTION WITH HUMAN IGA1, AS ASSESSED BY PRECIPITATION AND LATEX  
AGGLUTINATION. *Molecular Immunology* **25**, 69-83, doi:10.1016/0161-5890(88)90092-2  
(1988).
- 27 Bao, Y. & Newburg, D. S. Capillary electrophoresis of acidic oligosaccharides from human milk.  
*Electrophoresis* **29**, 2508-2515, doi:10.1002/elps.200700873 (2008).
- 28 Gupta, G., Gemma, E., Oscarson, S. & Surolia, A. Defining substrate interactions with  
calreticulin: an isothermal titration calorimetric study. *Glycoconjugate Journal* **25**, 797-802,  
doi:10.1007/s10719-008-9151-7 (2008).
- 29 Tomme, P., Creagh, A. L., Kilburn, D. G. & Haynes, C. A. Interaction of polysaccharides with  
the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC .1. Binding specificity and  
calorimetric analysis. *Biochemistry* **35**, 13885-13894, doi:10.1021/bi961185i (1996).
- 30 Tateno, H., Nakamura-Tsuruta, S. & Hirabayashi, J. Frontal affinity chromatography: sugar-  
protein interactions. *Nature Protocols* **2**, 2529-2537, doi:10.1038/nprot.2007.357 (2007).
- 31 Zhang, Y. L., Li, Q. A., Rodriguez, L. G. & Gildersleeve, J. C. An Array-Based Method To  
Identify Multivalent Inhibitors. *Journal of the American Chemical Society* **132**, 9653-9662,  
doi:10.1021/ja100608w (2010).
- 32 Oyelaran, O., Li, Q., Farnsworth, D. & Gildersleeve, J. C. Microarrays with Varying  
Carbohydrate Density Reveal Distinct Subpopulations of Serum Antibodies. *Journal of Proteome*  
*Research* **8**, 3529-3538, doi:10.1021/pr9002245 (2009).

- 33 Karamanska, R. *et al.* Surface plasmon resonance imaging for real-time, label-free analysis of protein interactions with carbohydrate microarrays. *Glycoconjugate Journal* **25**, 69-74, doi:10.1007/s10719-007-9047-y (2008).
- 34 Hsu, K. L., Gildersleeve, J. C. & Mahal, L. K. A simple strategy for the creation of a recombinant lectin microarray. *Molecular Biosystems* **4**, 654-662, doi:10.1039/b800725j (2008).
- 35 Wittig, I. & Schagger, H. Advantages and limitations of clear-native PAGE. *Proteomics* **5**, 4338-4346, doi:10.1002/pmic.200500081 (2005).
- 36 Zhang, J. Q. *et al.* Single Molecule Detection of Nitric Oxide Enabled by d(AT)(15) DNA Adsorbed to Near Infrared Fluorescent Single-Walled Carbon Nanotubes. *Journal of the American Chemical Society* **133**, 567-581, doi:10.1021/ja1084942 (2011).
- 37 Boghossian, A. A. *et al.* The chemical dynamics of nanosensors capable of single molecule detection. *J Chem Phys* **135** (2011).
- 38 Ahmed, H. *Principles and reactions of protein extraction, purification, and characterization.* (CRC Press, 2005).
- 39 Brege, J. J., Gallaway, C. & Barron, A. R. Fluorescence Quenching of Single-Walled Carbon Nanotubes with Transition-Metal Ions. *Journal of Physical Chemistry C* **113**, 4270-4276, doi:10.1021/jp808667b (2009).
- 40 Jin, H. *et al.* Divalent ion and thermally induced DNA conformational polymorphism on single-walled carbon nanotubes. *Macromolecules* **40**, 6731-6739, doi:10.1021/ma070608t (2007).
- 41 Oshannessy, D. J., Brighamburke, M., Soneson, K. K., Hensley, P. & Brooks, I. Determination of rate and equilibrium binding constants for macromolecular interactions using surface-plasmon resonance: use of non-linear least-squares analysis methods *Analytical Biochemistry* **212**, 457-468 (1993).
- 42 Nieba, L. *et al.* BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. *Analytical Biochemistry* **252**, 217-228 (1997).
- 43 Perret, S. *et al.* Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of *Pseudomonas aeruginosa*. *Biochem. J.* **389**, 325-332, doi:10.1042/bj20050079 (2005).
- 44 Yoon, H. *et al.* Periplasmic Binding Proteins as Optical Modulators of Single-Walled Carbon Nanotube Fluorescence: Amplifying a Nanoscale Actuator. *Angewandte Chemie-International Edition* **50**, 1828-1831, doi:10.1002/art.30229 (2011).
- 45 Saarela, S., WesterlundWikstrom, B., Rhen, M. & Korhonen, T. K. The GafD protein of the G (F17) fimbrial complex confers adhesiveness of *Escherichia coli* to laminin. *Infection and Immunity* **64**, 2857-2860 (1996).
- 46 Tanskanen, J. *et al.* The gaf fimbrial gene cluster of *Escherichia coli* expresses a full-size and a truncated soluble adhesin protein. *Journal of Bacteriology* **183**, 512-519 (2001).
- 47 Merckel, M. C. *et al.* The structural basis of receptor-binding by *Escherichia coli* associated with diarrhea and septicemia. *Journal of Molecular Biology* **331**, 897-905, doi:10.1016/s0022-2836(03)00841-6 (2003).
- 48 Carrillo, L. D., Krishnamoorthy, L. & Mahal, L. K. A cellular FRET-based sensor for beta-O-GlcNAc, a dynamic carbohydrate modification involved in signaling. *Journal of the American Chemical Society* **128**, 14768-14769, doi:10.1021/ja065835+ (2006).
- 49 Jin, H., Heller, D. A., Kim, J. H. & Strano, M. S. Stochastic Analysis of Stepwise Fluorescence Quenching Reactions on Single-Walled Carbon Nanotubes: Single Molecule Sensors. *Nano Letters* **8**, 4299-4304, doi:10.1021/nl802010z (2008).
- 50 Jin, H. *et al.* Detection of single-molecule H<sub>2</sub>O<sub>2</sub> signalling from epidermal growth factor receptor using fluorescent single-walled carbon nanotubes. *Nature Nanotechnology* **5**, 302-U381, doi:10.1038/nnano.2010.24 (2010).
- 51 Kim, J. H. *et al.* The rational design of nitric oxide selectivity in single-walled carbon nanotube near-infrared fluorescence sensors for biological detection. *Nature Chemistry* **1**, 473-481, doi:10.1038/nchem.332 (2009).

- 52 Heller, D. A. *et al.* Peptide secondary structure modulates single-walled carbon nanotube  
fluorescence as a chaperone sensor for nitroaromatics. *Proceedings of the National Academy of  
Sciences of the United States of America* **108**, 8544-8549, doi:10.1073/pnas.1005512108 (2011).
- 53 Reuel, N. F. *et al.* NoRSE: Noise Reduction and State Evaluator for High-Frequency Single  
Event Traces. *Bioinformatics*, doi:10.1093/bioinformatics/btr632 (2011).
- 54 Chung, S. H. & Kennedy, R. A. FORWARD-BACKWARD NONLINEAR FILTERING  
TECHNIQUE FOR EXTRACTING SMALL BIOLOGICAL SIGNALS FROM NOISE. *Journal  
of Neuroscience Methods* **40**, 71-86, doi:10.1016/0165-0270(91)90118-j (1991).
- 55 Reuel, N. F. *et al.* NoRSE: Noise Reduction and State Evaluator for High-Frequency Single  
Event Traces. *Bioinformatics* (Revisions submitted 2011).



## 4. SWNT Glycan Sensors Applied to Biomanufacturing Antibodies

Some of the work, text and figures presented in this chapter are reprinted or adapted from Nigel F. Reuel, *et al.* ACS Nano 2013 7 (9), 7472-7482 (reprinted under thesis use allowance from the American Chemical Society © 2013).

### 4.1 Background and Motivation

The concept of vastly multiplexing analyte detection using arrays of independently addressable sensors is ubiquitous in the literature,<sup>1,2</sup> following successful demonstrations of the DNA<sup>3,4</sup> and protein<sup>5,6</sup> microarrays. This approach has been a prevailing motivation for further reduction in array size to nanometer dimensions.<sup>7-9</sup> Such nanosensor arrays, however, have other important capabilities that are not as well recognized, even when functionalized for just a single analyte. For example, many biological analytes, including antibodies, demonstrate a distribution of dissociation constants even in relatively purified form.<sup>10-12</sup> In this work, we demonstrate that an array of sensors can reconstruct this important distribution *via* sampling a large number of independent interactions. Such arrays can also quantify weakly-affined interactions by recording a large number of rare binding events. Lastly, a nanosensor array can also characterize and differentiate biosynthesis around single cells and colonies, enabling the label-free selection of more productive strains, as we are the first to show. These new properties have the potential to greatly enhance process analytics for biomanufacturing applications.

Improved analytical technology<sup>13-15</sup> for the rapidly increasing production of clinical recombinant antibodies<sup>16-20</sup> is an area of great interest. In fact, several recent studies highlight the need for improved (1) clonal selection,<sup>21-23</sup> (2) glycan analysis,<sup>24-29</sup> and (3) determining the affinity distribution or heterogeneity of the expressed product.<sup>10-12</sup> Cell line generation and selecting culture parameters typically take over a year with current assays<sup>30</sup> (the clonal selection part is ~6 weeks but can also take multiple rounds to find the best candidate) and often cell candidates are only picked based on static measurements of productivity.<sup>22,23</sup> Glycosylation patterns can easily change due to processing conditions (media components,<sup>31-33</sup> temperature,<sup>34</sup> pH,<sup>35</sup> pCO<sub>2</sub>,<sup>36</sup> dissolved oxygen,<sup>37</sup> cell density,<sup>38</sup> duration,<sup>31</sup> *etc.*) and they have a dramatic effect on the pharmacokinetics and immunogenicity of the resulting drug.<sup>39,40</sup> Current titer and glycosylation analytical technologies such as ELISA and tandem LC/MS systems respectively can deliver exquisite detail at the expense of much time and reagent. Furthermore their processing steps are prohibitive to any on-line process use. The trend to milliliter-sized bioreactors<sup>41</sup> for upstream process optimization will also require platforms that can accurately assay much lower protein and glycan quantities. Finally, harsh biomanufacturing process conditions (pH, temperature, mixing) and variability

in cell production result in heterogeneous products with a distribution of binding affinities and currently there is no convenient platform on which to measure the dissociation constant ( $K_D$ ) distribution. Emerging nanoengineered sensors<sup>42</sup> fabricated in massive (10,000+ sites) arrays could provide solutions to these three areas of biomanufacturing analytics. In the previous chapter we have described carbon-nanotube based optical sensors for single protein<sup>43</sup> and single glycan<sup>44</sup> detection. However, in this chapter our focus is instead on applying these sensors to antibody production and on previously unacknowledged properties of nanosensor arrays in general, as described above. We demonstrate completely new functions of such arrays for assaying binding heterogeneity, weakly-affined hypermannosylation detection, and determining local cell productivity of biomanufactured products.

## 4.2 Methods

### 4.2.1 SWNT Sensor and Gel Platform Fabrication

SWNT are suspended in chitosan as before.<sup>44</sup> In brief 3 mg of purified HiPCO SWNT (Unidym) are added to 20ml of chitosan suspension (0.25 wt% in water containing 1 vol% acetic acid – Sigma). The resulting mixture was tip sonicated (1/4" tip Cole Parmer, Model CV18) at 10W for 45 minutes in an ice bath and table-top centrifuged three times at 13.2 RPM for 90 min each, while collecting the suspended SWNT supernatant and discarding the aggregate pellet after each cycle. The SWNT was then mixed at a 50:50 volume ratio with the polyacrylamide mixture for casting as the top layer. The amount of monomer (acrylamide) and cross-linker (N,N'-Methylenebisacrylamide – both Sigma) are specified using standard %T %C nomenclature where %T refers to the overall weight % of polymer (monomer and crosslinker) in the solution and %C refers to the wt% of the total polymer that is cross linker. The optimal surface gel was found to be 3%T and 1%C. A substrate gel (6%T 1%C) was also prepared. TEMED (Tetramethylethylenediamine) was added in at 0.7 vol% in both the top and bottom gel solutions to stabilize the radical reaction. A fresh initiator solution of 1 wt% Ammonium persulfate (APS) was made immediately prior to each gel batch. The APS, bottom and top gel solutions, and substrate chips (8 chamber Lab-Tek by Nunc) were degassed in the glove box antechamber to remove absorbed and dissolved oxygen. Within the nitrogen controlled glove box (MBraun LABstar), 1 vol% of the APS solution was added to the substrate gel to initiate the polymerization and it was immediately cast (100ul to each well) and then allowed to cure for one hour. The top gel was then initiated with 1 vol% APS and immediately spotted at 20 ul to each gel surface and allowed to cure for 1 hour.

The functionalization steps of the chitosan wrapped SWNT is also similar to our previous work.<sup>44</sup> In brief, the amine groups of the chitosan are reacted with succinic anhydride (133 mM in PBS 7.4 buffer



- Sigma) overnight and then washed thoroughly with water. The carboxylic acid functional groups are then activated with 100mM EDC and 520mM NHS (Sigma) in MES Buffer pH 4.7 (Pierce) for 2 hours. After washing with water thoroughly the gels are then reacted with 34mM N $\alpha$ ,N $\alpha$ -Bis(carboxymethyl)-L-lysine hydrate (Sigma) in PBS 7.4 buffer overnight. The gels are then washed and incubated with a 100mM nickel sulfate solution for 20 minutes. These chips are then washed thoroughly in water and stored in water.

#### 4.2.2 Poroelastic Relaxation Indentation and Dextran Release Curves

The gel pore size was first evaluated by poroelastic indentation by AFM as explained in the literature.<sup>45</sup> In brief, a short silicon tip with a 45  $\mu\text{m}$  polystyrene sphere (Novascan) is fitted on the AFM (Asylum Research – MFP3D) and the IgorPro software indentation panel is used to drive the tip into the gel at a specified distance and record the force over time (details below). This was done in replicates for multiple sites on each gel type. A custom Matlab algorithm is then used to analyze the force relaxation curves and determine the average pore size (Appendix 4.1).

The top acrylamide gel layer containing the SWNT sensors was tuned to a maximum pore size. This was done by protecting the radical polymerization in a nitrogen-filled glove box from quenching oxygen species. We found that solutions with as low as 3 wt% polymer (97% water) were able to crosslink in this environment (NOTE on standard nomenclature: %T is the total weight percent of polymer (monomer + crosslinker) in solution and %C is the percentage of that polymer that is crosslinker). The effect of crosslinker concentration on pore size has been established qualitatively in literature with TEM imaging,<sup>46</sup> but we probed the pore size of our optimized gels experimentally to ensure that large antibodies could diffuse to the sensor sites.

The pore size was rigorously probed using a recent hydrogel characterization technique called microscale poroelastic indentation.<sup>45</sup> In brief, a microsphere that was orders of magnitude larger than the average pore size ( $R = 22.5 \mu\text{m}$ ) is indented into the gel at a specified distance ( $h$ ) and held while the displaced solvent leaks into surrounding regions. The resulting force *versus* time curve (Fig 4.1b) exhibited an initial spike ( $F_0$ ) that relates to the shear modulus of the gel ( $G$  in Eq 4.1) and then relaxed to an equilibrium state ( $F_\infty$ ) which relates to the diffusion constant of the displaced fluid. The force relaxation curve was obtained for multiple penetration depths ( $h = 8, 6, \text{ and } 4 \mu\text{m}$ ), normalized (Fig 4.1c) and fit by a numerically solved model for an indented sphere geometry<sup>47</sup> to determine the diffusion constant ( $D$  in Eq 4.2) and Poisson ratio ( $\nu_s$  in Eq 4.3). These parameters and the viscosity of water ( $\eta$ ) were then used to solve for the average pore size of the network ( $\xi_p$  in Eq 4.4).<sup>48</sup> These measurements

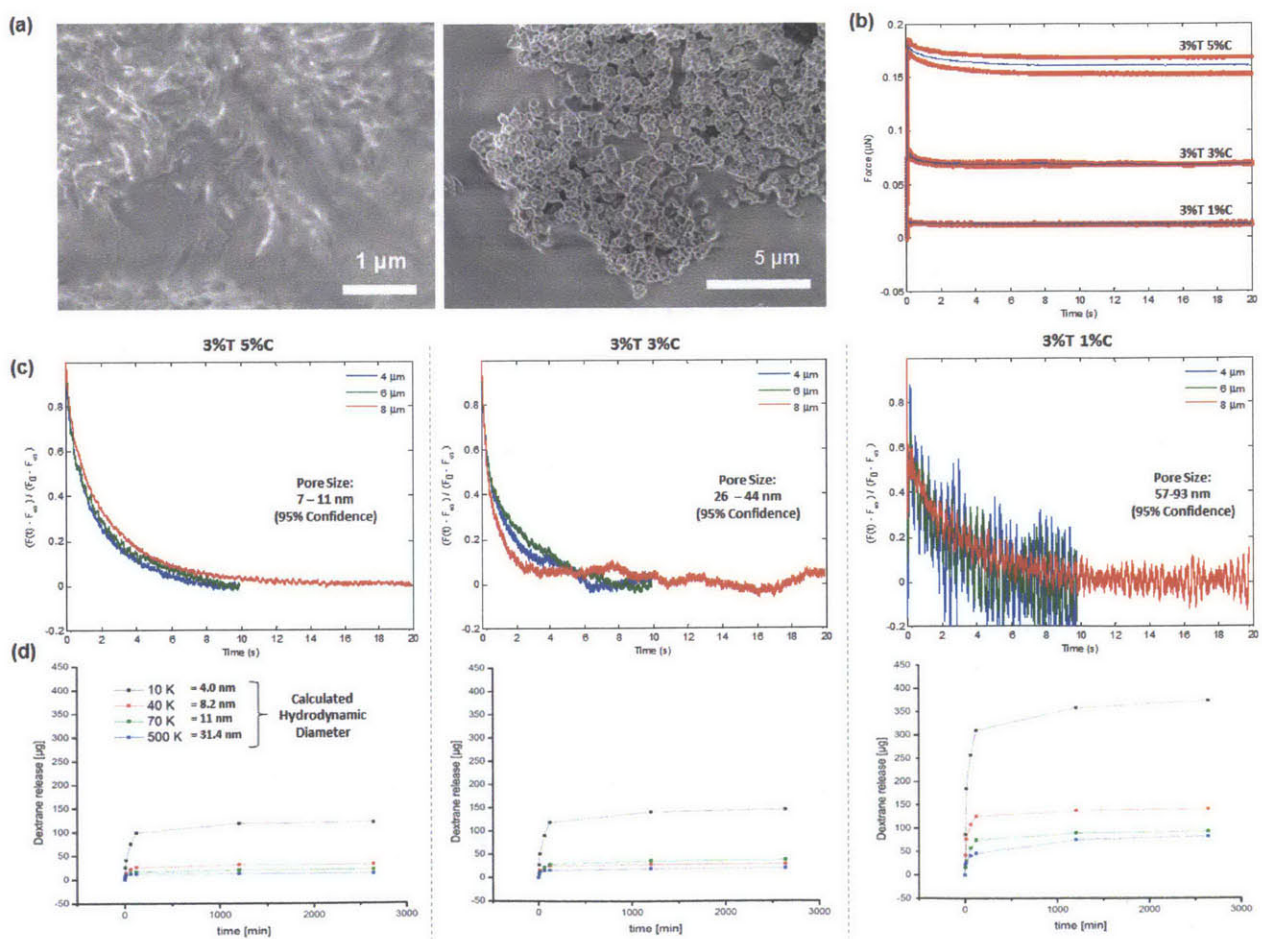
yielded an average pore size of 57-93 nm (95% confidence interval) which is 5-9 times the calculated size (~10-11nm) of a hydrated IgG antibody<sup>49</sup> (code available in Appendix 4.1).

$$F_0 = \frac{16}{3} G(\sqrt{Rh})h \quad \text{Eq. 4.1}$$

$$\frac{F(t) - F_\infty}{F_0 - F_\infty} = 0.491 \exp\left(-0.908 \sqrt{\frac{Dt}{Rh}}\right) + 0.509 \exp\left(-1.679 \frac{Dt}{Rh}\right) \quad \text{Eq. 4.2}$$

$$\frac{F_0}{F_\infty} = 2(1 - v_s) \quad \text{Eq. 4.3}$$

$$\xi_p = 2 \left( \frac{\eta D(1 - 2v_s)}{2 G(1 - v_s)} \right)^{\frac{1}{2}} \quad \text{Eq. 4.4}$$



**Figure 4.1** - Characterization of the polyacrylamide sensor gel network. a) SEM images showing fluorescent SWNT clusters (slow dried sample – left panel) and some structural information of the structures (flash dried – right

panel). b) Absolute force indentation curves at 8 $\mu$ m for each of the gel compositions. c) Normalized force relaxation curves for PRI analysis and resulting pore size calculations. d) Elution profiles of different diameter FITC-coated dextran particles from the gel.

FITC-conjugated dextran particles (Invitrogen) of various sizes (10, 40, 70, and 500 kD) were also absorbed into 150 $\mu$ l cylindrical gel plugs over 48 hours. The impregnated gels were then removed, washed, and inserted into clean water. The release of the FITC particles was observed by sampling the exterior fluid and assaying the FITC content with a plate reader. Using standard curves, the release can then be presented as cumulative mass release over time (Fig 4.1d). These curves showed that our largest dextran probe (31.4nm in diameter) was able to freely diffuse into the 3%T 1%C optimized gel.

#### 4.2.3 Data collection on nIR Inverted Microscope

The SWNT sensor data presented in this paper was collected on a custom inverted microscope (Zeiss D.1 Observer) that was fitted with a 660 nm laser (Crystal Laser, 100 mW). A 20x planar objective (Zeiss) was used and the emission intensities were recorded by a nitrogen-cooled InGaAs array (Princeton Instruments). Win Spec software (Princeton Instruments) was used to collect the SWNT emission and saved as an image stack TIF file. This file was then analyzed using the Matlab code presented in Supplement 3. Analyte samples were added to the sensor gels by hand, applying the 100 $\mu$ l sample to the lower right corner of the well as not to place the plastic pipette tip in the laser beam path.

To prepare a sensor gel for testing, it was first thoroughly washed with PBS to exchange the buffer and then allowed to incubate with the his-tag sensor protein (Protein A (Abcam) or PSA lectin (Vector Labs – conjugated to His-tag peptide (Abbiotec) *via* Traut's reagent and SMCC linker (Pierce)) at 500 $\mu$ g/ml overnight. The gel is again washed thoroughly with PBS and then fitted on the microscope for testing.

#### 4.2.4 HEK Cell Line Generation and CHO Origin

A tricistronic expression cassette pLB2-CMV-GFP-TA99 was created using 2A skip peptides.<sup>50</sup> The light and heavy chain sequences of TA99, a murine IgG2a antibody<sup>51</sup>, were linked by a T2A sequence. The expression cassette was cloned into the lentiviral vector, pLB2,<sup>52</sup> modified with a CMV promoter driving GFP-F2A expression, creating the complete plasmid sequence of pLB2-CMV-GFP-F2A-LC-T2A-HC. All cloning was performed using overlap extension PCR. HEK- GFP-TA99 cells were generated using a modified version of a previously described protocol.<sup>53</sup> Briefly, HEK-293FT cells (Invitrogen) were transfected with the following plasmids: pLB2-CMV-GFP-TA99, pCMV-dR8.91,<sup>54</sup> and

pCMV-VSV-G<sup>55</sup> at a mass ratio of 2:1:1 using PEI. After 24 hours, fresh media was exchanged. 48 and 72 hours later, supernatant containing lentiviral particles was harvested. HEK-293 cells were transduced twice for 24 hours by incubation with freshly harvested supernatant supplemented with protamine sulfate at 5 µg/mL. GFP positive cells were selected to a purity of greater than 95% using flow fluorescence activated cell sorting at the Koch Institute Flow Cytometry Core.

The CHO cell line has been cultured by the Love *et al.* group as received (courtesy of D. Burton, Scripps Research Institute). Details on the generation of this cell line can be found in the following paper.<sup>56</sup>

#### **4.2.5 Cell Passaging and Culture**

For standard culture the medias used were DMEM (with 4.5 g/L glucose, 10% heat inactivated FBS (Invitrogen), 2 mM L-glutamine, 100 U/ml Penicillin, 100 U/ml Streptomycin – rest Sigma) and GMEM (same additives - Sigma) for the HEK and CHO cultures respectively. For cultures used in experiments, a serum free media was used for growth and as a buffer in the sensor gel (Invitrogen Freestyle 293). To passage the cells, they were allowed to grow to confluence, washed with PBS, and then released with Trypsin (0.05% w/0.53 mM EDTA). The cells were then pelleted, resuspended in fresh media and diluted at a 1:5 ratio. The cells were passaged every 2-3 days and discarded after the 20<sup>th</sup> passage.

#### **4.2.6 Hypermannosylation CHO Culture Experiment**

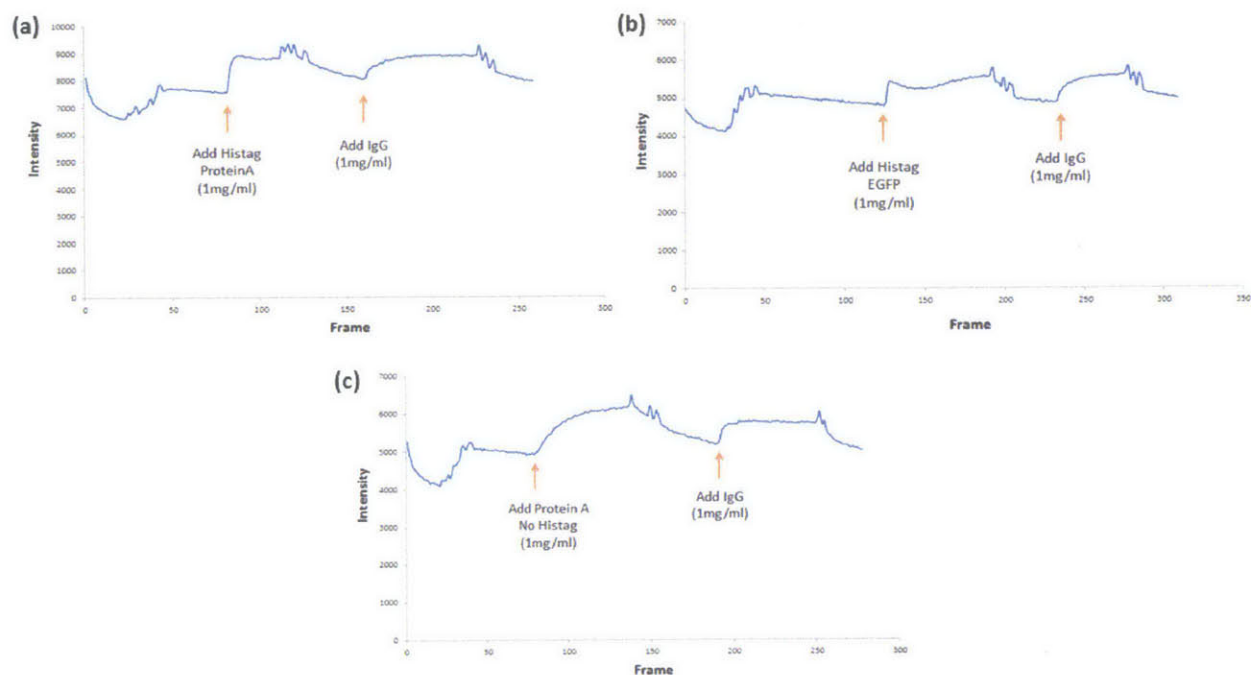
CHO cells were seeded at equal density in small culture flasks (25cm<sup>2</sup> – Sarstedt) and allowed to grow to confluence with regular GMEM media (overnight). The growth media was then exchanged with 3ml serum free media (Freestlye 293 Invitrogen) and the cells were allowed to produce for 24 hours. The media was then saved and a fresh 3ml of serum-free media was added for the next 24 cycle. This was repeated for 10 days. The IgG content was measured by ELISA (ICL Lab, Inc.) and the samples were diluted to 10ng/ml in Freestyle to run on the PSA-incubated SWNT gels.

### **4.3 Results and Discussion**

#### **4.3.1 Sensor Fabrication, Detection Method, and Mechanism**

Recall that the original SWNT sensors for protein and glycan recognition (Chapter 3) were cast in a chitosan hydrogel. Not only were the desired attributes of large pore size and local SWNT

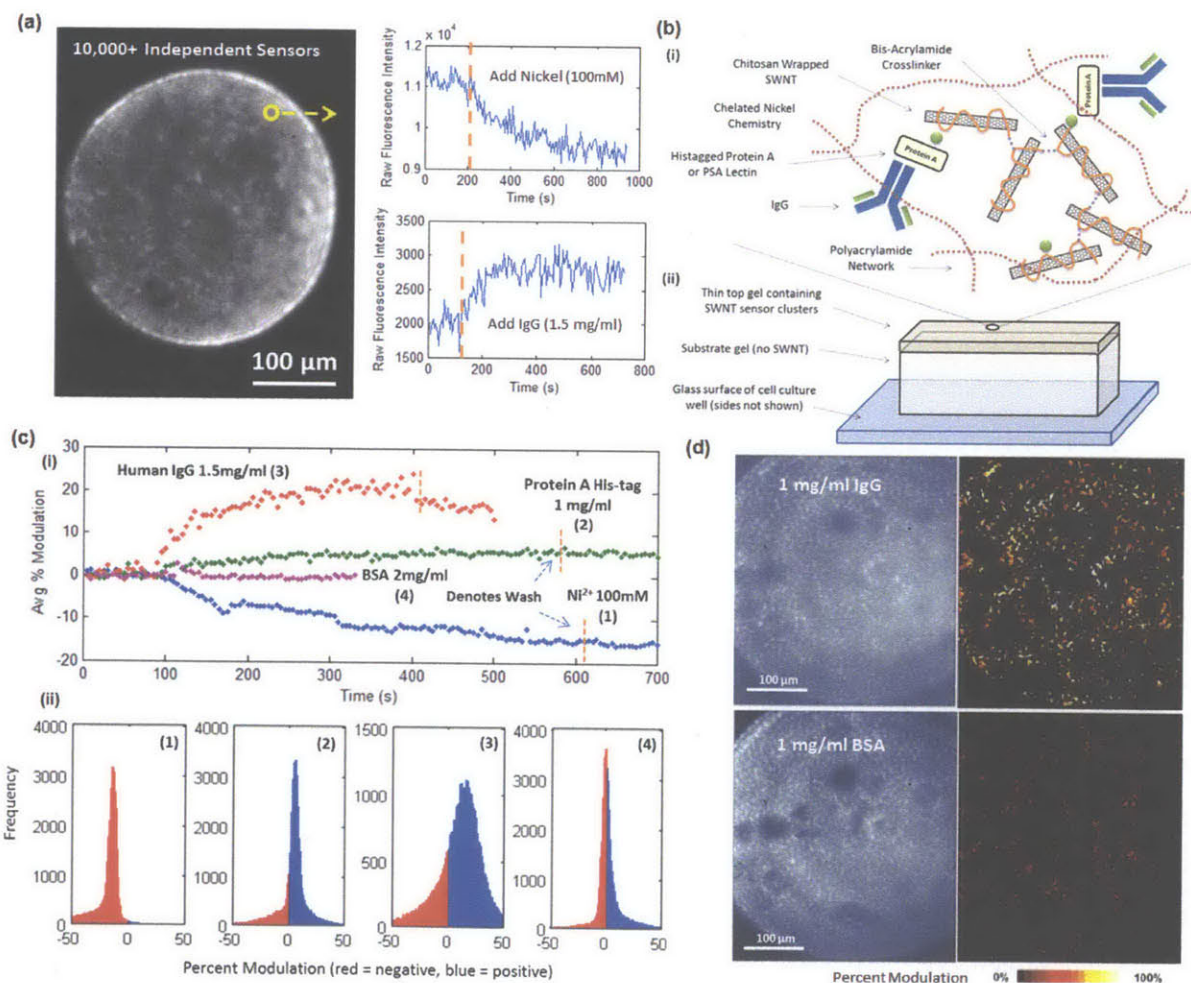
functionalization hard to control in this gel, but the chitosan proved to exacerbate the non-specific binding of the more ‘sticky’ IgG and Protein A species. Adding Protein A-Histag followed by Human IgG (Fig 4.2a) resulted in the expected loading curves but when other proteins were swapped for Protein A, like EFGP-Histag, we still observed a loading curve from the IgG (Fig 4.2b). Furthermore Protein A would also non-specifically bind to the gel, as Protein A without a Histag also resulted in binding (Fig 4.2c).



**Figure 4.2** – Protein A and IgG nonspecific response on original SWNT sensor platform using chitosan hydrogel (Chapter 3)

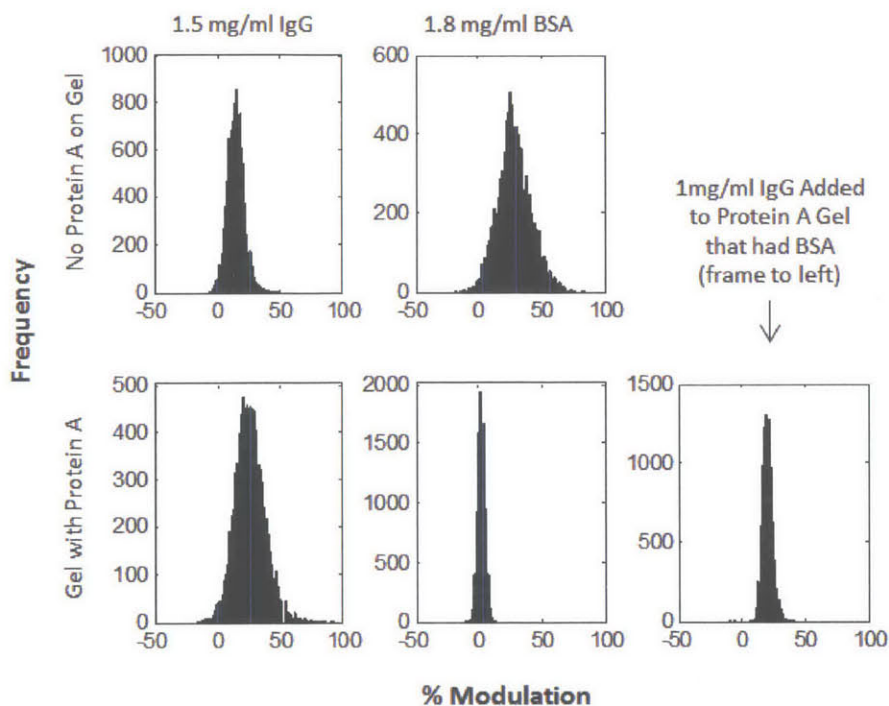
In this work single-walled carbon nanotubes (SWNT) are arrayed (Fig 4.3a) in a multi-layer, highly-porous (60-90nm) polyacrylamide hydrogel to reduce nonspecific binding<sup>46,57</sup> and allow for fast diffusion of the IgG analytes to the SWNT sensors (Fig 4.3b). A basal gel with no SWNT is used on the glass substrate to position the thin sensor layer in a separate focal plane and eliminate the background fluorescence caused by glass impurities. The platform is excited by a 660nm laser (Crystal Laser) on an inverted microscope (Zeiss D.1) and the nIR emission is collected as an image stack on a 256x320 pixel InGaAs array (Princeton Instruments Acton Array). The SWNT are suspended in chitosan and have been chemically modified as before<sup>44</sup> to display chelated nickel groups that act as both the docking site for Histagged capture proteins (Protein A and PSA Lectin here) and as the signal transducer (proximity quencher). In brief, the nanotube acts as an optical switch, brightening as antibodies bind to the capture protein. The ensemble response of the array is created by averaging the intensity values over the entire array for each time point and is analogous to other bioassay techniques like ELISA and Biacore® SPR

measurements. Divalent nickel, protein A, and IgG addition (100mM, 1 mg/ml, 1.5 mg/ml) cause a decrease, increase, and additional increase respectively (Fig 4.3ci). Subsequent washing of the gel surface shows negligible effects on the ensemble signal, which we interpret as absence of unbinding. Adding a BSA control (2mg/ml) does not elicit a sensor response (Fig 4.3ci). Alternatively the responses of all SWNT pixels can be monitored as histograms of percent modulation ( $I_o - I_{Final} / I_o$ ) (Fig 4.3cii). Sensor specificity can also be observed qualitatively with a nIR heatmap filtered to the 10,000 most responsive sites (Fig 4.3d). To render the sensor array specific, we observed that the sensor protein (Protein A or PSA lectin) must first be docked to the chelated nickel; otherwise BSA can elicit a response (Figure 4.4). Although the ensemble response can be efficiently used to construct calibration curves and monitor titer and glycosylation trends, the averaging loses the valuable information about the affinity distribution recorded by each of the individual sensor elements.



**Figure 4.3** – Nanosensor array fabrication and detection methods. a) nIR micrograph of nanosensor array – each SWNT-illuminated pixel records changes in intensity upon analyte binding, such as nickel quenching or signal return from IgG adsorption shown. b) Multi-layer, poly-acrylamide gel platform used to immobilize the SWNT

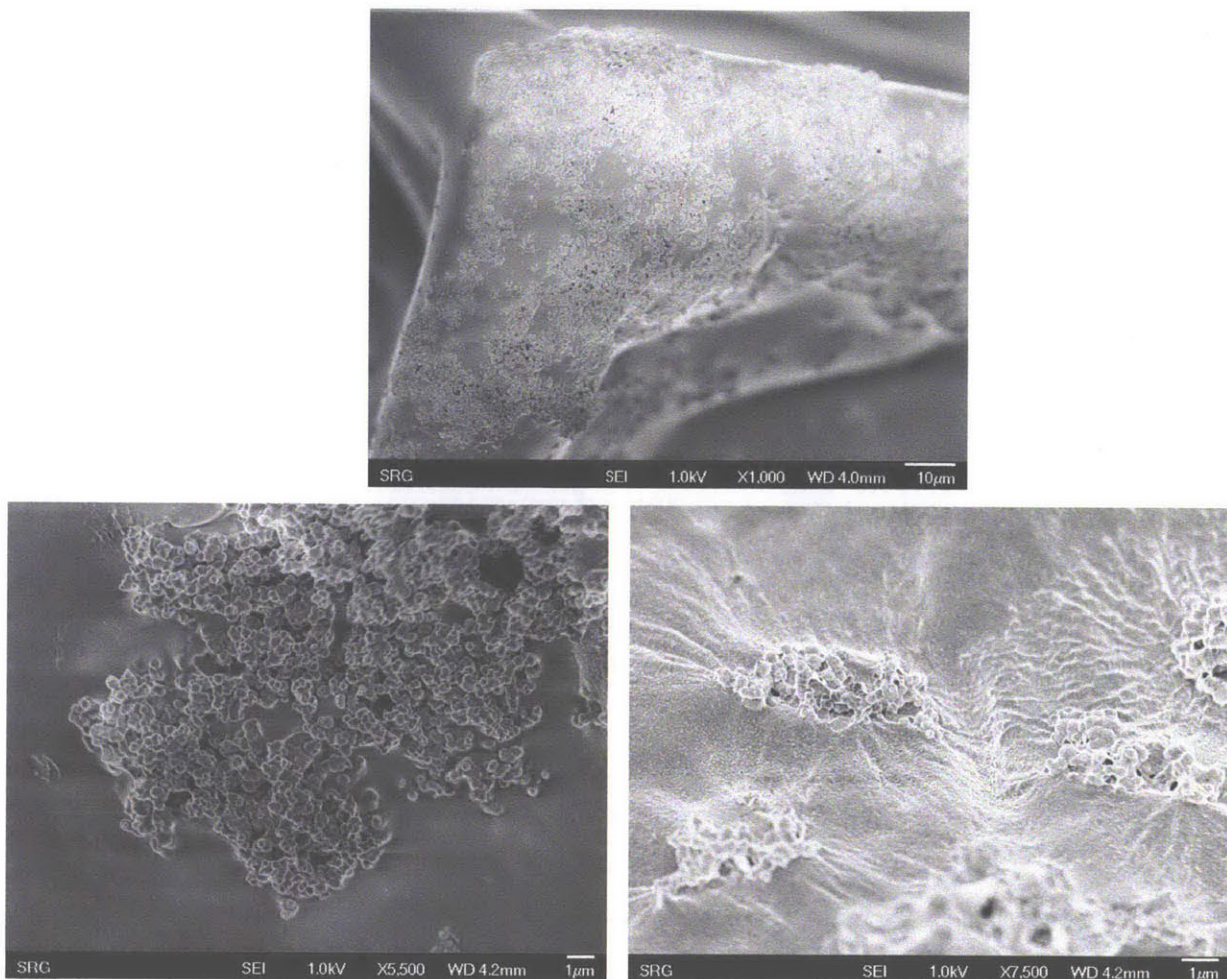
sensors and provide a porous network for IgG diffusion. SWNT sensors are wrapped in chitosan, nickel-chelated, and functionalized with a capture protein (Protein A or PSA – Lectin). c) Ensemble (i) and all-points histogram (ii) response to 100mM nickel (1), 1mg/ml Protein A (2), 1.5mg/ml Human IgG (3) and 2mg/ml BSA (4) additions. d) Selectivity of the sensor array (nIR micrograph of starting intensities shown) as demonstrated by a heat map reporting the percent modulation response of the 10,000 top responding pixels after analyte addition.



**Figure 4.4** – His-tagged sensor protein needed to reduce nonspecific binding - data presented as distribution of all SWNT sensor sites.

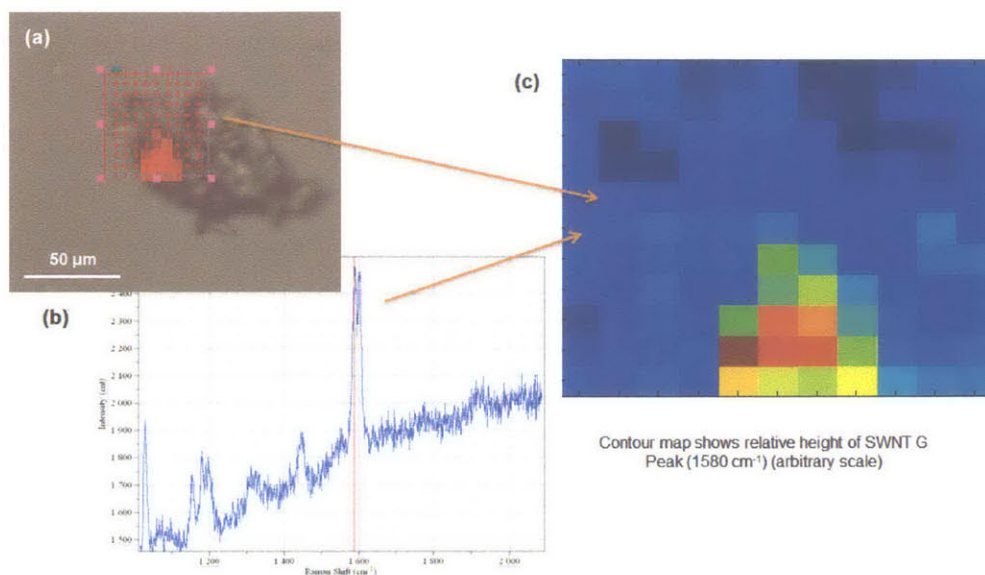
In an attempt to understand the variation between sensor gel batches we did substantial characterization of the polyacrylamide hydrogel network. The thin, top polyacrylamide gel entraps the SWNT sensors in a sparsely cross-linked network. If the chitosan wrapped SWNT are added to the solution containing acrylamide monomers and bis-acrylamide crosslinker immediately prior to radical polymerization initiated by ammonium persulfate, the resulting gel has a homogeneously distributed SWNT pattern. Conversely if the SWNT are able to interact with the monomer and crosslinker solution during the bottom gel curing time, the amine groups on the chitosan-wrapped SWNT undergo linear step growth Michael Addition Polymerization with the bisacrylamide cross-linker (forming poly amido amines) leading to larger fluorescent SWNT clusters. The cluster sizes are dependent on the interaction time of these two constituents before starting the radical polymerization that consumes the remaining crosslinkers and immobilizes the clusters. These clusters can be clearly seen in the nIR micrographs (Fig 4.3a). The clusters visible in the top gel nIR micrographs were characterized with SEM (Fig 4.5) and the

presence of SWNT was confirmed with Raman mapping (Fig 4.6) and TEM (Fig 4.7). The SEM images were created from flash dried samples. We found that this preserved the structural integrity of the underlying gel and some of the clusters in the top gel. We found that gels with SWNT clusters were found more responsive to IgG than the gels with sparse, evenly distributed SWNT. This is likely due to binding events affecting a greater population of SWNT when in cluster formations *versus* interacting with only a few SWNT when patterned as more sparse elements.

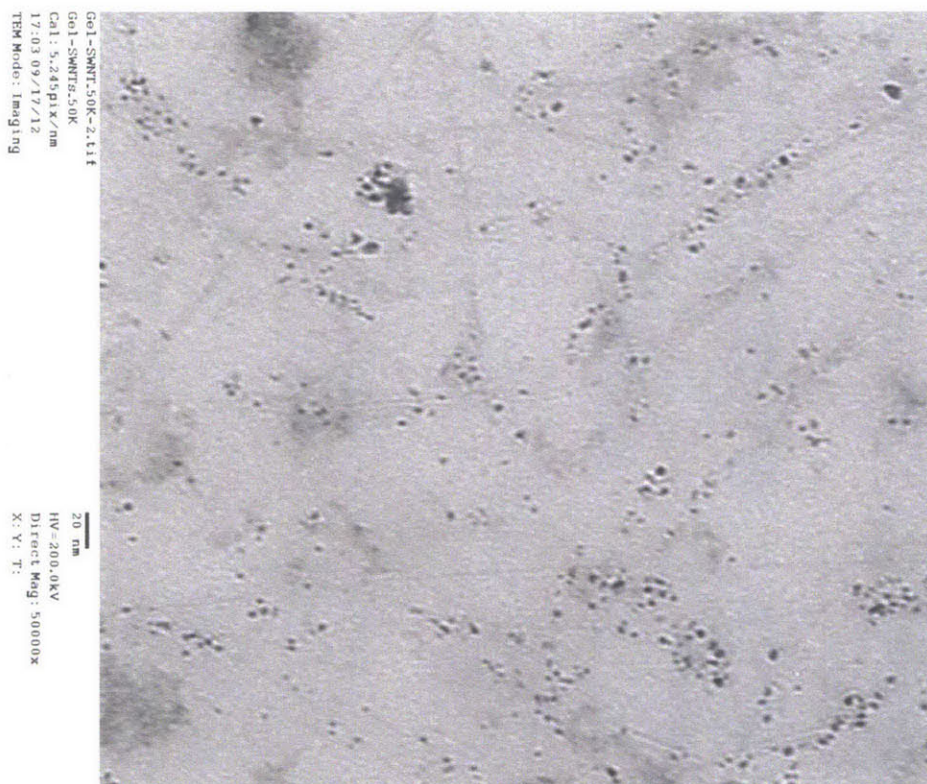


**Figure 4.5** – SEM images of clusters present in the top gel. The smooth background is the substrate gel. One can also see the collapsed surface gel surrounding the clusters.



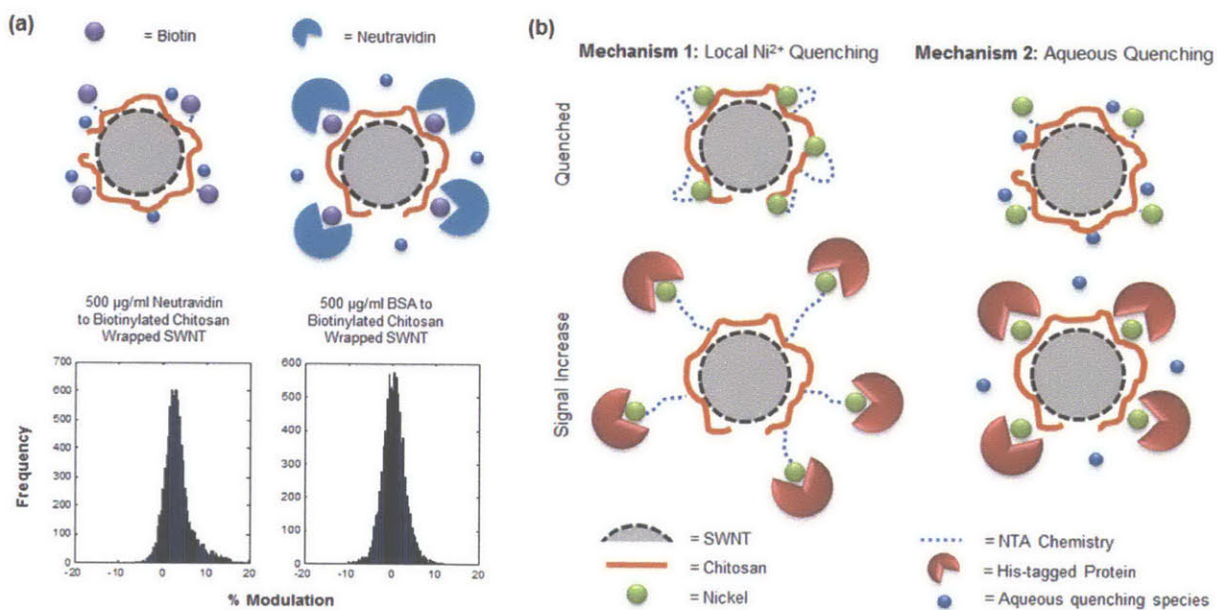


**Figure 4.6** – Raman mapping of clusters reveals SWNT presence. A large cluster is found with the visible objective (a), the SWNT G Peak ( $1580\text{ cm}^{-1}$ ) is probed (b) and a Raman map is created over a section of the cluster revealing a greater concentration of SWNT (c).



**Figure 4.7** – TEM image of top surface gel absorbed and dried to a holey carbon grid. The porosity and structure of this dried gel will differ from the ones cast in the microwell format; however there is evidence of the dried polymer (darker regions) associating with the SWNT (light fibrils).

In our previous work,<sup>43,44</sup> we postulated that the chelated nickel used to bind the His-tag sensor protein is responsible for the modulation of the SWNT fluorescence. From our experiments it is clear that nickel quenches chitosan-wrapped SWNT in solution and immobilized on a surface (Fig 4.3d(i)). Divalent nickel cations have been shown to quench anionic surfactant wrapped SWNT<sup>58</sup> and quantum dots.<sup>59</sup> The proximity of the chelated nickel groups alter the local electronic environment and offer non-radiative decay pathways for the SWNT exciton. To explore if this is the only mechanism at work, we conceived of an experiment to replace the nickel with another small molecule proxy – biotin. The chitosan wrapped SWNT are biotinylated with a commercially available NHS linker (Pierce EZ-Link NHS-Biotin) and then exposed to neutravidin and a BSA control (500µg/ml). The specific response is again a positive modulation and the control is null (Fig 4.8a) although the turn-on response is approximately 20% of what we typically see when nickel is present. This supports two interoperating mechanisms (Fig 4.8b): (1) the sensor proteins are originally quenched by chelated nickel groups and the return of signal occurs when the nickel is displaced upon binding of the sensor protein or IgG, and (2) the chitosan wrapped SWNT is originally quenched by aqueous quenching species (water molecules,<sup>60</sup> dissolved oxygen<sup>34</sup> and protons<sup>61</sup> have all been shown to quench exposed portions of suspended SWNT) and the bound macromolecule displaces these species to cause a return in signal. We ruled out a macroscopic swelling mechanism (as we have observed in previous work with PVA hydrogels<sup>62</sup>) as the recorded movies of SWNT arrays during testing show that the SWNT sensors are immobilized and do not move on the pixel resolution recorded (1.2 µm per pixel).



**Figure 4.8** - The role of nickel in SWNT sensor modulation. a) Testing the current hypothesis of nickel being the crucial transduction component by swapping it out with another small molecule binding site – biotin. Histograms

show sensor responses of the biotinylated chitosan-wrapped SWNT to nuertravidin and BSA. b) Two proposed quenching mechanisms at work – 1) the nickel acts as a quencher providing electronic states for non-radiative decay of the SWNT exciton, 2) the bound macromolecule displaces aqueous quenching species such as nickel and water from the SWNT surface.

### 4.3.2 Modeling a Nanosensor Array to Measure $K_D$ Distributions

Consider an analyte with a Gaussian distribution of  $K_D$  and variance ( $\sigma_t^2$ ) assayed on a  $1 \text{ mm}^2$  total area array of otherwise independently addressable sensors. Obviously, as the total area is subdivided further and further into a population of independent sensor regions, the exact  $K_D$  distribution can be stochastically observed and recovered as the area of a single sensor approaches the projected area of the analyte. Before this limit is reached, however, if each sensor element reports an average of the interactions from multiple molecules, the measured sample variance ( $\sigma_m^2$ ) is much smaller than the actual value (Fig 4.9a insert). For a Gaussian distribution of  $K_D$ , this decay of ( $\sigma_m / \sigma_t$ ) scales as  $1/N^{1/2}$  (see derivation below), where  $N$  is the number of molecules averaged on each sensor site (Fig 4.9a). This relation informs the effect of concentration and sensor area and identifies the regime where nanosensor arrays can effectively report on the variance of  $K_D$  (Fig 4.9b – assumes  $10\text{nm}^2$  area for antibodies). Our demonstration platform in this work uses nanotubes arranged in  $1.4\mu\text{m}^2$  pixels in the concentration ranges of nM to  $\mu\text{M}$ , thus determining distributions of  $K_D$  among samples is feasible when the intrinsic variance of the sensor response is low.

$$N = \text{number of sites in full distribution} \quad \text{Eq 4.5}$$

$$N_{avg} = \text{number of sites averaged together from full distribution} \quad \text{Eq 4.6}$$

$$\sigma_t^2 = \frac{1}{N} [(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_N - \mu)^2] \text{ where } \mu = \frac{1}{N} (x_1 + \dots + x_N) \quad \text{Eq 4.7}$$

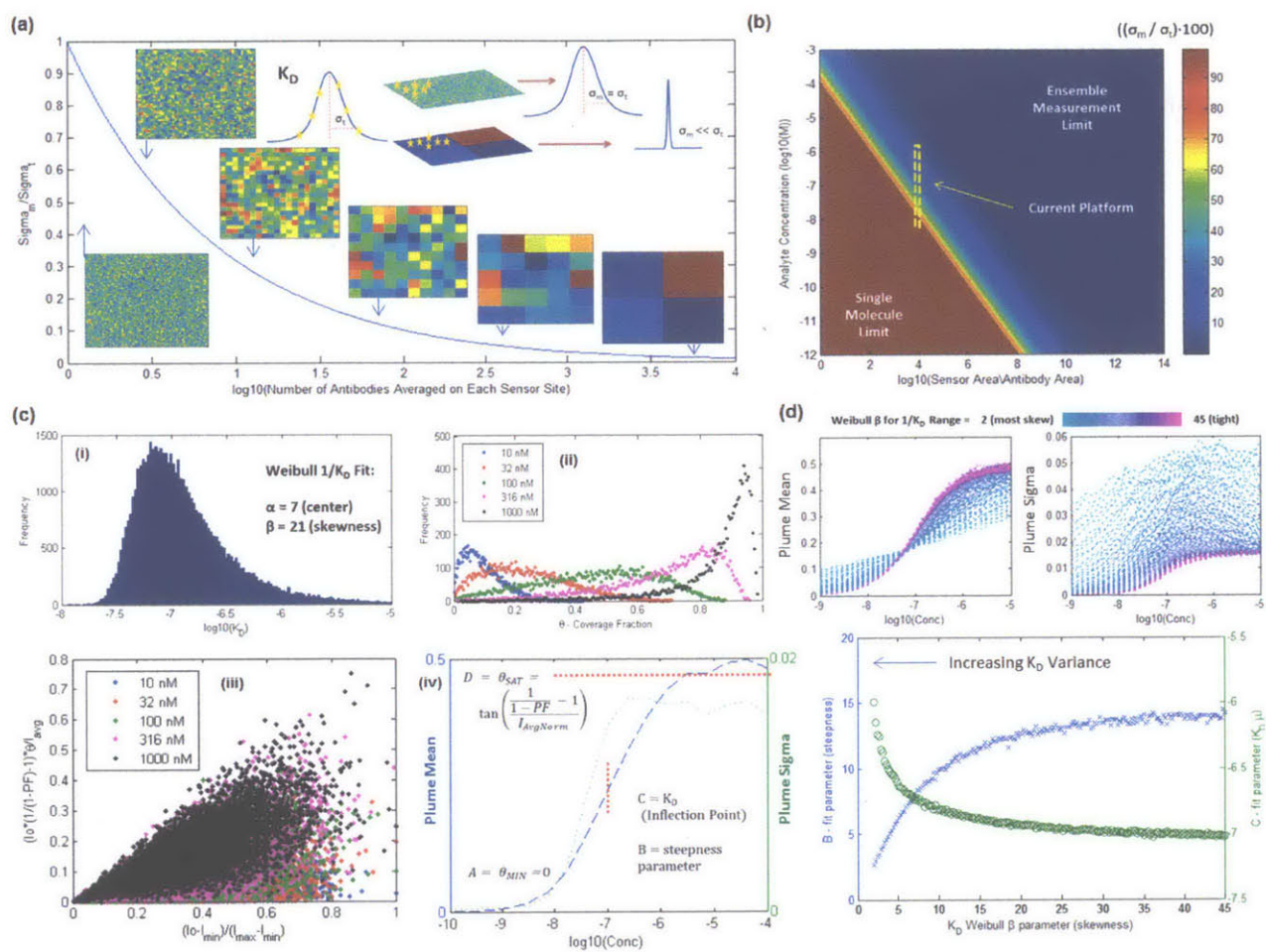
$$\sigma_m^2 = \frac{1}{N/N_{avg}} [(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_N - \mu)^2] \text{ where } \mu = \frac{1}{N} (x_1 + \dots + x_N) \quad \text{Eq 4.8}$$

$$\frac{\sigma_m^2}{\sigma_t^2} = \frac{\frac{1}{N/N_{avg}}}{\frac{1}{N}} = \frac{1}{N_{avg}} \quad \text{Eq 4.9}$$

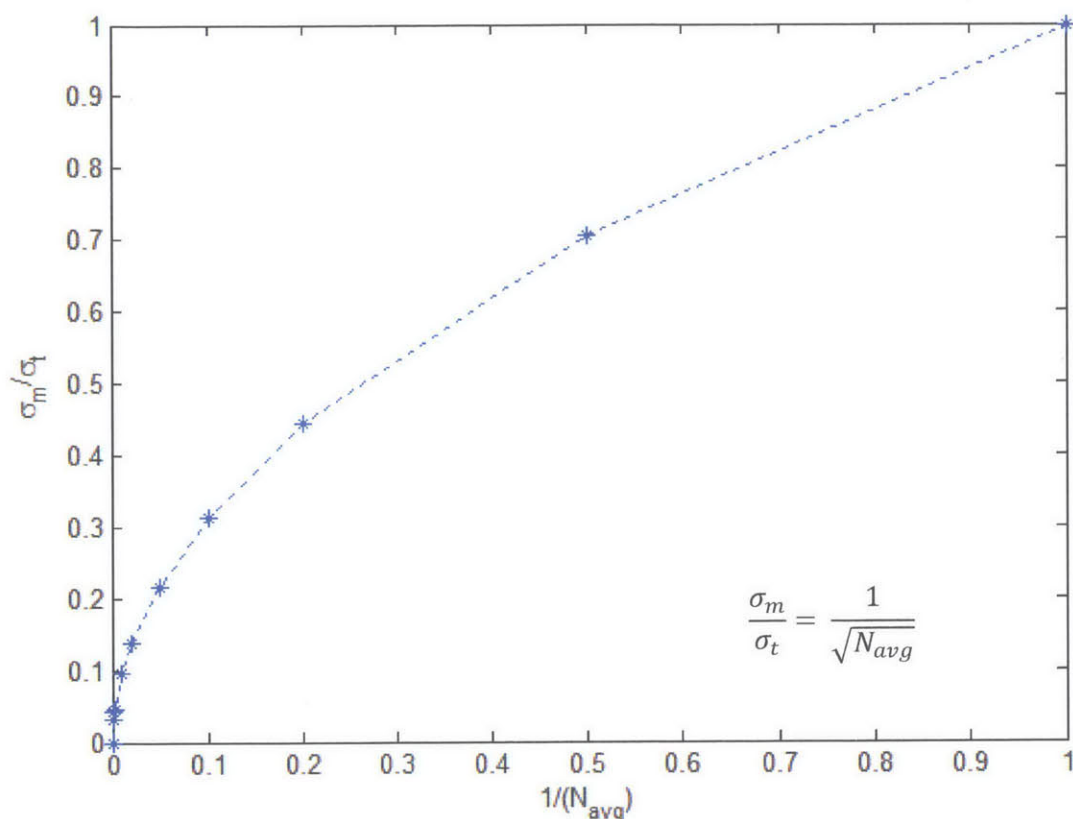
$$\frac{\sigma_m}{\sigma_t} = \frac{1}{\sqrt{N_{avg}}} \quad \text{Eq 4.10}$$

To verify the derivation of equations 5-10 we ran a simulation in which you assume a Gaussian distribution with  $\sigma_t$  and then create a new distribution by selecting a number of points ( $N_{avg}$ ) from the original distribution and then measure this distribution's new standard deviation ( $\sigma_m$ ). By running the

simulation multiple times and averaging the simulation results, the relationship found above clearly emerges (Fig 4.10).



**Figure 4.9** – Modeling a nanosensor array for measuring  $K_D$  distributions. a) Assuming a Gaussian shaped  $K_D$  distribution with known variance ( $\sigma_1^2$ ), the effect of averaging number of molecules ( $N$ ) on a single sensor site on measured variance ( $\sigma_m^2$ ). b) Design regime where  $\sigma_1^2$  can be reconstructed showing two spectrum limits (ensemble and single molecule detection) and where our current nanosensor platform operates. c) Modeling nanotube array response – (i) assuming a Weibull (skewed)  $K_D$  distribution, (ii) Langmuir coverage fraction to determine extent of nanosensor modulation, (iii) resulting simulated response ‘plume’ normalized by average sensor intensity ( $I_{avg}$ ) where each point represents a sensor site, and (iv) calibration curve from fitting the plume angle mean and standard deviation ( $\theta_\mu$  and  $\theta_\sigma$ ) with a bivariate Gaussian distribution. d) Effect of changing  $K_D$  distribution skewness ( $\beta$ ) on resulting  $\theta_\mu$  and  $\theta_\sigma$  calibration curves as well as fit parameters ‘B’ (steepness of calibration curve) and ‘C’ ( $K_D$  mean).



**Figure 4.10** – Simulation of averaging a number of points ( $N_{avg}$ ) from a given distribution with standard deviation  $\sigma_t$  and determining the ratio of the new standard deviation ( $\sigma_m$ ) with the original. The  $-1/2$  power law determined above emerges.

A model is first needed to map our experimental array response back to an initial variance in binding. To simulate the array response, a more realistic distribution of  $K_D$  for an antibody is used. The Weibull distribution has been proposed for antibody affinity due to its ability to effectively present skewness in the probability distribution function using two shape parameters ( $\alpha$  and  $\beta$ ).<sup>12</sup> Antibody affinity is intuitively non-symmetric as changes in the optimal protein will more often result in reduced binding and rarely enhance the product. Thus an antibody with positive skew in  $\log_{10}(K_D)$  is modeled (tail with higher  $K_D$  values or analytes with less affinity) and the probability density function (PDF) is expressed as:

$$\phi_1(x; \alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta} \text{ where } x = \log_{10}\left(\frac{1}{K_D}\right)$$

Eq. 4.11

This distribution is initially centered at  $K_D = 100$  nM (can be translated for other values) such that  $\alpha = 7$  and  $\beta$  can vary from 2 to 45+ with 2 being a very large tail (high variance) and 45+ being an essentially monodisperse distribution (Fig 4.9c(i)). The sensor response is modeled by the Langmuir equation (Eq 4.12) where the coverage fraction ( $\theta_L$ ) is determined by  $1/K_D$  and sample concentration (C) (Fig 4.9c(ii)). In this case  $\theta_L$  represents the extent to which a nanotube sensor is turned on by a small number of local interacting molecules during the time of light acquisition (1s per frame).

$$\theta_L(\phi_1, C) = \frac{\left(\frac{1}{K_D}\right)C}{1 + \left(\frac{1}{K_D}\right)C} \quad \text{where } \left(\frac{1}{K_D}\right) \text{ is sampled from } \phi_1(x; \alpha, \beta)$$

Eq. 4.12

Two additional inputs to the sensor response are the starting intensity of the SWNT ( $I_0$ ) and the percent functionalization (PF). Both will determine the extent to which a given sensor can modulate. Both can be fit with probability density functions (PDF) from experimental data. An all points histogram of pixel intensity (Fig 4.11a) reveals two distinct populations of pixels – those that are not illuminated by the laser spot and the pixels that contain SWNT emission. The starting SWNT intensities for each sensor gel can be fit by a Weibull distribution (Fig 4.11b) and vary in their mean values ( $\alpha$ ) and skewness of bright SWNT ( $\beta$ ) depending on processing conditions, age of SWNT suspension *etc.* However, this does not affect the simulation results as variation in the SWNT  $I_0$  distribution translates to shifting of the center of the plume distribution in the radial distance when performing the Gaussian bivariate fit in polar coordinates. The plume angle is used to report the calibration curve, and not the radial position, thus any positively skewed Weibull distribution can be used. For the simulation (Appendix 4.2) we used the following PDF fit (Fig 4.11b) to the first data set (nickel quench in Fig 4.3a,d):

$$\phi_2(x; \alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta} \quad \text{where } x = I_0, \alpha = 6872, \text{ and } \beta = 2.2$$

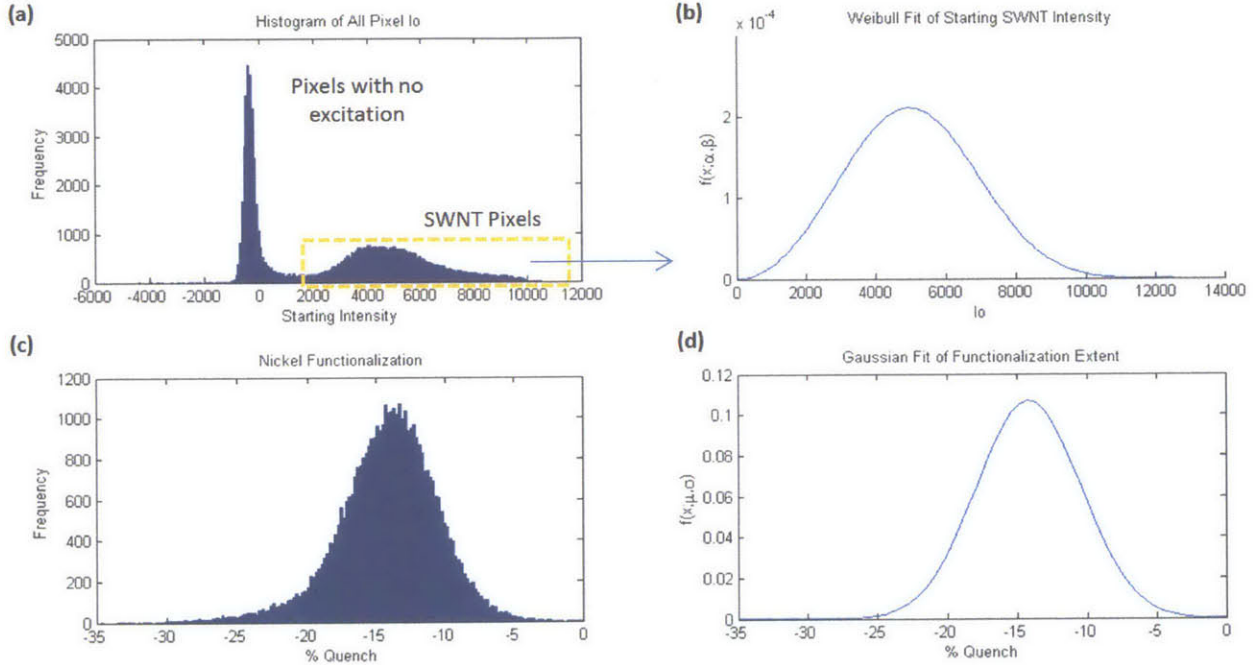
Eq. 4.13

The extent of sensor functionalization was determined by monitoring the extent of quenching caused by nickel addition. This provides a reasonable measure to the accessibility and functionalization of each nanotube. The percent quenching was found to be invariant in SWNT length (banded when

plotted vs.  $I_0$ ) and approximated well by a Gaussian distribution (Fig 4.11c-d) with the following PDF (Fig 4.11d):

$$\phi_3(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ where } x = PF, \mu = -14.2, \text{ and } \sigma = 13.9$$

Eq. 4.14



**Figure 4.11** – Probability distribution functions of sensor starting intensity ( $I_0$ ) and percent functionalization (PF) found from experimental data. a) Histogram of all pixels in nickel quench experiment (Fig 1a of manuscript). Population is fit with a Weibull PDF (b) and can be created for each experiment. c) Histogram of quenching percent after nickel addition - population is found to be conserved amongst separated gels and is fit with a Gaussian PDF (d).

It is important to note that the number of binding sites on the sensor protein is also imbedded in the functionalization variance (if all nanotubes were uniformly wrapped, functionalized, and had the same number of binding sites, this variable would be a constant). To stochastically simulate the response of an ensemble, for example, individual sensor responses,  $R$ , can be obtained as a coverage fraction, intensity, and percent functionalization value ( $\theta$ ,  $I_0$ , and PF,) randomly generated from their respective PDFs ( $\phi_1$ ,  $\phi_2$ ,  $\phi_3$ ) and subsequently evaluated as:

$$R(\theta_L(\phi_1, C), \phi_2(I_o), \phi_3(PF)) = \frac{I_o \left( \frac{1}{1-PF} - 1 \right) \theta_L}{I_{avg}}$$

Eq. 4.15

The response is normalized by the average intensity ( $I_{avg}$ ) as each experimental platform's overall intensity may change due to variance in the experimental setup or quality of SWNT. This normalization allows for clear comparison of distinct arrays assayed at different concentrations. By plotting against the normalized starting intensity ( $(I_o - I_{min}) / (I_{max} - I_{min})$ ), characteristic response 'plumes' are observed (Fig 4.9c(iii)). The modeled 'plumes,' are fit well by a bivariate Gaussian distribution in polar coordinates ( $\theta, R$ ) and the mean 'plume' angle ( $\theta_\mu$ ) is used to create the calibration curve (Fig 4.9c(iv)). The standard deviation in the plume angle ( $\theta_\sigma$ ) also has an interesting dependence on starting  $K_D$  skew ( $\beta$ ) and is reported (Fig 4.9c(iv)). The ( $\theta_\mu$ ) calibration curve is fit by a four parameter logistics curve (used for many other bioassays that exhibit a signal saturation – i.e. ELISA) however in this case two parameters are known from experimental conditions ( $A = \theta_{min} = 0$  and  $D = \theta_{SAT} = \text{Eq. 4.17}$ ).

$$\text{Plume } \theta_\mu(x) = \left( \frac{A - D}{1 + \left( \left( \frac{x}{C} \right)^B \right)} \right) + D \quad \text{where } x = \log_{10}(C)$$

Eq. 4.16

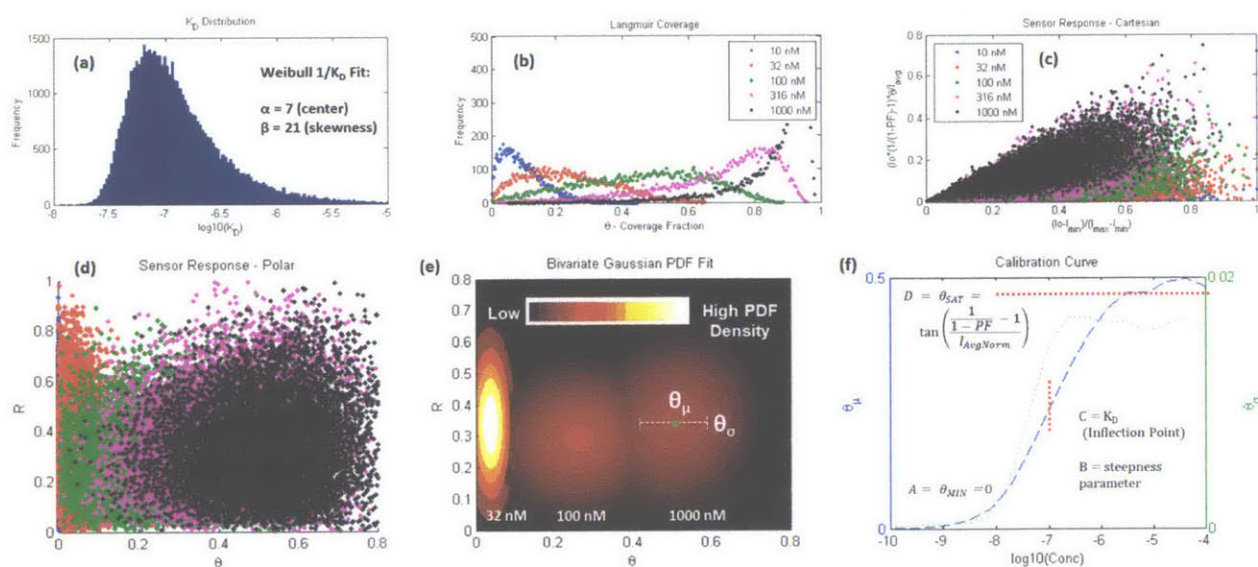
$$D = \theta_{SAT} = \tan \left( \frac{1}{I_{AvgNorm}} \left( \frac{1}{1-PF} - 1 \right) \right)$$

Eq. 4.17

The parameter B is a measure of 'steepness' and C is equal to the inflection point at  $K_D$ . By analyzing the model at many different extents of  $K_D$  skewness ( $2 < \beta < 45$ ), the effect of  $K_D$  variance on the fit parameters were found (Fig 4.9d). As the distribution of antibody affinities becomes tighter ( $\beta$  increases), the nanosensor array yields a steeper calibration curve (Fig 4.13c fits this relation). An analytical relation between  $K_D$  variance ( $\beta$  – skewness parameter in  $\phi_1$ ) and the fit nanosensor response (B-fit parameter) is mathematically beyond the scope of this work (multiplying three PDFs may be



intractable) and thus we have used a numerical simulation to determine this relation. In a single iteration of the program, 10000 sensor sites are generated with random  $K_D$ ,  $I_0$ , and PF values generated from  $\phi_1$ ,  $\phi_2$ , and  $\phi_3$  respectively. The  $K_D$  values, centered at 100nM (Fig 4.12a), determine the coverage fraction coefficients ( $\theta_L$ ) (Fig 4.12b) as specified in the paper (Eq. 2). The response of each sensor is determined for 29 concentration points between 1nM and 10 $\mu$ M (Fig 4.12c). The responses are then fit in polar coordinates (Fig 4.12d) with a bivariate Gaussian distribution (Fig 4.12e). The plume angle mean and standard deviation ( $\theta_\mu$  and  $\theta_\sigma$ ) are then recorded for each concentration. The program completes 6 iterations to create a smooth calibration curve (Fig 4.12f). The initial  $K_D$  skewness parameter ( $\beta$ ) is then changed and the program is run again. This was done for 200 different  $\beta$  values spanning 2 to 45 and the results are shown in Figure 2d of the manuscript. The full code for this simulation can be found in Appendix 4.2. A movie showing how the frames in Fig 12 change for each  $\beta$  value can be found [here](#).

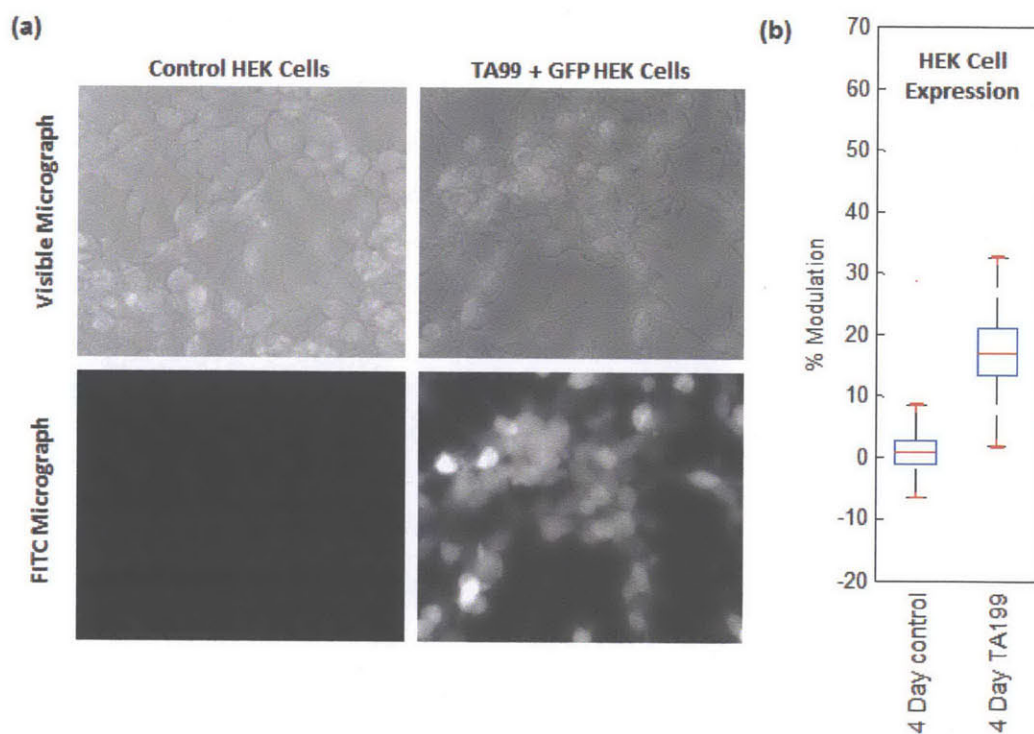


**Figure 4.12** – Simulation frames. a) Starting  $K_D$  distribution from 10,000 simulated points. b) Distributions of Langmuir coverage fraction coefficients ( $\theta_L$ ). Resulting response of all simulated points in cartesian (c) and polar coordinates (d). e) Bivariate Gaussian fit of plume angle mean and standard deviation ( $\theta_\mu$  and  $\theta_\sigma$ ). Simulation sensor points at many different concentrations results in calibration curve (f).

### 4.3.3 $K_D$ Distributions and Calibration Curves from Experimental Data

The array of nanotube sensors conjugated to His-tagged Protein A (Abcam) was used to assay three different samples of IgG with expected differences in affinity distributions: (1) commercial, lyophilized, polyclonal Human IgG reconstituted in PBS, (2) murine IgG (TA99) from an engineered human embryonic kidney (HEK) cell line, and (3) human IgG (b12) cultured from Chinese hamster ovary (CHO) cells. The custom engineered human embryonic kidney (HEK) cell line was transduced to

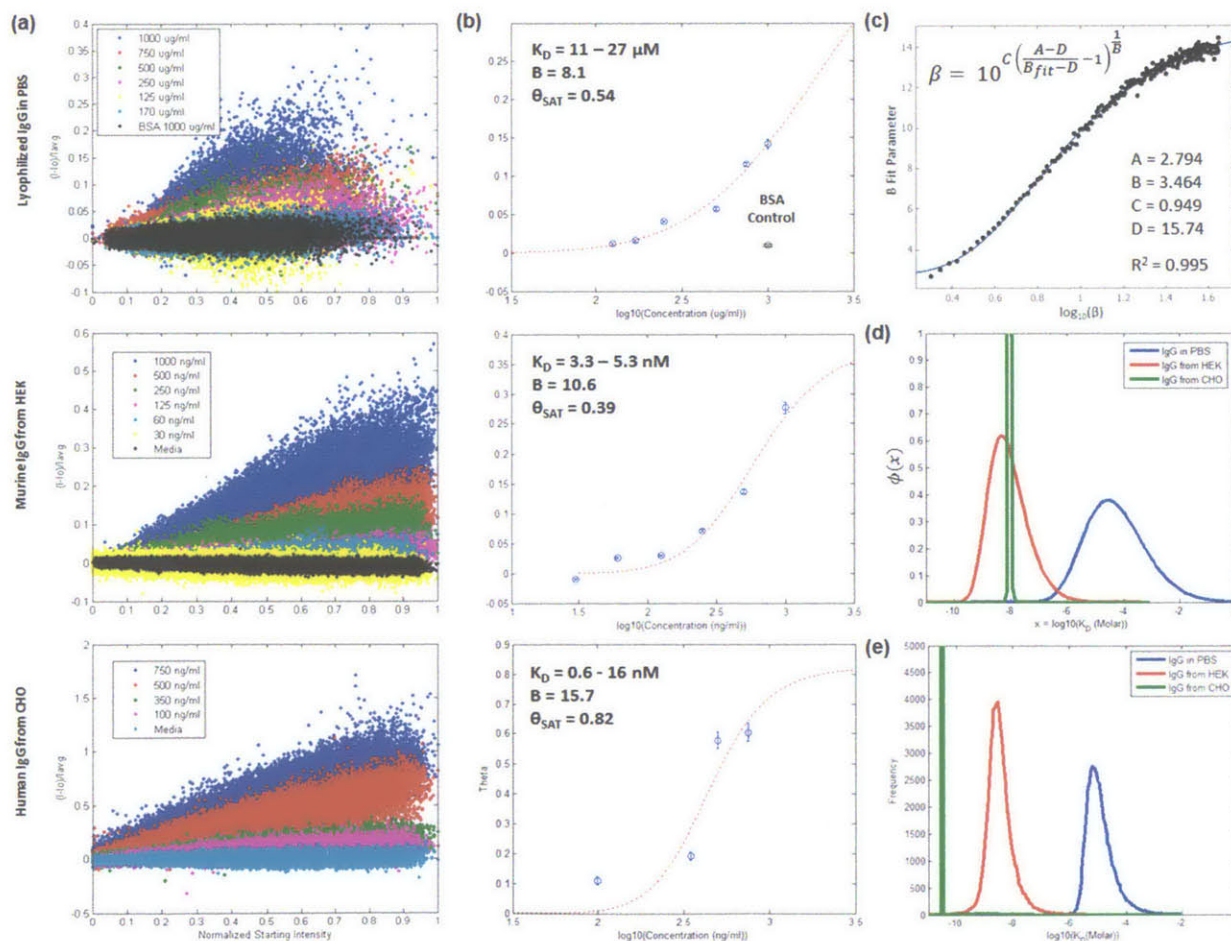
produce both GFP and a murine IgG (TA99). This provides a convenient control as the non-transfected cells can be visually inspected with a FITC filter to confirm the lack of a GFP signal (Fig 4.12a) and therefore assured lack of IgG production. The control HEK cells and TA99 producing HEK cells were seeded at equal density and grown for four days in serum free media (Invitrogen Freestyle 293). The cell extract supernatants of both were then screened showing a statistically significant increase in signal for the antibody producing cells (Fig 4.12b).



**Figure 4.12** - Cell culture extract supernatant studies. a) Engineered HEK cell line that is transduced to produce both murine IgG (TA99) and GFP. b) Response to control and and IgG cell line extract presented as box plot of percentage modulation values.

The data ‘plumes’ predicted by the model were observed for each of the systems (Fig 4.14a) and the resulting calibration curves (Figure 4.14b) yielded  $K_D$  mean values (95% confidence intervals for each: 11 – 27  $\mu\text{M}$ , 3.3 – 5.3 nM, and 0.6 - 16 nM) comparable to those found in literature for IgG-Protein A interactions (2-50 nM from SPR,<sup>63</sup> 34 nM from acoustic device<sup>64</sup>). The calibration curve also provided ‘B’ fit parameters (Eq. 4.16) that are related to the starting distribution skew parameter ( $\beta$ ) as solved from the model simulation (Fig 4.14c). With  $K_D$  mean ( $\alpha$ ) and skewness ( $\beta$ ), the affinity PDF of each system (Eq. 4.11) can be determined (Fig 4.14d). As expected, the freshly expressed Human IgG has the greatest affinity for Protein A with the least amount of variance. The murine antibody has a comparable  $K_D$  average, but much more predicted variance. This difference could be due to less efficient binding of murine IgG to Protein A as observed in literature.<sup>65</sup> Finally the lyophilized, polyclonal human IgG shows

a 1000x reduction in  $K_D$  and a much broader distribution of affinities likely due to denaturation damage<sup>66</sup> or freeze-thaw cycles<sup>67</sup> (vs. the freshly expressed CHO product). One limitation of the modeling approach is the a priori assumption of PDF form (in this case Weibull). The array of nanosensors can be used to report distributions directly, without any assumption of PDF form by calculating the Langmuir coverage from the response of each individual sensor (Eq. 4.12 and 4.15) and creating a histogram of  $K_D$  values (Fig 4.14e). To do this now, however, one assumes a constant percent functionalization value (PF = 0.14, mean of  $\phi_3$ ) and chooses a concentration value away from saturation. These assumptions are an approximation as a distribution of functionalization ( $\phi_3$ ) exists and each concentration does not align perfectly on the calibration curve, so each ‘plume’ will yield a slightly different  $K_D$  histogram with differing  $K_D$  mean values (as can be seen with the CHO data, green trace, in Fig 4.14e). This direct technique for measurement can be improved by better fabrication methods, reducing the intrinsic variance of length and functionalization. We do note however that the direct reporting approach also presents  $K_D$  distributions with similar shape and positive skew as our model assumed.

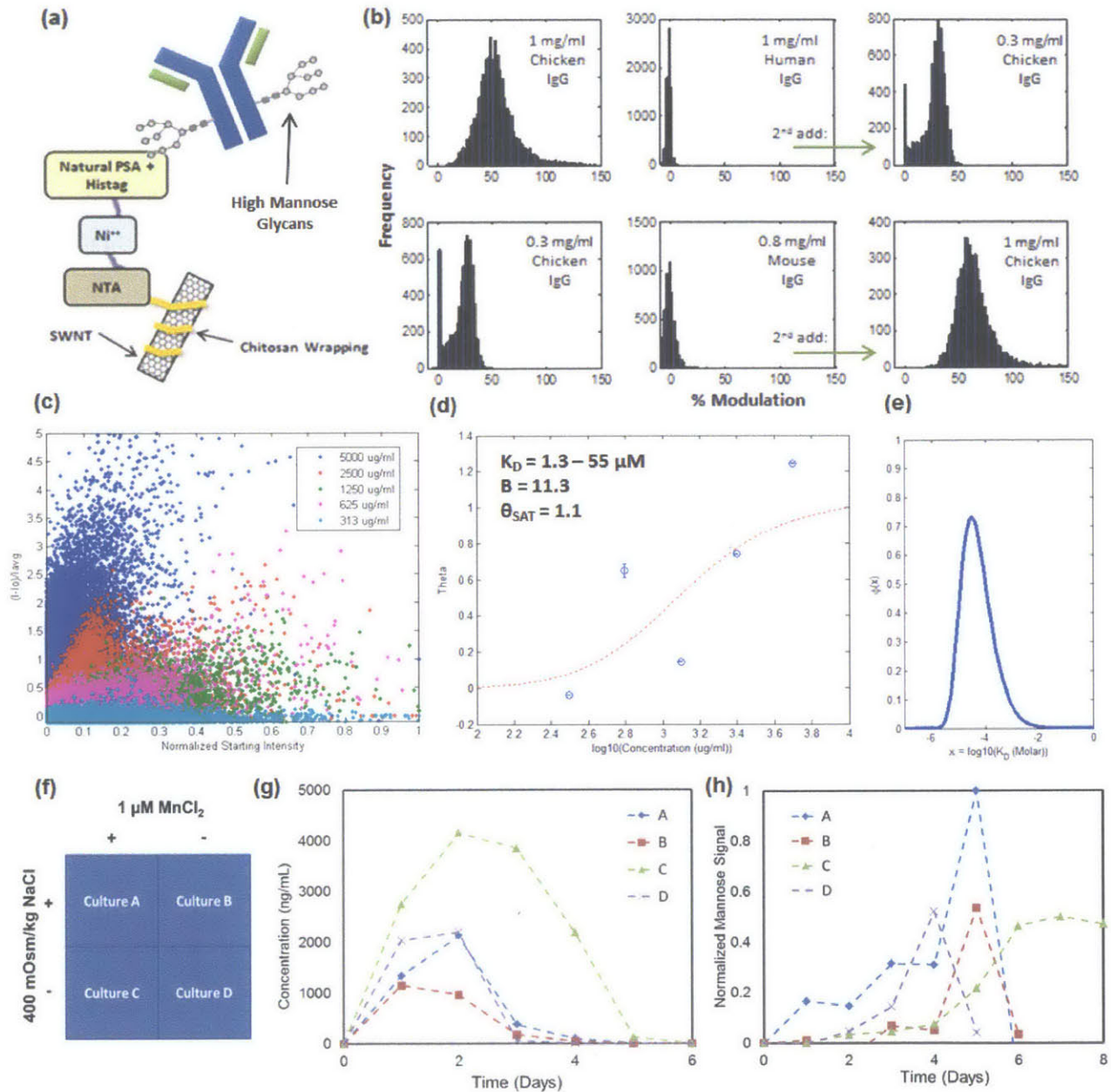


**Figure 4.14** – Experimental results from lyophilized IgG in PBS, murine IgG (TA99) from HEK cells, and human IgG from CHO cells screened on sensor arrays at various concentrations. a) All point sensor response ‘plumes’ and

fit calibration curves (b) yielding  $K_D$  95% confidence intervals, B-fit parameters, and angle saturation values ( $\theta_{SAT}$ ) calculated from the starting intensity distribution of each sensor batch (Eq. 5). c) Relation of B-fit parameter and  $K_D$  skewness parameter ( $\beta$ ) as found by simulation results. d) Measured  $K_D$  distributions using assumed Weibull PDF. e)  $K_D$  histograms directly calculated from sensor response assuming a constant functionalization value – revealing the approximate shape of the true PDF.

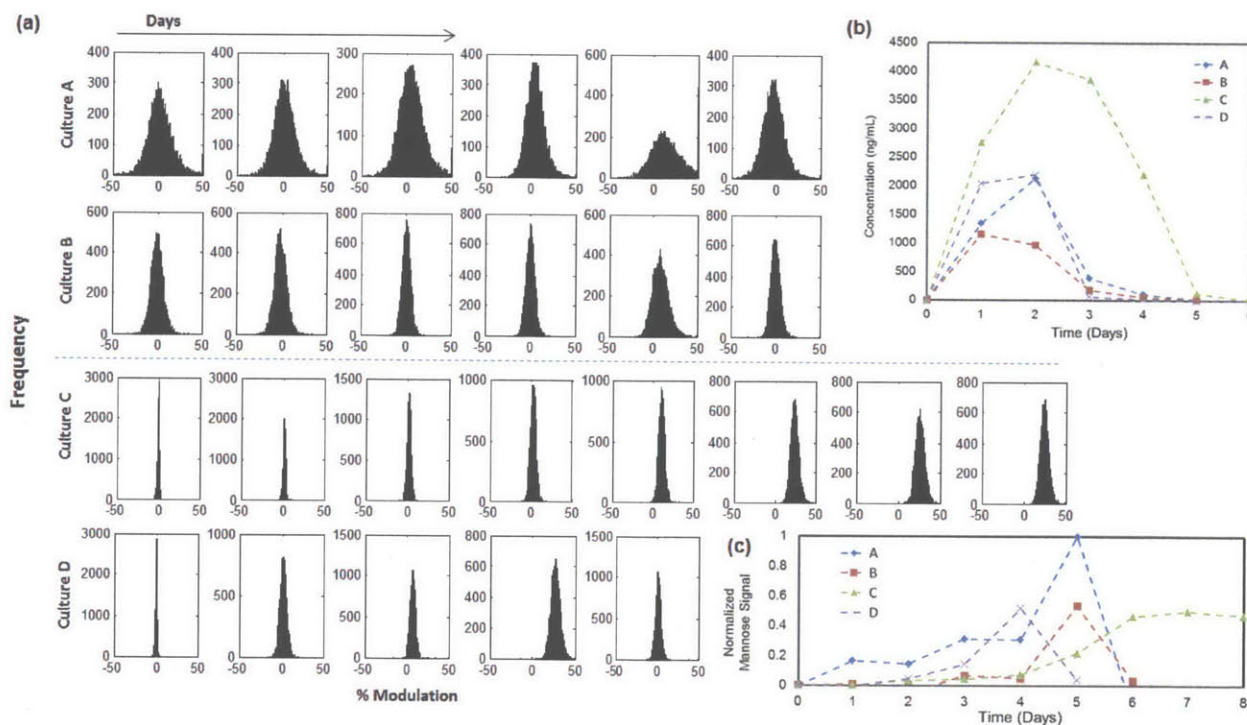
#### 4.4.4 Hypermannosylation Detection – Weakly affined interactions on nanosensor arrays

Another advantage of nanosensor arrays is their ability to report weak binding events – a greater number of individual sensor sites increases the probability of a detection event and this event is not averaged to null with other non-responsive sites, as in the case for an ensemble sensor. The label-free nature of the platform is also beneficial to detecting weakly-affined ligands since it requires no washing steps. By swapping out the His-tagged Protein A with a His-tagged, mannose-specific plant lectin, *Pisum sativum* agglutinin (PSA) the sensor platform can detect specifically high mannose content IgG (Fig 4.15a). Different species of IgG were initially used to test this concept. Chicken IgG contains an appreciable amount of glycoforms with high mannose content (>40% of population) whereas these are virtually absent in human and mouse IgG.<sup>28</sup> The SWNT sensor responses to human, mouse, and chicken IgG in PBS align with these findings and confirm that the platform can detect mannose species with the PSA lectin specifically (Fig 4.15b). A well-characterized, sample of IgG with high-mannose content from a fungal expression system (Novartis) was then used to further validate detection and obtain a calibration curve and  $K_D$  distribution as before (Fig 4.15c-e). The mannose IgG – PSA affinity ( $K_D = 1.3 - 55 \mu\text{M}$  95% confidence interval) is comparable to literature values ( $\mu\text{M} > K_D > \text{mM}$  for lectin-glycan interactions<sup>68</sup>) with a broad distribution (Fig 4.15e). This is expected as 60% of the IgG sample is aglycosylated (verified by MS analysis) and the remaining 40% bear differing lengths and structures of high mannose-type glycans (verified by released glycan analysis - Novartis).



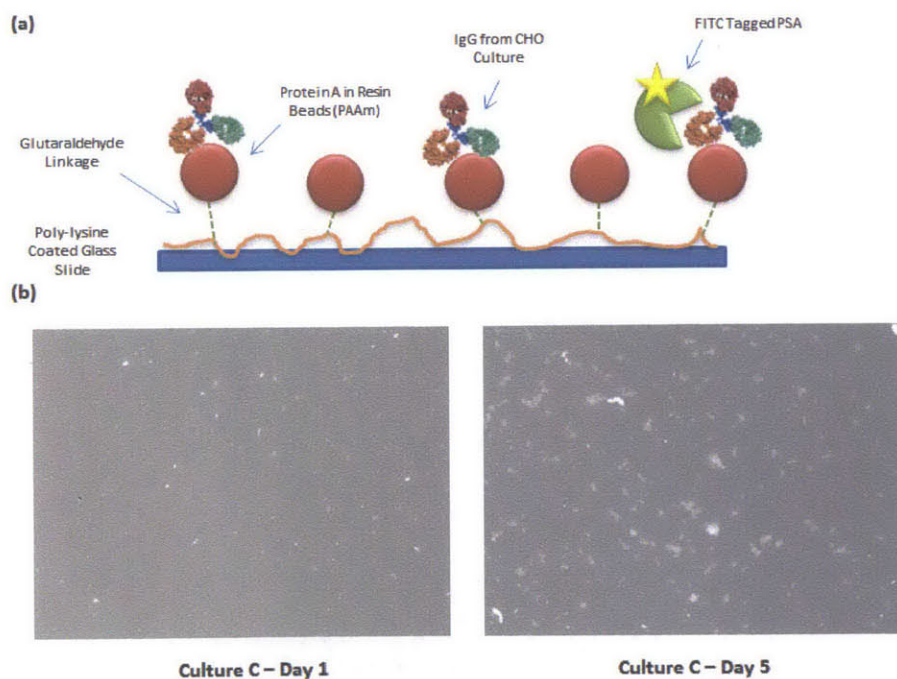
**Figure 4.15** – Hypermansylation detection on PSA-lectin conjugated sensor arrays. a) The weaker mannose-PSA lectin interactions can be transduced on independent nanosensors. b) Results of testing concept with chicken IgG which has > 40% high mannose glycoforms where human and mouse IgG do not contain these isoforms. c) All point ‘plume’ responses and calibration curve (d) as before with corresponding  $K_D$  distribution (e). f) Media compositions used to elicit hypermansylation in CHO cell culture. Culture was sampled for eight days showing expected changes in IgG titer from ELISA (g) and mannose content from SWNT sensor array (h).

It has also been shown that changing culture conditions such as media components can affect the onset and extent of IgG hypermannosylation. In order to further validate our platform, we designed an experiment for CHO cultures in which levels of NaCl were increased and a MnCl<sub>2</sub> additive was used while monitoring mannose content over time using traditional Peptide-N-Glycosidase F release and capillary electrophoresis.<sup>31</sup> We cultured four identical dishes of CHO cells in which we fed media compositions derived from this study (Fig 4.15f). The supernatants were collected after each 24 hour period, diluted to a standardized 10ng/ml IgG concentration and assayed on our PSA rendered sensor gels. The IgG concentrations were determined with ELISA (Fig 4.15g) and if below 10ng/ml, the sample was run at stock concentration. The resulting trends (Fig 4.15h) determined by the mean percent modulation from the nanosensor array distributions (Fig 4.16) match those found in the previous study: 1) increased mannose content as culture time increases, 2) increased mannose from higher NaCl osmolality, and 3) delayed onset of hypermannosylation from MnCl<sub>2</sub> additive.



**Figure 4.16** - The CHO cultures were fed different media compositions (Figure 4.15f) for ten days and then assayed for IgG concentration (Figure 4.16b) and diluted to 10ng/ml each. The samples were then assayed for mannose content. a) Raw trends in mannoseylation of each 24hr period diluted to 10ng/ml presented as distributions of all SWNT modulation. b) IgG titer for each of the culture days (by ELISA). c) General trends of mannoseylation derived from the distribution averages in (a).

The presence of mannose in these samples was confirmed by surface staining with fluorescently tagged PSA (Fig 4.17a). Protein A conjugated to a porous resin bead (Pierce 53139) was immobilized to a poly-lysine coated glass surface *via* glutaraldehyde in 2.5mm diameter Teflon patterned wells. The daily CHO samples diluted to 10ng/ml IgG (used in Fig 4.16 assay) were then spotted on the glass slide at 20µl. The IgG was allowed to interact for 2 hours and the chip was then washed in PBS. The chip was then spotted with 300µg/ml BSA to block non-specific sites and washed with PBS again. Finally FITC-conjugated PSA (Vector Labs FL-1051) was spotted on the sample at 25ug/ml and allowed to interact overnight. The samples were then washed with PBS and assayed on an inverted microscope reading the visible FITC emission. Qualitatively we observed more conjugated FITC-PSA on the days with the peak mannose found by the nanosensor platform (Fig 4.17b). This supports the finding of mannose in the IgG samples. The low concentration (< 4µg/ml) of the CHO hypermannosylation culture samples and low culture volumes (3 ml) make it very difficult to assay with current analytical techniques. We note that the sensitivity demonstrated by the array exceeded the capabilities of established glycan characterization tools; the same samples were captured on protein A columns, cleaved and interrogated using LC/MS but there was not adequate signal to resolve glycosylated species (Novartis).

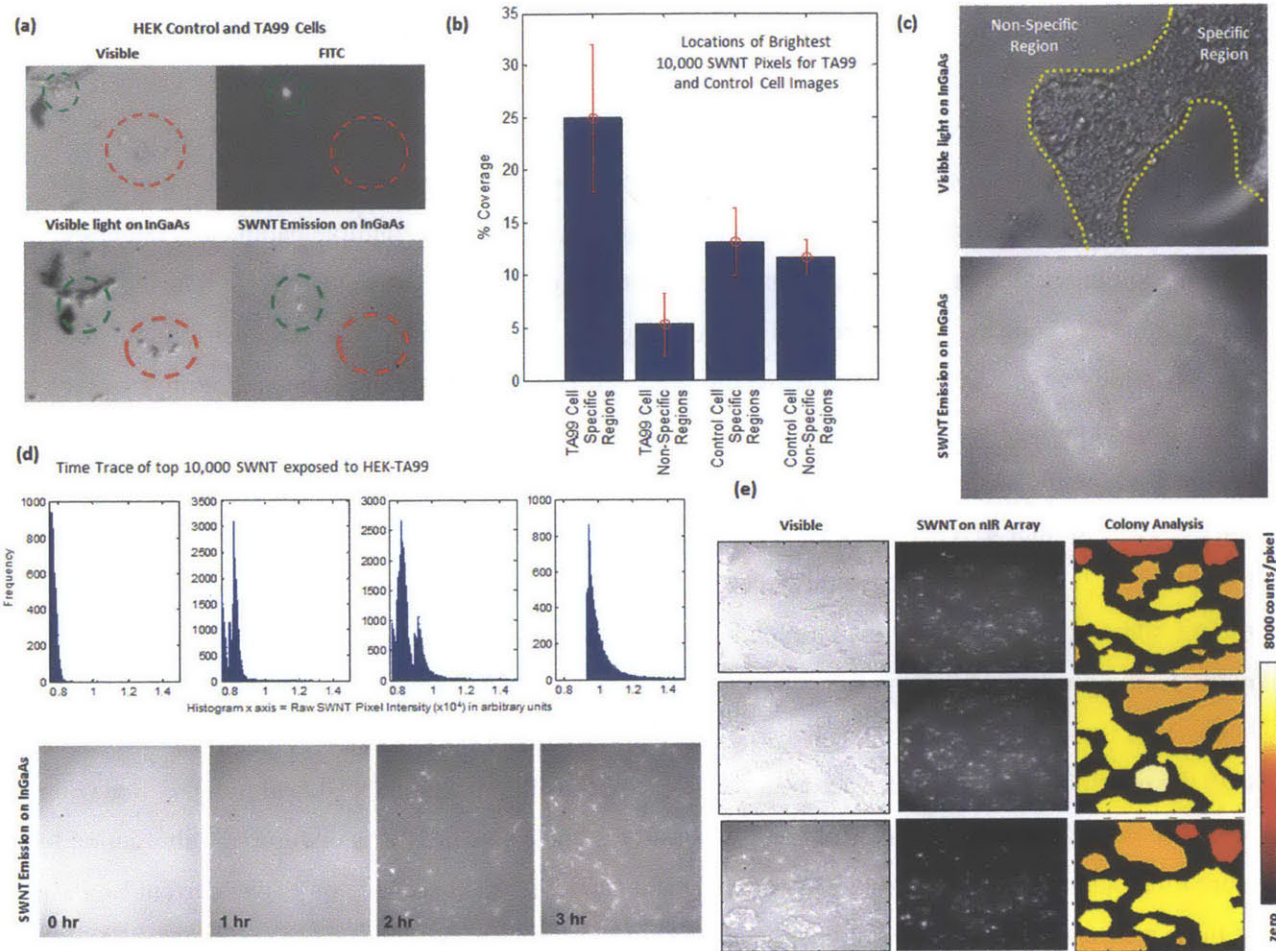


**Figure 4.17** – Surface staining experiment to confirm the presence of mannose in the CHO culture samples. a) A surface is treated to bind the IgG in the CHO samples and a FITC-labeled PSA lectin is used to stain the mannose content. b) Representative FITC fluorescent images of the stained surface showing a marked increase in mannose content as found by the SWNT nanosensors.

#### 4.4.5 Nanosensor Arrays for Monitoring Local Cell Colony Production

A gel with an imbedded array of nanosensors can be used to screen local production of cells. Unfortunately single cells are very difficult to culture for long periods of time on the current porous platform (little indications of healthy, single cell adherence). The single cells that did culture well, however, displayed colocalization of IgG production on a Protein A-incubated gel (Fig 4.18a). By seeding a greater number of cells, larger colonies of cells formed on the porous gel surface and were able to produce for longer time periods (24 hours). By analyzing images of control and IgG-producing HEK cells islands after 24 hours of production, we observed a statistical difference between the two profiles of production. By ranking the brightest 1000 SWNT and then querying their location, there is a greater localization of the bright SWNT under IgG-producing islands where as they are evenly or randomly distributed within and outside of the control cell islands (Fig 4.18b,c – Code in Appendix 4.3). We then plated HEK cells producing IgG on a gel and acquired multiple images of the nIR intensities at 0, 1, 2, and 3 hours. Histograms of the 1000 brightest SWNT pixels in these images show a ‘turn-on’ trend that is likely due to IgG production (Fig 4.18d). Finally, large HEK islands of HEK cells were allowed to grow overnight on a Protein A gel. The nIR response was clearly co-localized under each of the islands, and the response was summed, averaged over the island area and each island was ranked based on productivity (Fig 4.18e).



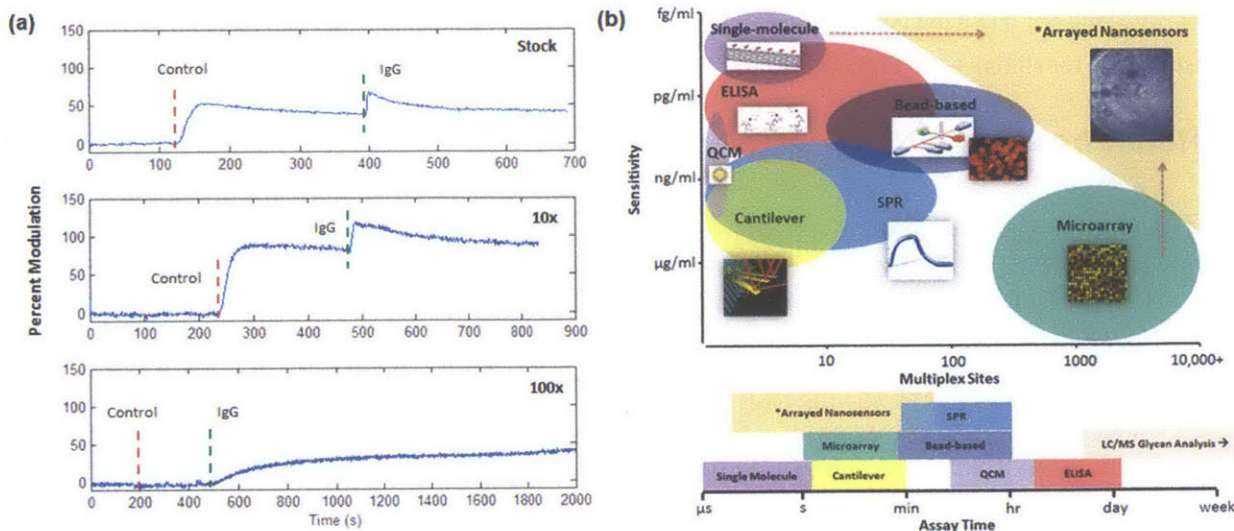


**Figure 4.18** - Sensor response to local cell production. a) Qualitative images of control and IgG producing cells showing colocalization of SWNT response. b) Location of top 1000 SWNT in IgG and control cell images, presented as coverage percentage of cell area (specific) and gel area (non-specific) in each image (c). d) Dynamic response of SWNT sensors to IgG producing cells plated on gel presented as distributions of top 1000 SWNT from images and micrographs of scaled SWNT intensity. e) Mapping visible IgG producing cell islands, the colocalized SWNT signal underneath, and ranking the islands' productivity based on intensity normalized by island area.

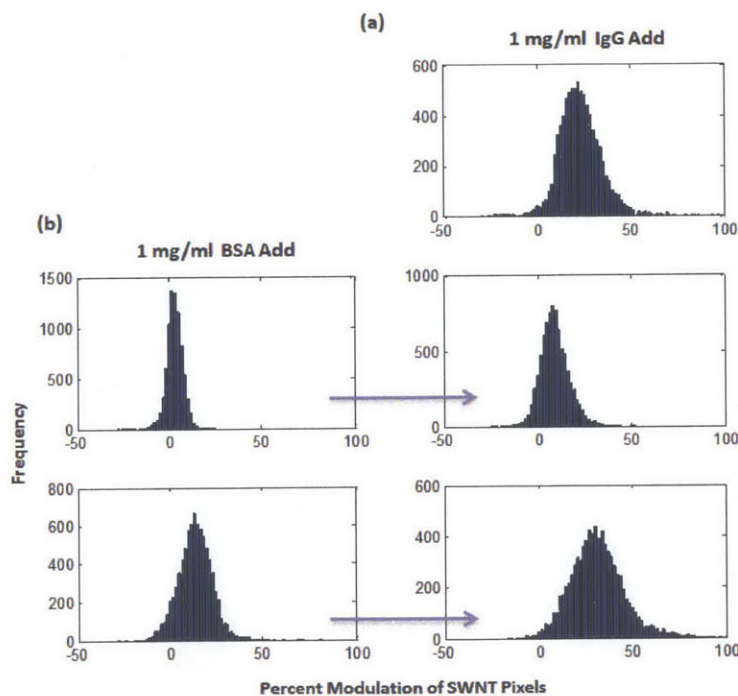
## 4.4 Conclusion

Here we have reported on carbon nanotube-based fluorescent sensor arrays for monitoring distributions of  $K_D$ , hypermannosylation, and local cellular production of IgG with clear implications in biomanufacturing. The platform was demonstrated with lyophilized IgG in PBS as well as characterizing freshly-expressed IgG in complex media from three different cellular expression systems: HEK, CHO, and a fungal cell line. The sensor array was rendered specific to mannose with PSA-lectin and trends in metabolically-induced hypermannosylation from a previous study were confirmed. Finally, local

production of IgG from HEK cell colonies cultured on sensor arrays was monitored. Better upstream colony selection could be performed with a sensor gel optimized for healthy cell culture. The cell colonies could be exposed to various culture and media conditions and their productivity and glycosylation patterns could be monitored in real time. This platform could lead to more rapid and informed selection of master cell lines and culture conditions based on multiple parameters rather than picking colonies based on static snapshots of productivity provided by current assays.<sup>23</sup> During production, sensor arrays in a microfluidic platform could monitor product titer,  $K_D$  distribution, and glycosylation by periodically sampling the bioreactor, filtering cellular components, diluting to a set level depending on the cell line's average productivity (Fig 6a), read the fluorescent signal, and then regenerate for the next sample. We have demonstrated that the Protein A gels can be regenerated using a pH 3.0 release wash, similar to regeneration of Protein A purification columns with little loss of sensitivity (Supplement 2h). Detection of mannose has been validated here but other glycans of interest (galactose, fucose, sialic acids, and non-human, immunogenic glycans like gal- $\alpha$  1,3-gal<sup>69</sup>) could also be detected by multiplexing portions of the nanosensor array with different His-tag lectins. The longstanding goal of nanosensor arrays is to preserve the sensitivity and analytical advantages of single-molecule nanosensors with the multiplexing ability of macroscale techniques, thus filling an untapped analytical regime (Figure 6b). The fast assay time (< 5 min) of nanosensor arrays could also provide a disruptive<sup>70</sup> alternative to the more time intensive ELISA and LC/MS analytics that are currently used (Fig 6b). The current limitations of these arrays are the intrinsic variances caused by non-automated production in small batches (16-32 gels per batch). Small variations in polymer casting time, initiator concentration, and washing procedures result in gels with varying levels of functionalization and sensitivity. A standardized gel from an automated printing/production system could reduce this variance and provide a robust tool for biomanufacturing analytics and beyond. In the following chapter we report on progress to this goal.



**Figure 4.19** – a) Ensemble response of high concentration IgG bioreactor supernatant delivered from a fungal cell expression system assayed at different dilutions along with samples from the non-transfected cell line that was cultured in parallel delivering a similar high background concentration of proteins. 100x dilution necessary to suppress background proteins from saturating the sensor array. b) Experimental domain addressed by arrayed nanosensors in comparison to existing technologies. Also mapped are approximate assay times of each technique.



**Figure 4.20** – Response of recycled gel to IgG (a) and to BSA followed by IgG (b). Gels had been tested with IgG, washed with pH 3.0 buffer (0.1M acetic acid), and then washed with PBS (1x pH 7.4). The IgG response is as before (Fig 4.20a) but the nonspecific response to BSA (Fig 4.20b) has increased. This is likely due to denatured

Protein A that did not retain the best specificity upon refolding. This effect could be studied and optimized to find the best process for eluting bound IgG while retaining specificity for future use. This will be an ongoing study.

## 4.5 References

- 1 Kingsmore, S. F. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nature Reviews Drug Discovery* 5, 310-320, doi:10.1038/nrd2006 (2006).
- 2 Nielsen, U. B. & Geierstanger, B. H. Multiplexed sandwich assays in microarray format. *Journal of Immunological Methods* 290, 107-120, doi:10.1016/j.jim.2004.04.012 (2004).
- 3 Epstein, J. R., Biran, I. & Walt, D. R. Fluorescence-based nucleic acid detection and microarrays. *Analytica Chimica Acta* 469, 3-36, doi:10.1016/s0003-2670(02)00030-2 (2002).
- 4 Cheng, M. M. C. *et al.* Nanotechnologies for biomolecular detection and medical diagnostics. *Current Opinion in Chemical Biology* 10, 11-19, doi:10.1016/j.cbpa.2006.01.006 (2006).
- 5 Schweitzer, B. & Kingsmore, S. F. Measuring proteins on microarrays. *Current Opinion in Biotechnology* 13, 14-19, doi:10.1016/s0958-1669(02)00278-1 (2002).
- 6 Wilson, D. S. & Nock, S. Recent developments in protein microarray technology. *Angewandte Chemie-International Edition* 42, 494-500, doi:10.1002/anie.200390150 (2003).
- 7 Niemeyer, C. M. Functional devices from DNA and proteins. *Nano Today* 2, 42-52, doi:10.1016/s1748-0132(07)70058-0 (2007).
- 8 Tam, J. M., Song, L. & Walt, D. R. DNA detection on ultrahigh-density optical fiber-based nanoarrays. *Biosensors & Bioelectronics* 24, 2488-2493, doi:10.1016/j.bios.2008.12.034 (2009).
- 9 Tan, C. P., Cipriany, B. R., Lin, D. M. & Craighead, H. G. Nanoscale Resolution, Multicomponent Biomolecular Arrays Generated By Aligned Printing With Parylene Peel-Off. *Nano Letters* 10, 719-725, doi:10.1021/nl903968s (2010).
- 10 Werblin, T. P. & Siskind, G. W. DISTRIBUTION OF ANTIBODY AFFINITIES - TECHNIQUE OF MEASUREMENT. *Immunochemistry* 9, 987-&, doi:10.1016/0019-2791(72)90110-3 (1972).
- 11 Pierson, L., Allauzen, S., Blumenthal, M. & Rosenberg, A. An automated method for determination of antibody affinity distribution functions with nanogram quantities. *Journal of Immunological Methods* 211, 97-109, doi:10.1016/s0022-1759(97)00204-4 (1998).
- 12 Steensgaard, J., Steward, M. W. & Frich, J. R. THE SIGNIFICANCE OF ANTIBODY-AFFINITY HETEROGENEITY IN ANTIGEN-ANTIBODY REACTIONS DEMONSTRATED BY COMPUTER-SIMULATION. *Molecular Immunology* 17, 689-698, doi:10.1016/0161-5890(80)90138-8 (1980).
- 13 Henriques, J. G., Buziol, S., Stocker, E., Voogd, A. & Menezes, J. C. in *Optical Sensor Systems in Biotechnology* Vol. 116 *Advances in Biochemical Engineering-Biotechnology* (ed G. Rao) 73-97 (2009).
- 14 Pizarro, S. A., Dinges, R., Adams, R., Sanchez, A. & Winter, C. Biomufacturing Process Analytical Technology (PAT) Application for Downstream Processing: Using Dissolved Oxygen as an Indicator of Product Quality for a Protein Refolding Reaction. *Biotechnology and Bioengineering* 104, 340-351, doi:10.1002/bit.22382 (2009).
- 15 Rodrigues, M. E., Costa, A. R., Henriques, M., Azeredo, J. & Oliveira, R. Technological Progresses in Monoclonal Antibody Production Systems. *Biotechnology Progress* 26, 332-351, doi:10.1002/btpr.348 (2010).
- 16 Holliger, P. & Hudson, P. J. Engineered antibody fragments and the rise of single domains. *Nature Biotechnology* 23, 1126-1136, doi:10.1038/nbt1142 (2005).
- 17 Butler, M. Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. *Applied Microbiology and Biotechnology* 68, 283-291, doi:10.1007/s00253-005-1980-8 (2005).

- 18 Farid, S. S. Process economics of industrial monoclonal antibody manufacture. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 848, 8-18, doi:10.1016/j.jchromb.2006.07.037 (2007).
- 19 Glennie, M. J. & van de Winkel, J. G. J. Renaissance of cancer therapeutic antibodies. *Drug Discovery Today* 8, 503-510, doi:10.1016/s1359-6446(03)02714-4 (2003).
- 20 Adams, G. P. & Weiner, L. M. Monoclonal antibody therapy of cancer. *Nature Biotechnology* 23, 1147-1157, doi:10.1035/nbt1137 (2005).
- 21 Hacker, D. L., De Jesus, M. & Wurm, F. M. 25 years of recombinant proteins from reactor-grown cells — Where do we go from here? *Biotechnology Advances* 27, 1023-1027, doi:<http://dx.doi.org/10.1016/j.biotechadv.2009.05.008> (2009).
- 22 Browne, S. M. & Al-Rubeai, M. Selection methods for high-producing mammalian cell lines. *Trends in Biotechnology* 25, 425-432, doi:<http://dx.doi.org/10.1016/j.tibtech.2007.07.002> (2007).
- 23 Burke, J. F. *et al.* Using ClonePix FL to Assess Monoclonality. *Genetic Engineering & Biotechnology News* 29, 38-39 (2009).
- 24 Elliott, S. *et al.* Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nature Biotechnology* 21, 414-421, doi:10.1038/nbt799 (2003).
- 25 Hermeling, S., Crommelin, D. J. A., Schellekens, H. & Jiskoot, W. Structure-immunogenicity relationships of therapeutic proteins. *Pharmaceutical Research* 21, 897-903, doi:10.1023/B:PHAM.0000029275.41323.a6 (2004).
- 26 Jefferis, R. Glycosylation of recombinant antibody therapeutics. *Biotechnology Progress* 21, 11-16, doi:10.1021/bp040016j (2005).
- 27 Jefferis, R. Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery* 8, 226-234, doi:10.1038/nrd2804 (2009).
- 28 Raju, T. S., Briggs, J. B., Borge, S. M. & Jones, A. J. S. Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology* 10, 477-486, doi:10.1093/glycob/10.5.477 (2000).
- 29 van Berkel, P. H. C. *et al.* N-Linked Glycosylation is an Important Parameter for Optimal Selection of Cell Lines Producing Biopharmaceutical Human IgG. *Biotechnology Progress* 25, 244-251, doi:10.1021/bp.92 (2009).
- 30 Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology* 22, 1393-1398, doi:10.1038/nbt1026 (2004).
- 31 Pacis, E., Yu, M., Autsen, J., Bayer, R. & Li, F. Effects of Cell Culture Conditions on Antibody N-linked Glycosylation-What Affects High Mannose 5 Glycoform. *Biotechnology and Bioengineering* 108, 2348-2358, doi:10.1002/bit.23200 (2011).
- 32 Gramer, M. J. *et al.* Modulation of Antibody Galactosylation Through Feeding of Uridine, Manganese Chloride, and Galactose. *Biotechnology and Bioengineering* 108, 1591-1602, doi:10.1002/bit.23075 (2011).
- 33 Lee, S. Y. *et al.* Effect of process change from perfusion to fed-batch on product comparability for biosimilar monoclonal antibody. *Process Biochemistry* 47, 1411-1418, doi:10.1016/j.procbio.2012.05.017 (2012).
- 34 Ahn, W. S., Jeon, J. J., Jeong, Y. R., Lee, S. J. & Yoon, S. K. Effect of Culture Temperature on Erythropoietin Production and Glycosylation in a Perfusion Culture of Recombinant CHO Cells. *Biotechnology and Bioengineering* 101, 1234-1244, doi:10.1002/bit.22006 (2008).
- 35 Muthing, J. *et al.* Effects of buffering conditions and culture pH on production rates and glycosylation of clinical phase I anti-melanoma mouse IgG3 monoclonal antibody R24. *Biotechnology and Bioengineering* 83, 321-334, doi:10.1002/bit.10673 (2003).
- 36 Schmelzer, A. E. & Miller, W. M. Hyperosmotic stress and elevated pCO<sub>2</sub> alter monoclonal antibody charge distribution and monosaccharide content. *Biotechnology Progress* 18, 346-353, doi:10.1021/bp010187d (2002).

- 37 Serrato, J. A., Palomares, L. A., Meneses-Acosta, A. & Ramirez, O. T. Heterogeneous conditions in dissolved oxygen affect N-glycosylation but not productivity of a monoclonal antibody in hybridoma cultures. *Biotechnology and Bioengineering* 88, 176-188, doi:10.1002/bit.20232 (2004).
- 38 Majid, F. A. A., Butler, M. & Al-Rubeai, M. Glycosylation of an immunoglobulin produced from a murine hybridoma cell line: The effect of culture mode and the anti-apoptotic gene, bcl-2. *Biotechnology and Bioengineering* 97, 156-169, doi:10.1002/bit.21207 (2007).
- 39 Bumbaca, D., Boswell, C. A., Fielder, P. J. & Khawli, L. A. Physicochemical and Biochemical Factors Influencing the Pharmacokinetics of Antibody Therapeutics. *Aaps Journal* 14, 554-558, doi:10.1208/s12248-012-9369-y (2012).
- 40 Goetze, A. M. *et al.* High-mannose glycans on the Fc region of therapeutic IgG antibodies increase serum clearance in humans. *Glycobiology* 21, 949-959, doi:10.1093/glycob/cwr027 (2011).
- 41 Lee, K. S., Boccazzi, P., Sinsky, A. J. & Ram, R. J. Microfluidic chemostat and turbidostat with flow rate, oxygen, and temperature control for dynamic continuous culture. *Lab on a Chip* 11, 1730-1739, doi:10.1039/c1lc20019d (2011).
- 42 Reuel, N. F., Mu, B., Zhang, J., Hinckley, A. & Strano, M. S. Nanoengineered glycan sensors enabling native glycoprofiling for medicinal applications: towards profiling glycoproteins without labeling or liberation steps. *Chemical Society Reviews* 41, 5744-5779, doi:10.1039/c2cs35142k (2012).
- 43 Ahn, J. H. *et al.* Label-Free, Single Protein Detection on a Near-Infrared Fluorescent Single-Walled Carbon Nanotube/Protein Microarray Fabricated by Cell-Free Synthesis. *Nano Letters* 11, 2743-2752, doi:10.1021/nl201033d (2011).
- 44 Reuel, N. F. *et al.* Transduction of Glycan-Lectin Binding Using Near-Infrared Fluorescent Single-Walled Carbon Nanotubes for Glycan Profiling. *Journal of the American Chemical Society* 133, 17923-17933, doi:10.1021/ja2074938 (2011).
- 45 Kalcioğlu, Z. I., Mahmoodian, R., Hu, Y. H., Suo, Z. G. & Van Vliet, K. J. From macro- to microscale poroelastic characterization of polymeric hydrogels via indentation. *Soft Matter* 8, 3393-3398, doi:10.1039/c2sm06825g (2012).
- 46 Ruchel, R., Steere, R. L. & Erbe, E. F. TRANSMISSION-ELECTRON MICROSCOPIC OBSERVATIONS OF FREEZE-ETCHED POLYACRYLAMIDE GELS. *Journal of Chromatography* 166, 563-575, doi:10.1016/s0021-9673(00)95641-3 (1978).
- 47 Hu, Y. H., Zhao, X. H., Vlassak, J. J. & Suo, Z. G. Using indentation to characterize the poroelasticity of gels. *Applied Physics Letters* 96, doi:121904 10.1063/1.3370354 (2010).
- 48 Chan, E. P., Hu, Y. H., Johnson, P. M., Suo, Z. G. & Stafford, C. M. Spherical indentation testing of poroelastic relaxations in thin hydrogel layers. *Soft Matter* 8, 1492-1498, doi:10.1039/c1sm06514a (2012).
- 49 Armstrong, J. K., Wenby, R. B., Meiselman, H. J. & Fisher, T. C. The hydrodynamic radii of macromolecules and their effect on red blood cell aggregation. *Biophysical Journal* 87, 4259-4270, doi:10.1529/biophysj.104.047746 (2004).
- 50 Hu, T., Fu, Q., Chen, P., Zhang, K. & Guo, D. Y. Generation of a stable mammalian cell line for simultaneous expression of multiple genes by using 2A peptide-based lentiviral vector. *Biotechnology Letters* 31, 353-359, doi:10.1007/s10529-008-9882-3 (2009).
- 51 Thomson, T. M., Mattes, M. J., Roux, L., Old, L. J. & Lloyd, K. O. PIGMENTATION-ASSOCIATED GLYCOPROTEIN OF HUMAN MELANOMAS AND MELANOCYTES - DEFINITION WITH A MOUSE MONOCLONAL-ANTIBODY. *Journal of Investigative Dermatology* 85, 169-174, doi:10.1111/1523-1747.ep12276608 (1985).
- 52 Stern, P. *et al.* A system for Cre-regulated RNA interference in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 105, 13895-13900, doi:10.1073/pnas.0806907105 (2008).
- 53 Kuroda, H., Kutner, R. H., Bazan, N. G. & Reiser, J. Simplified lentivirus vector production in protein-free media using polyethylenimine-mediated transfection. *Journal of Virological Methods* 157, 113-121, doi:10.1016/j.jviromet.2008.11.021 (2009).

- 54 Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nature Biotechnology* 15, 871-875, doi:10.1038/nbt0997-871 (1997).
- 55 Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *Journal of Virology* 72, 8463-8471 (1998).
- 56 Hezareh, M., Hessel, A. J., Jensen, R. C., van de Winkel, J. G. J. & Parren, P. Effector function activities of a panel of mutants of a broadly neutralizing antibody against human immunodeficiency virus type 1. *Journal of Virology* 75, 12161-12168, doi:10.1128/jvi.75.24.12161-12168.2001 (2001).
- 57 Wang, J., Gonzalez, A. D. & Ugaz, V. M. Tailoring Bulk Transport in Hydrogels through Control of Polydispersity in the Nanoscale Pore Size Distribution. *Advanced Materials* 20, 4482-4489, doi:10.1002/adma.200801303 (2008).
- 58 Brege, J. J., Gallaway, C. & Barron, A. R. Fluorescence Quenching of Single-Walled Carbon Nanotubes with Transition-Metal Ions. *Journal of Physical Chemistry C* 113, 4270-4276, doi:10.1021/jp808667b (2009).
- 59 Wu, P. & Yan, X. P. Ni(2+)-modulated homocysteine-capped CdTe quantum dots as a turn-on photoluminescent sensor for detecting histidine in biological fluids. *Biosensors & Bioelectronics* 26, 485-490, doi:10.1016/j.bios.2010.07.068 (2010).
- 60 Strano, M. S. *et al.* The Role of Surfactant Adsorption during Ultrasonication in the Dispersion of Single-Walled Carbon Nanotubes. *Journal of Nanoscience and Nanotechnology* 3, 81-86, doi:10.1166/jnn.2003.194 (2003).
- 61 Blackburn, J. L. *et al.* Protonation effects on the branching ratio in photoexcited single-walled carbon nanotube dispersions. *Nano Letters* 8, 1047-1054, doi:10.1021/nl072809g (2008).
- 62 Barone, P. W. *et al.* Modulation of Single-Walled Carbon Nanotube Photoluminescence by Hydrogel Swelling. *Acs Nano* 3, 3869-3877, doi:10.1021/nn901025x (2009).
- 63 Nohldén, S. *Affinity Determination of Protein A Domains to IgG subclasses by Surface Plasmon Resonance* Masters thesis, Linköpings universitet, (2008).
- 64 Saha, K., Bender, F. & Gizeli, E. Comparative study of IgG binding to proteins G and A: Nonequilibrium kinetic and binding constant determination with the acoustic waveguide device. *Analytical Chemistry* 75, 835-842, doi:10.1021/ac0204911 (2003).
- 65 Richman, D. D., Cleveland, P. H., Oxman, M. N. & Johnson, K. M. THE BINDING OF STAPHYLOCOCCAL PROTEIN-A BY THE SERA OF DIFFERENT ANIMAL SPECIES. *Journal of Immunology* 128, 2300-2305 (1982).
- 66 Schersch, K. *et al.* Systematic Investigation of the Effect of Lyophilizate Collapse on Pharmaceutically Relevant Proteins I: Stability after Freeze-Drying. *Journal of Pharmaceutical Sciences* 99, 2256-2278, doi:10.1002/jps.22000 (2010).
- 67 Burry, R. W. Ch. 2, 7-16 (Springer, 2010).
- 68 Hirabayashi, J. Concept, strategy and realization of lectin-based glycan profiling. *Journal of Biochemistry* 144, 139-147, doi:10.1093/jb/mvn043 (2008).
- 69 Chung, C. H. *et al.* Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *New England Journal of Medicine* 358, 1109-1117 (2008).
- 70 Bower, J. L. & Christensen, C. M. DISRUPTIVE TECHNOLOGIES - CATCHING THE WAVE. *Harvard Business Review* 73, 43-53 (1995).

## Appendix 4.1

This code was used to analyze the force relaxation curves presented in the manuscript to identify the hydrogel pore size.

```
function AFM_PRI_Analysis_v4_8um
% Coded by Nigel F. Reuel on 8.7.2012
```

```

% Used to analyze the PRI curves derived from AFM as the Van Vliet group
% has demonstrated - Analyzes the curves to find the following:
% 1. G = shear modulus
% 2. D = diffusion rate
% 3. Vs = Poisson's ratio
% 4. PS = Pore Size
%
% 8.8.2012 v2 provides 95% error limits on the pore calculation
% 8.8.2012 v3 corrects for drift present in low crosslink gels
% 8.27.12 v4 adjusted for new data set (8.23.12 experiments)
%
% ----- 8um Indentation Step ----- 30s indentation
%
% (1) Read in the raw data:
%
% Number of files, time of indentation, and data rate
NF = 18;
TI = 20; % In seconds
% Determine the recording rate:
Record_temp = dlmread('3p5p_8um_1.txt');
Time1 = Record_temp(5,1);
Time2 = Record_temp(4,1);
DR = 1/(Time1-Time2); % In Hertz
%
RD_8um = zeros(TI*DR,NF);
Nrows = floor(TI*DR);
for i = 1:NF
    Data_temp = dlmread(['3p5p_8um_',int2str(i),'.txt']);
    % Check and correct for drift, using last 5000 time points:
    L = floor(TI*DR);
    L2 = 5000;
    yp = Data_temp(L-L2+1:L,2);
    xp = (1:L2)';
    p = polyfit(xp,yp,1);
    plot(xp,yp)
    %p1 = slope and p2 = intercept
    % Correct for slope:
    p2 = 0;
    Data_Temp_Cor = zeros(L,1);
    for j = 1:L
        Data_Temp_Cor(j,1) = Data_temp(j,2) - p2*(j-1);
    end
    RD_8um(:,i) = Data_Temp_Cor(1:Nrows,1);
end

% Center the values so the initial Force is zero
CenterD_8um = zeros(Nrows,NF);
for i = 1:NF
    for j = 1:Nrows
        CenterD_8um(j,i) = RD_8um(j,i) - RD_8um(1,i);
    end
end

% Try plotting together to see what they look like:
X = (1:Nrows)';
plot(X,CenterD_8um)
%
% Good replication in 1-14

```



```

Avg8umCurve = zeros(Nrows,1);
ConfPosCurve = zeros(Nrows,1);
ConfNegCurve = zeros(Nrows,1);
SR = 1;
ER = 14;
for i = 1:Nrows
    Avg8umCurve(i,1) = mean(CenterD_8um(i,SR:ER));
    ConfPosCurve(i,1) = mean(CenterD_8um(i,SR:ER)) +
2*std(CenterD_8um(i,SR:ER));
    ConfNegCurve(i,1) = mean(CenterD_8um(i,SR:ER)) -
2*std(CenterD_8um(i,SR:ER));
end
% Plot the averaged curve with error bars
plot(X,Avg8umCurve,X,ConfPosCurve,'r--',X,ConfNegCurve,'r--')
csvwrite('AvgF_8um.csv',Avg8umCurve)
csvwrite('AvgFP_8um.csv',ConfPosCurve)
csvwrite('AvgFN_8um.csv',ConfNegCurve)
%
% Plot the force relax curve (w/ error bars)
[Fzero index] = max(Avg8umCurve);
Finf = mean(Avg8umCurve(Nrows-100:Nrows,1));
F = zeros(1:Nrows-index+1,1);
for i = index:Nrows
    Ft = Avg8umCurve(i,1);
    F(i-index+1,1) = (Ft - Finf)/(Fzero-Finf);
end
Xtime = ((1:Nrows-index+1)*1/DR)';
% ---- High error
[FzeroH indexH] = max(ConfPosCurve);
FinfH = mean(ConfPosCurve(Nrows-100:Nrows,1));
FH = zeros(1:Nrows-indexH+1,1);
for i = indexH:Nrows
    FtH = ConfPosCurve(i,1);
    FH(i-indexH+1,1) = (FtH - FinfH)/(FzeroH-FinfH);
end
XtimeH = ((1:Nrows-indexH+1)*1/DR)';
% ---- Low error
[FzeroL indexL] = max(ConfNegCurve);
FinfL = mean(ConfNegCurve(Nrows-100:Nrows,1));
FL = zeros(1:Nrows-indexL+1,1);
for i = indexL:Nrows
    FtL = ConfNegCurve(i,1);
    FL(i-indexL+1,1) = (FtL - FinfL)/(FzeroL-FinfL);
end
XtimeL = ((1:Nrows-indexL+1)*1/DR)';
% Plot together...
plot(Xtime,F,XtimeH,FH,'r--',XtimeL,FL,'r--')
% Save the F and X file
csvwrite('Force_8um.csv',F);
csvwrite('Time_8um.csv',Xtime);
% Solve for tau
R = 22.5*10^-6; %Radius of the bead
h = 8*10^-6; % Depth of the indentation
a = (R*h)^(1/2);
eta = 1*10^-3; % (N s/m2)
D0 = .000000002;
% ----- Calculated value -----

```

```

D = fminsearch(@ForceFunc,D0, [], F, a, Xtime);
G = Fzero*3/16/a/h;
Vs = 1-Fzero/Finf/2;
PS = 2*(eta/2*D*(1-2*Vs)/G/(1-Vs))^(1/2)*10^9 %Reported in nanometers!
% ----- High error value -----
DH = fminsearch(@ForceFunc,D0, [], FH, a, XtimeH);
GH = FzeroH*3/16/a/h;
VsH = 1-FzeroH/FinfH/2;
PSH = 2*(eta/2*DH*(1-2*VsH)/GH/(1-VsH))^(1/2)*10^9 %Reported in nanometers!
% ----- Low error value -----
DL = fminsearch(@ForceFunc,D0, [], FL, a, XtimeL);
GL = FzeroL*3/16/a/h;
VsL = 1-FzeroL/FinfL/2;
PSL = 2*(eta/2*DL*(1-2*VsL)/GL/(1-VsL))^(1/2)*10^9 %Reported in nanometers!
end

```

```

function FZet = ForceFunc(D0,F,a,Xtime)
rows = length(F);
ErrorSQR = 0;
for i = 1:rows
    tau = D0*Xtime(i,1)/a^2;
    ErrorSQR = (F(i,1) - (0.491*exp(-.908*(tau)^(1/2)) + 0.509*exp(-
1.679*tau)))^2 + ErrorSQR;
end
FZet = ErrorSQR;
end

```

## Appendix 4.2

The following code was used to simulate the effect of the KD distribution skewness parameter ( $\beta$ ) on the calibration fit parameter.

```

function Simulation_Effect_of_Beta
% Coded by Nigel F. Reuel on 1.8.12
% Used to explain:
% 1. How going from KD --> K --> LangTheta --> Plume of Data
% 2. How to fit the plume of data
% 3. Effect of changing KD distribution skewness parameter (Beta) on fit
%
% Number of sensor points to simulate:
Npoints = 10000;
% Number of concentration points between those that are graphically
represented!
BP = 4;
Nconc = BP*6+5;
logC = linspace(-9,-5,Nconc); % ## <----- Change Conc Range here!!
Lt = zeros(Npoints,Nconc);
Kvec = zeros(Npoints,1);
KDGlobalMean = 7;
%
% ##### KD Beta Parameter Loop Start
BetaMin = 2;
BetaMax = 45;

```

```

nB = 200; % ## <-----Change number of Beta parameters
here
BetaSpan = linspace(BetaMin,BetaMax,nB);
ThetaSAVE = zeros(nB,Nconc);
SigmaSAVE = zeros(nB,Nconc);
ThetaSAVEstd = zeros(nB,Nconc);
SigmaSAVEstd = zeros(nB,Nconc);
for Bloop = 1:nB
    Beta = BetaSpan(1,Bloop);

% ##### Replicate Loop START #####

NR = 6; % <----- Number of Replicates here
MuTheta = zeros(NR,Nconc);
S22 = zeros(NR,Nconc);
for Rep = 1:NR

for i = 1:Nconc
    for j = 1:Npoints
        K = 10^(wblrnd(KDGlobalMean,Beta));
        C = 10^(logC(1,i));
        Lt(j,i) = K*C/(1+K*C);
        Kvec(j,1) = K;
    end
    if Rep == NR
        if i == 1
            bins = linspace(-11.1,-2.9,200);
            KDL = log10(1./Kvec);
            [n x] = hist(KDL,bins);
            subplot(2,3,1) % <-----Plot 1
            bar(x,n)
            xlim([-11 -3])
            xlabel('log10(K_D)')
            ylabel('Frequency')
            title('K_D Distribution')
            text(-7,max(n)*.95,[' \leftarrow \beta = ',int2str(Beta)])
        end
        subplot(2,3,2) % <----- Plot 2
        [n x] = hist(Lt(:,i),200);
        if i == 1+BP
            plot(x,n,'b.')
        elseif i == 2*BP+2
            plot(x,n,'r.')
        elseif i == 3*BP+3
            plot(x,n,'g.')
        elseif i == 4*BP+4
            plot(x,n,'m.')
        elseif i == 5*BP+5
            plot(x,n,'k.')
        end
        xlim([0 1])
        hold on
    end
end
end
%Cspan = 10.^(logC)*10^9; % Reported in nM
if Rep == NR

```

```

    %hleg1 = legend([int2str(Cspan(1,1+BP)), '
nM'], [int2str(Cspan(1,2*BP+2)), ' nM'], [int2str(Cspan(1,3*BP+3)), '
nM'], [int2str(Cspan(1,4*BP+4)), ' nM'], [int2str(Cspan(1,5*BP+5)), ' nM']);
    %set(hleg1, 'Location', 'NorthWest')
    title('Langmuir Coverage')
    ylabel('Frequency')
    xlabel('\theta - Coverage Fraction')
    hold off
end
%
NSWNT = Npoints;
Length = zeros(NSWNT,Nconc);
Response = zeros(NSWNT,Nconc);
RespNorm = zeros(NSWNT,Nconc);
INorm = zeros(NSWNT,Nconc);
% Create matrix of SWNT lengths, then do the calculation:
for i = 1:Nconc
    for j = 1:NSWNT
        Length(j,i) = wblrnd(6871.7,2.2);
    end
end
% Values to Normalize SWNT Length
MaxL = max(Length);
MinL = min(Length);
MeanL = mean(Length);
%MeanLNorm = (MeanL - MinL) ./ (MaxL - MinL)
for i = 1:Nconc
    for j = 1:NSWNT
        Theta = Lt(j,i);
        L = Length(j,i);
        Response(j,i) = Theta*L;
        RespNorm(j,i) = Theta*L/MeanL(1,i);
        INorm(j,i) = (L - MinL(1,i)) / (MaxL(1,i) - MinL(1,i));
    end
end

end
% Plot the expected response at different concentrations
%subplot(3,3,6) %
%plot(INorm(:,1),RespNorm(:,1),'.b',INorm(:,2),RespNorm(:,2),'.r',INorm(:,3),
RespNorm(:,3),'.g',INorm(:,4),RespNorm(:,4),'.m',INorm(:,5),RespNorm(:,5),'.k
')
%legend([int2str(Cspan(1,1)), ' nM'], [int2str(Cspan(1,2)), '
nM'], [int2str(Cspan(1,3)), ' nM'], [int2str(Cspan(1,4)), '
nM'], [int2str(Cspan(1,5)), ' nM'])
%title('Simulated Sensor Response - No Functionalization Parameter')
%ylabel('(I_o*\theta)/I_a_v_g')
%xlabel('(I_o - I_m_i_n)/(I_m_a_x-I_m_i_n)')
% But the SWNT have been functionalized to a certain extent
PM_Ni = csvread('PM_Nickel.csv').*100;
%[n x] = hist(PM_Ni,200);
%subplot(3,3,7) %
%bar(x,n)
%xlim([-35 0])
%title('Nickel Functionalization')
%ylabel('Frequency')
%xlabel('% Quench')
%
```

```

% Read in the nickel quenching data to create your Gaussian fit:
%
Gfit = gmdistribution.fit(PM_Ni,1);
%GMean = Gfit.mu
%GSigma = Gfit.Sigma;
%Gm = mean(PM_Ni);
%Gs = std(PM_Ni);
%Gfit2 = gmdistribution(Gm,Gs,1);
%subplot(3,3,8) %
%X = linspace(-50,10,100)';
%Y = pdf(Gfit,X);
%plot(X,Y)
%xlim([-35 0])
%xlabel('% Quench')
%ylabel('f(x;\mu,\sigma)')
%title('Gaussian Fit of Functionalization Extent')
% Now plot what the experimental ouput would look like with
% functionalization distribution
FuncResponse = zeros(NSWNT,Nconc);
FuncRespNorm = zeros(NSWNT,Nconc);
for i = 1:Nconc
    for j = 1:NSWNT
        Theta = Lt(j,i);
        Io = Length(j,i);
        FuncRange = Io*(1/(1-random(Gfit)/-100) - 1);
        FuncResponse(j,i) = Theta*FuncRange;
        FuncRespNorm(j,i) = Theta*FuncRange/MeanL(1,i);
    end
end
if Rep == NR
    subplot(2,3,3) % <-----Plot 3

plot(INorm(:,1+BP),FuncRespNorm(:,1+BP),'.b',INorm(:,2*BP+2),FuncRespNorm(:,2
*BP+2),'.r',INorm(:,3*BP+3),FuncRespNorm(:,3*BP+3),'.g',INorm(:,4*BP+4),FuncR
espNorm(:,4*BP+4),'.m',INorm(:,5*BP+5),FuncRespNorm(:,5*BP+5),'.k')
    %hleg2 = legend([int2str(Cspan(1,1+BP)), '
nM'], [int2str(Cspan(1,2*BP+2)), ' nM'], [int2str(Cspan(1,3*BP+3)), '
nM'], [int2str(Cspan(1,4*BP+4)), ' nM'], [int2str(Cspan(1,5*BP+5)), ' nM']);
    %set(hleg2,'Location','NorthWest')
    title('Sensor Response - Cartesian')
    ylabel('(Io*(1/(1-PF))-1)*\theta/I_a_v_g')
    xlabel('(Io-I_m_i_n)/(I_m_a_x-I_m_i_n)')
    ylim([0 1.2])
end
%
% Now plot what this plume reduces to with your radial analysis
%{
ThetaMean = zeros(1,Nconc);
L = zeros(1,Nconc);
U = zeros(1,Nconc);
ThetaSig = zeros(1,Nconc);
Lsig = zeros(1,Nconc);
Usig = zeros(1,Nconc);
%}
MixGausData = zeros(3,2,Nconc);
for i = 1:Nconc
    Xn = INorm(:,i);

```

```

Yn = FuncRespNorm(:,i);
[THETA1 RHO1] = cart2pol(Xn,Yn);
% Plot according to the point density (to see what this looks like)...
%n1 = 100;
%n2 = 100;
%[z,C] = hist3([RHO1 THETA1],[n1 n2]);
%xb = C{1,1}; % RHO Values
%yb = C{1,2}; % THETA Values
clear X
X(:,1) = THETA1;
X(:,2) = RHO1;
if Rep == NR
    subplot(2,3,4) %<----- Plot 4
    if i == 1+BP
        plot(THETA1, RHO1, '.b')
        hold on
        Cloud1 = gmdistribution.fit(X,1);
        MU = Cloud1.mu;
        SIG = Cloud1.Sigma;
        MixGausData(1,1:2,i) = MU;
        MixGausData(2:3,1:2,i) = SIG(:, :,1);
        %ezcontourf(@(x,y)pdf(Cloud1,[x y]),[0 .8],[0 1]);

    elseif i == 2*BP+2
        plot(THETA1, RHO1, '.r')
        Cloud2 = gmdistribution.fit(X,1);
        MU = Cloud2.mu;
        SIG = Cloud2.Sigma;
        MixGausData(1,1:2,i) = MU;
        MixGausData(2:3,1:2,i) = SIG(:, :,1);
        %ezcontourf(@(x,y)pdf(Cloud2,[x y]),[0 .8],[0 1]);

    elseif i == 3*BP+3
        plot(THETA1, RHO1, '.g')
        Cloud3 = gmdistribution.fit(X,1);
        MU = Cloud3.mu;
        SIG = Cloud3.Sigma;
        MixGausData(1,1:2,i) = MU;
        MixGausData(2:3,1:2,i) = SIG(:, :,1);
        %ezcontourf(@(x,y)pdf(Cloud3,[x y]),[0 .8],[0 1]);

    elseif i == 4*BP+4
        plot(THETA1, RHO1, '.m')
        Cloud4 = gmdistribution.fit(X,1);
        MU = Cloud4.mu;
        SIG = Cloud4.Sigma;
        MixGausData(1,1:2,i) = MU;
        MixGausData(2:3,1:2,i) = SIG(:, :,1);
        %ezcontourf(@(x,y)pdf(Cloud4,[x y]),[0 .8],[0 1]);

    elseif i == 5*BP+5
        plot(THETA1, RHO1, '.k')
        Cloud5 = gmdistribution.fit(X,1);
        MU = Cloud5.mu;
        SIG = Cloud5.Sigma;
        MixGausData(1,1:2,i) = MU;

```

```

        MixGausData(2:3,1:2,i) = SIG(:, :, 1);
        %ezcontourf(@(x,y)pdf(Cloud5, [x y]), [0 .8], [0 1]);

    else
        CloudNull = gmdistribution.fit(X,1);
        MU = CloudNull.mu;
        SIG = CloudNull.Sigma;
        MixGausData(1,1:2,i) = MU;
        MixGausData(2:3,1:2,i) = SIG(:, :, 1);

    end
    hold on
else
    CloudNull = gmdistribution.fit(X,1);
    MU = CloudNull.mu;
    SIG = CloudNull.Sigma;
    MixGausData(1,1:2,i) = MU;
    MixGausData(2:3,1:2,i) = SIG(:, :, 1);
end

end

if Rep == NR
    xlim([0 .75])
    ylim([0 1])
    title('Sensor Response - Polar')
    hold off
    xlabel('\theta')
    ylabel('R')
    % Represent polar fits as bivariate distributions
    subplot(2,3,5)
    h = ezcontourf(@(x,y)pdf(Cloud1, [x y]), [0 .8], [0 .8], 100);
    set(h, 'LineStyle', 'none');
    hold on
    h = ezcontourf(@(x,y)pdf(Cloud3, [x y]), [0 .8], [0 .8], 100);
    set(h, 'LineStyle', 'none');
    hold on
    h = ezcontourf(@(x,y)pdf(Cloud5, [x y]), [0 .8], [0 .8], 100);
    set(h, 'LineStyle', 'none');
    hold on
    colormap(hot)
    title('Bivariate Gaussian PDF Fit')
    xlabel('\theta')
    ylabel('R')
    hold off
end
%
% Now in your last plot, plot the parameters of interest (Rho-Mu and Rho
% sigmall)
%
for i = 1:Nconc
    MuTheta(Rep,i) = MixGausData(1,1,i);
    S22(Rep,i) = MixGausData(2,1,i);
end

```

```

Rep
% ##### End for Replicate Loop #####
end
X = logC;
subplot(2,3,6)
[AX,H1,H2] = plotyy(X,mean(MuTheta),X,mean(S22),'plot');
set(get(AX(1),'Ylabel'),'String','\theta_\mu')
set(get(AX(2),'Ylabel'),'String','\theta_\sigma')
set(AX(1),'YLim',[0 .6])
set(AX(2),'YLim',[0 .1])
xlabel('log10(Conc)')
title('Calibration Curve')
set(H1,'LineStyle','--')
set(H2,'LineStyle',':')
%
% Set up the picture, save and close!

set(gcf,'Position',[5 5 1000 1000])
figure_name_fig = [int2str(Bloop),'.png'];
print(gcf,'-dpng',figure_name_fig);
close
ThetaSAVE(Bloop,:) = mean(MuTheta);
SigmaSAVE(Bloop,:) = mean(S22);
ThetaSAVEstd(Bloop,:) = std(MuTheta);
SigmaSAVEstd(Bloop,:) = std(S22);
Bloop
% ##### End for BetaLoop
end
% Save the calibration curve data:
csvwrite('ThetaMu_mean.csv',ThetaSAVE);
csvwrite('ThetaMu_std.csv',ThetaSAVEstd);
csvwrite('ThetaSig_mean.csv',SigmaSAVE);
csvwrite('ThetaSig_std.csv',SigmaSAVEstd);

end

```

### Appendix 4.3

The following code was used to analyze the local cell production images. First the images of control cells and IgG producing cells were analyzed to find the top 10000 SWNT and the cell regions were recorded using the ROI matlab function. These locations were then used to determine the coverage of the SWNT in the cell regions (specific) and outside the cell regions (un-specific) using the CellAnalysis Function (below).

The time response of the IgG producing cells was determined again using the top 10,000 SWNT pixels and plotting their absolute intensity distributions over time for the series of images collected using the CellTime function (below).



Finally, the AnalyzeCellPic function was used to plot the average intensity of cell islands as a contour plot as presented in Figure 6 of the main manuscript.

```
function CellAnalysis
% Coded by Nigel F. Reuel on 9.19.2012
% This program does the basic comparison analysis of our control and IgG
% producing cells
%
NFiles = (1:21);
PSpec = zeros(1:21,1);
PNSpec = zeros(1:21,1);
for i = NFiles
    CellLoc = csvread(['K',int2str(i),'_CellLoc.csv']);
    SWNTLoc = csvread(['SLoc_Rank10000_',int2str(i),'.csv']);
    CellCount = sum(sum(CellLoc));
    OutCellCount = 320*256 - CellCount;
    SWNTinCell = 0;
    SWNToutCell = 0;
    for j = 1:320
        for k = 1:256
            if SWNTLoc(j,k) > 0 && CellLoc(j,k) > 0;
                SWNTinCell = SWNTinCell + 1;
            elseif SWNTLoc(j,k) > 0 && CellLoc(j,k) == 0;
                SWNToutCell = SWNToutCell + 1;
            end
        end
    end
    PSpec(i,1) = SWNTinCell/CellCount*100; % Percent Coverage
    PNSpec(i,1) = SWNToutCell/OutCellCount*100; % Percent Coverage
end
X = (1:21)';
plot(X,PSpec, X,PNSpec)
Data = zeros(4,1); % 1 = IgG Cell specific, 2 = IgG Cell non specific, 3 =
Control Cell Specific, 4 = Control Cell non specific
Error = zeros(4,1);
Data(1,1) = mean(PSpec([13:16 18:21],1));
Data(2,1) = mean(PNSpec([13:16 18:21],1));
Data(3,1) = mean(PSpec(1:7,1));
Data(4,1) = mean(PNSpec(1:7,1));
Error(1,1) = std(PSpec([13:16 18:21],1));
Error(2,1) = std(PNSpec([13:16 18:21],1));
Error(3,1) = std(PSpec(1:7,1));
Error(4,1) = std(PNSpec(1:7,1));
bar(Data)
hold on
errorbar(Data,Error,'or')
hold off
ylabel('% Coverage')

end

function CellTime
% Coded by Nigel Reuel on 9.19.2012
% This code analyzes the time traces of the cells
%
```

```

PL_0 = 0;
PL_1 = 0;
PL_2 = 0;
PL_3 = 0;
    count1 = 1;
    count2 = 1;
    count3 = 1;
    count4 = 1;
for i = 1:19
    PLData = imread(['HS',int2str(i),'.tif']);
    SWNTLoc = csvread(['SLoc_Rank1000_',int2str(i),'.csv']);

    for j = 1:320
        for k = 1:256
            if SWNTLoc(j,k) > 0
                if i < 3                % Time Zero
                    PL_0(count1,1) = PLData(j,k);
                    count1 = count1 + 1;
                elseif i <10            % Time 1-1.5 hr
                    PL_1(count2,1) = PLData(j,k);
                    count2 = count2 + 1;
                elseif i <18            % Time 2-2.5 hr
                    PL_2(count3,1) = PLData(j,k);
                    count3 = count3 + 1;
                else                    % Time 3-3.5 hr
                    PL_3(count4,1) = PLData(j,k);
                    count4 = count4 + 1;
                end
            end
        end
    end
end
% Construct histograms for these time points
X = linspace(5000,30000,1000);
%
subplot(1,4,1)
n = hist(PL_0,X);
bar(X,n)
xlim([7500 15000])
%
subplot(1,4,2)
n = hist(PL_1,X);
bar(X,n)
xlim([7500 15000])
%
subplot(1,4,3)
n = hist(PL_2,X);
bar(X,n)
xlim([7500 15000])
%
subplot(1,4,4)
n = hist(PL_3,X);
bar(X,n)
xlim([7500 15000])

```

```
function AnalyzeCellPic
```

```

% Coded by Nigel F. Reuel on 9.12.2012
% This code will analyze a single cell picture (*.tif)
% For Figure 6 of biomanufacturing paper
%
%}

% Plot the Data together!!
NC = 10;
% Data file number
FN = 3;
%
CData = zeros(320,256);
PLData = imread(['S',int2str(FN),'.tif'],1);
Cell = imread(['C',int2str(FN),'.tif'],1);
for i = 1:NC
    Ccount = 0;
    Stotal = 0;
    Cloc = csvread(['C',int2str(FN),'_CellIsland',int2str(i),'Loc.csv']);

    for j = 1:320
        for k = 1:256
            if Cloc(j,k) > 0
                Ccount = Ccount + 1;
                Stotal = Stotal + PLData(j,k);
            end
        end
    end
    AvgSig = Stotal/Ccount;
    for j = 1:320
        for k = 1:256
            if Cloc(j,k) > 0
                CData(j,k) = AvgSig;
            end
        end
    end
end
contourf(CData)
set(gca, 'Clim', [0 8000])
colormap(hot)
m = 10;
imagesc(Cell)
colormap(gray)
m = 10;
imagesc(PLData)
m = 10;
end

```



## 5. Portable Detection of Carbon Nanotube Optical Sensors

### 5.1 Background and Motivation

Several emergent nanosensor platforms are capable of sensitive and selective detection of analytes without prior labeling, including fluorescent single walled carbon nanotubes<sup>1</sup> for glucose<sup>2,3</sup>, protein<sup>4</sup>, glycoprotein<sup>5</sup>, neurotransmitters<sup>6</sup>, and nitric oxide<sup>7</sup>. Recent work has shown the broad ability to tune the selectivity of these types of sensors based on the rational design of the absorbed polymer phase (Corona Phase Molecular Recognition (CoPhMoRe)<sup>8</sup>). However, challenges in methodology and instrumentation to produce broadly applicable research tools have remained unaddressed. Herein, we describe a novel hydrogel preparation and detection methodology coupled with low cost optics and integrated data processing that addresses this specific challenge.

Single walled carbon nanotubes (SWNT) have clear advantages for label-free, sensitive detection due to their unique photophysical properties, fluorescent stability, and single molecule sensitivity<sup>1</sup>. However, as with other emergent nanosensor platforms<sup>9</sup>, there is a need for novel methodologies that allow for widespread tool development. The constraints of size and capital intensive photo-excitation sources, optics, and detectors represent one challenge. Others include environmental sensitivity to temperature, humidity, vibration, contamination and electromagnetic interference. In this chapter, we describe a novel methodology, demonstrated using examples from protein and neurotransmitter detection, using fluorescent SWNT sensors in the context of a stand-alone, field-portable instrumentation platform with low total cost below \$3000, to highlight applicability.

### 5.2 Methods

#### 5.2.1 Fabrication of Gel Modules for Portable Sensing of SWNT Sensors

A movie of this fabrication process can be viewed [here](#).<sup>a</sup> Attach a batch (30-60) of ½” glass slides (Warner Instrument) to a PDMS holder and lightly scratch the centers with a diamond tip scribe to help the hydrogel adhere to the surface. Melt 2mg/ml (0.2 wt%) of agarose in water, try not to boil (agarose solution will become clear). Let cool to ~40°C and cast 4ul gel spots on the center of the chips. Let gels cure in a humid environment for 30 min. Place 10 ul of desired SWNT sensor solution on each gel. Place in 37°C incubator for 10 min. Remove from incubator and let cool for 10 min. Wash off

---

<sup>a</sup> <http://www.youtube.com/watch?v=6bhBKGUdXY>

excess SWNT sensor solution and place sensor gels in holding buffer or perform extra functionalization chemistry.

### **5.2.2 Functionalization of Chitosan Wrapped SWNT for IgG Sensing**

The chemistry to functionalize chitosan wrapped nanotubes for protein sensing is similar to before<sup>10</sup> except for more efficient coupling of succinic anhydride with a high capacity buffer and the pre-chelation of copper to NTA before conjugation to the SWNT sensor. First deprotonate the chitosan wrapped SWNT in the sensor gels by soaking in 0.01M NaOH. Wash with water and add 5mg/ml of succinic anhydride (Sigma) in 500 mM PBS, pH 8.0. Shake gently overnight. Wash with water. Activate carboxylic acid with 20mg/ml and 60mg/ml of EDC and NHS respectively (Sigma) in MES buffer, pH 4.5 (Pierce). Let sit for 2 hours (bubbles will be visible). Carefully remove the fluid as the chips will tend to float due to bubble formation and wash with water three times. Pre-chelate the copper to the NTA-N (Sigma) as presented in the literature<sup>11</sup> and react with the NHS esters in a buffer of pH 7-8 (half-life of NHS ester is too short above pH 8.5 to achieve good conversion). Wash in PBS or experimental buffer immediately before experiment.

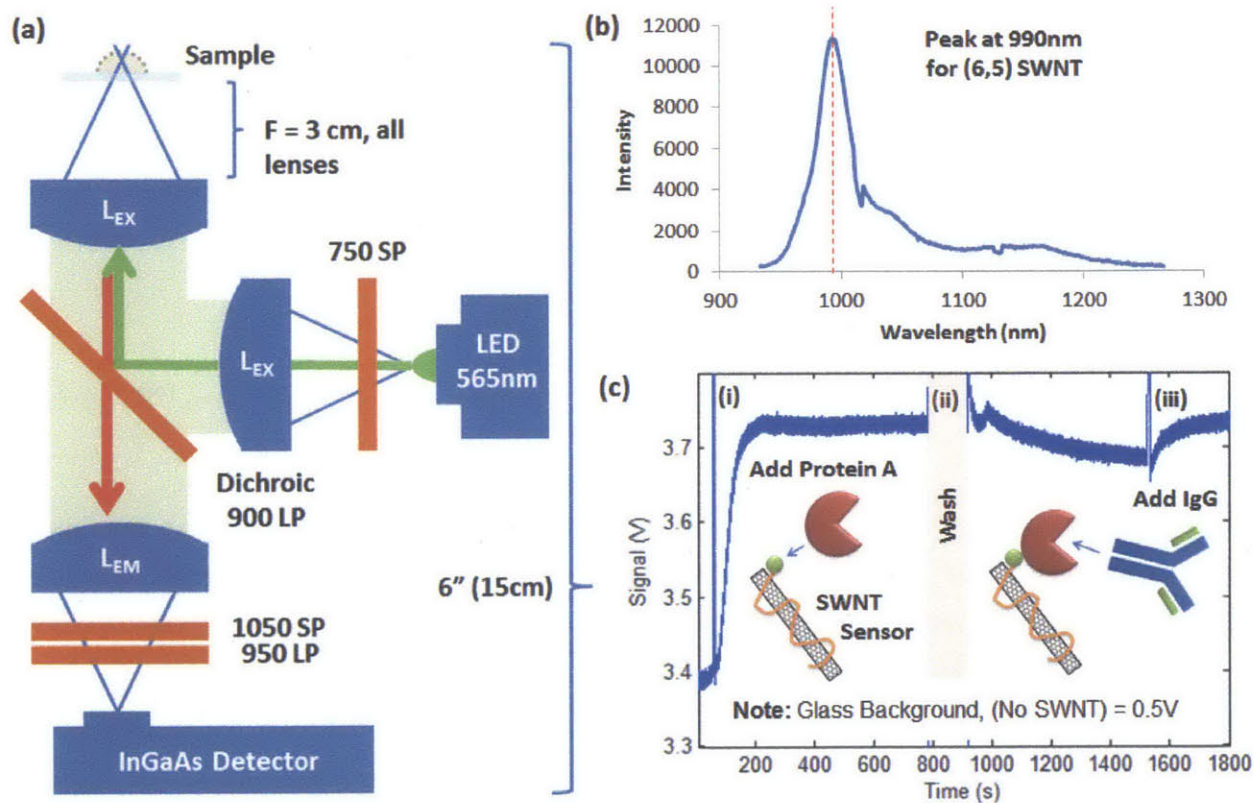
### **5.2.3 Preparation of (GT)15-SWCNT for dopamine detection**

First, 1 mg of as-synthesized single walled carbon nanotubes (SWeNT SG65i, CoMoCAT-process, SouthWest NanoTechnologies, USA) was placed in an 1.5 ml centrifugation tube. Then 2 mg of (GT)15-ssDNA (IDT, USA) was dissolved in 1 ml of 0.1 M NaCl solution and added to the carbon nanotubes. The carbon nanotubes were suspended by direct probe-tip sonication (Cole Parmer, 40 % amplitude, 3 mm tip) for 10 min in an ice bath. Afterwards, the sample was centrifuged twice (16000g) for 90 min to remove aggregates and collect the supernatant. Finally, samples were analyzed by UV-Vis spectroscopy to confirm spectral features of carbon nanotubes and determine a concentration.

## **5.3 Results**

Previous data presented on these SWNT based optical sensors<sup>5,8</sup> were collected with custom, near infrared optical microscope systems; they included a floated optics stage, liquid nitrogen cooled InGaAs detector, and a laser for excitation. The total capital cost (>\$250k) and size (> 1m<sup>3</sup>) limits broad applicability as a scientific tool. The detection platform presented here (Fig 5.1a) has a similar optics pathway to an inverted microscope but is designed for SWNT detection in a more compact footprint (< 0.015m<sup>3</sup>) and the cost of instrumentation is lower (<\$3k see Table 5.1). The platform presented in this communication has been optimized for the (6,5) SWNT chirality (specific nanotube wrapping pattern)

with a high-power LED at 565nm to excite on resonance<sup>12</sup> and filters to collect the 990nm emission peak (Fig 5.1b). Raw data from the detector (Fig 5.1c) reveals the strong signal to background (S/B > 6) and low electronic noise ( $\sigma = 0.003$  volts) such that even low modulating SWNT sensors (like antibody detection<sup>10</sup>) have clearly transduced events. A detector was found that could electronically amplify the SWNT signal without the constraint of cooling and that would not be damaged upon over saturation, thus enabling normal, counter-top operation in ambient lighting.



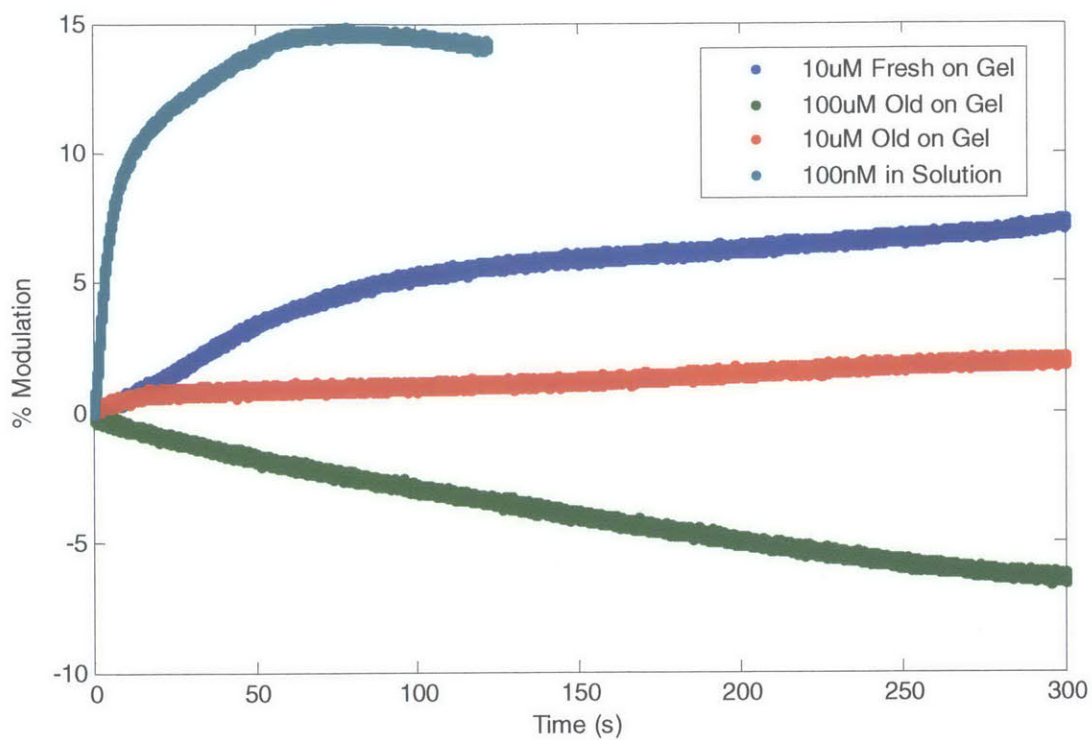
**Figure 5.1** – Platform for portable detection of nanotubes. a) Optical pathway and components for the <\$3k detection system for nanotube sensors (SP = short pass filter, LP = long pass filter, L<sub>EX</sub> = excitation lens, L<sub>EM</sub> = emission lens, green line = excitation pathway, red line = emission pathway). b) Emission spectrum from chitosan-wrapped SWNT excited by 200mW 561nm laser (Opto Engine) with five second integration time (sample was diluted 10x from 30mg/L) revealing the 990nm emission peak. c) Raw response from portable detector for antibody detection showing loading of Histag Protein A (i), wash followed by partial unloading of Protein A (ii), and IgG loading (iii).

**Table 5.1** – Cost of goods analysis for portable detector parts

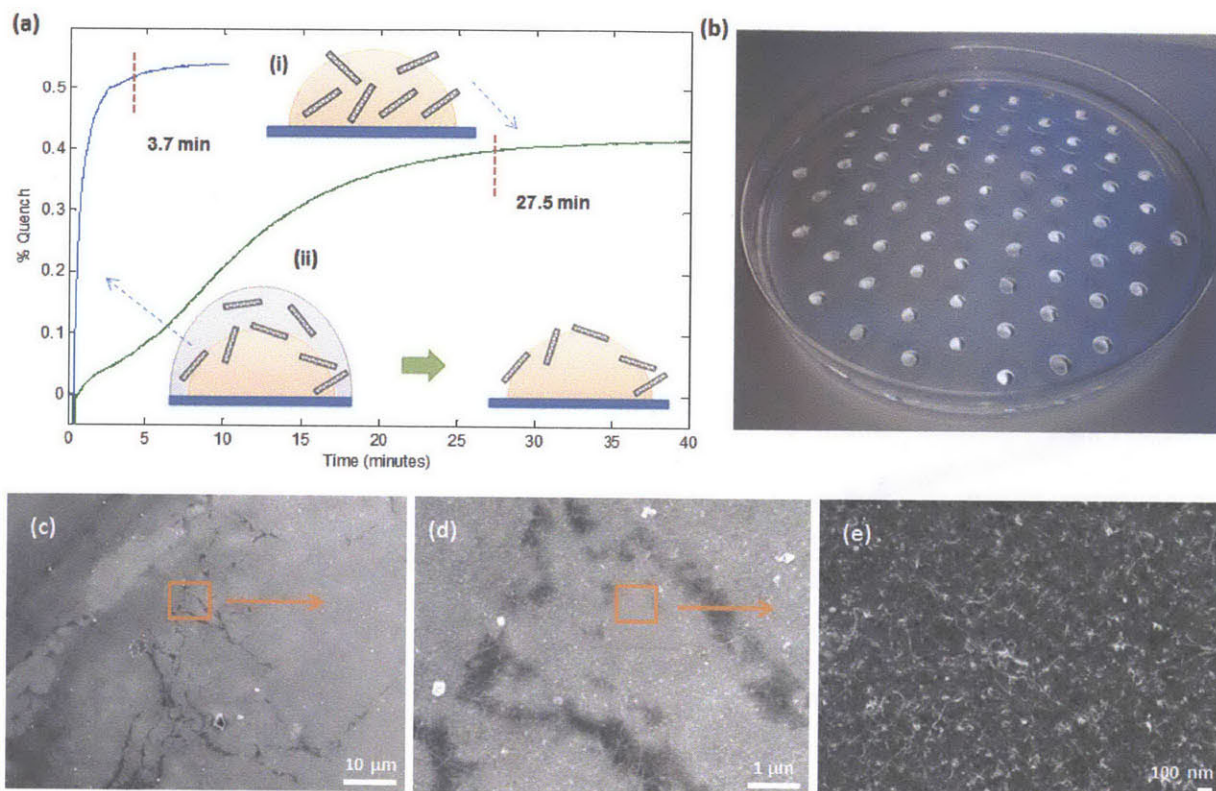
Shorthand Name	Catalog Number and Vendor	Unit Price
Excitation lenses	MAP103030-A 1:1 Matched Achr. Pair, f1=30 mm, f2=30 mm, BBAR 400-700 nm (Thorlabs)	\$168
Dichroic	DMLP900R (Thorlabs)	\$360
LED	M565L2 (Thorlabs)	\$455
Excitation Filter	FES0750 (Thorlabs)	\$73
Dichroic Cage	CM1-DCH (Thorlabs)	\$150
Emission lens	MAP103030-B 1:1 Matched Achr. Pair, f1=30 mm, f2=30 mm, BBAR 650-1050 nm (Thorlabs) - Use one	\$84
Emission Filters Tube	NT 66-230 (LPx2), NT 64-330 (SPx1) from Edmund Optics	\$460
Framework	1/3"x1" (3x) + 1/2"x1" (2x) + 1"x1" (2x) + 1/3"x1/2" (1x) all Thorlabs [SM1L03, SM1L05, SM1L10, SM05L03]	\$104
Detector	PDF10C (Thorlabs)	\$790
DAQ	U6 from LabJack	\$299
Housing + Base	User made (cardboard/case/base to secure detector on)	\$50
<b>Total =</b>		<b>\$2,993</b>

A new method for immobilizing SWNT sensors was also developed to decrease response time with the portable detection platform, thus enabling higher throughput. The portable detector minimizes the effects of vibrational noise by interrogating the sensors in a larger focal volume ( $\sim 0.125\text{mm}^3$  created by two achromatic doublets Fig 5.1a). For sensors immobilized in hydrogels, this can result in longer diffusion times of the analyte to reach all of the illuminated sensors (see difference in dopamine detection with SWNT sensors in gel vs. solution Fig 5.2), especially in the case of large molecular weight analytes such as proteins. The response time decreases  $>7x$  when the sensors are placed in the top surface (casting the hydrogel and then reheating to  $37^\circ\text{C}$  in the presence of SWNT for surface adsorption) *versus* the sensors being placed throughout the hydrogel (mixed in the melted agarose before cooling) as shown by response curves to a protein tagged with quenching moieties (BSA-pyrene Fig 5.3a). The fabrication steps have also been simplified (Methods) to enable scalable and automated production of large gel batches (Fig 5.3b) with a mesh of SWNT sensors absorbed to the top surface (Fig 5.3c-e).





**Figure 5.2** – Response of dopamine selective SWNT sensors embedded in agarose hydrogel for portable detection *versus* solution on the singleplex portable detector. The detector also shows the long-term, reversed response of dopamine sensing noted before<sup>6</sup> – as the analyte sits in solution for 15 min and polymerizes the response begins to degrade and eventually response goes from positive modulation to quenching (blue to green trace).



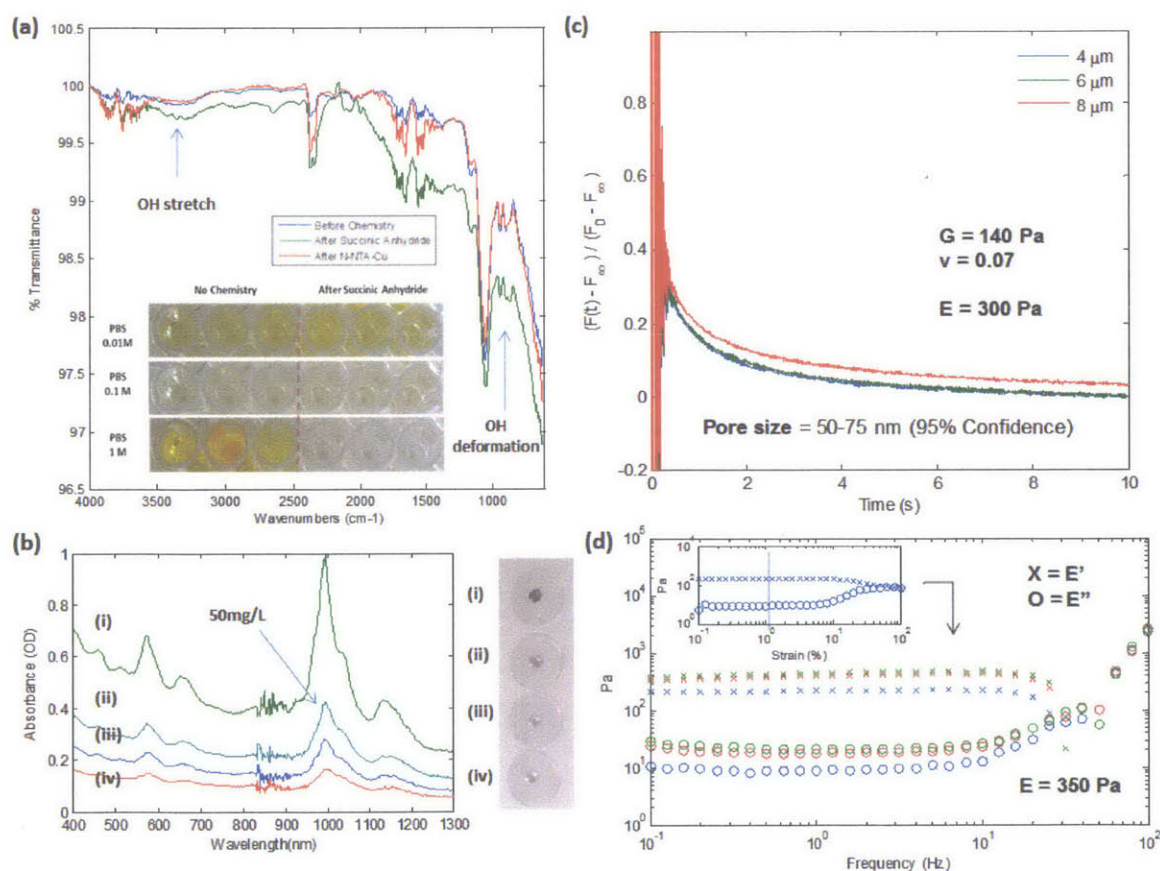
**Figure 5.3** – Agarose hydrogels optimized for portable sensing and automated fabrication. a) Gels can be cast with the nanotubes fully dispersed throughout the gel (i) or absorbed on the top surface (ii). The quenching response time to BSA-pyrene is improved 7x by absorption to the top surface. b) Large batches of gels can be fabricated in ambient conditions with the new protocol. c-e) SEM images reveal the SWNT network in the top surface of the gel.

The resulting hydrogel sensors for portable detection were characterized to ensure consistency in functionalization and composition. In the case of protein detection, following our originally developed laboratory protocol<sup>5,10</sup>, the amines of the chitosan-wrapped nanotubes are conjugated to a chelating group for a divalent cation that acts as both the signal quenching moiety and docking site for a His-tagged binding protein (antibody, lectin, or receptor protein). Both colorimetric assay (TNBSA)<sup>13</sup> and FTIR<sup>11</sup> can be used to track the completion of the amine conjugation to succinic anhydride (introducing COOH groups Fig 5.4a). We found that a high capacity buffer is necessary for efficient completion of these steps (online methods). While screening sensor gels for protein detection made from 15 different SNWT suspensions, an optimal SWNT concentration was found (50mg/ml) for incubation in the top gel surface (Fig 5.4b). Below this concentration (Fig 5.4b iii,iv) the SWNT emission intensity fell below the detection limit of the portable detector and above this concentration (Fig 5.4b i) the signal saturated, decreasing the magnitude of response (also a denser nanotube network on the gel surface inhibits protein entry). We find that reproducible hydrogel synthesis requires precise measurement of the mechanical

properties, from which one can infer the cross linking density and all other constructive properties. Pore size was measured with poroelastic relaxation indentation (PRI<sup>10,14,15</sup>) and was found to be 50-75nm, enough for large analytes like hydrated IgG molecules.<sup>10</sup> PRI also yields the shear modulus (G) and Poisson ratio ( $\nu$ ) which can be used to estimate Young's modulus (E – Eq 5.1). The consistency of the gel composition batch to batch was measured using a stress-controlled rheometer (Fig 5.4d), yielding the storage modulus (E') and loss modulus (E''), which also can be used to measure Young's modulus (Eq 5.2). Both techniques yield a similar measure of the compliant material (300 and 350 Pa respectively).

$$E = 2G(1 + \nu) \quad \text{Eq (5.1)}$$

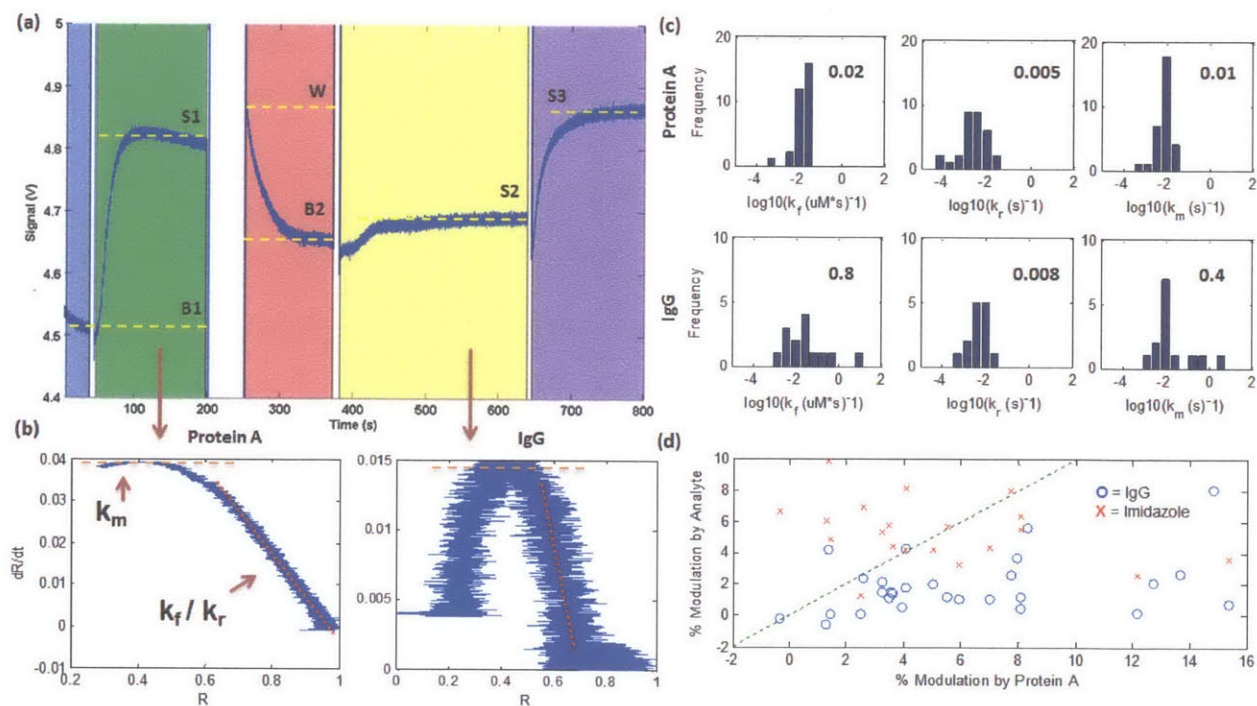
$$E = (E'^2 + E''^2)^{\frac{1}{2}} \quad \text{Eq (5.2)}$$



**Figure 5.4** – Characterization of the sensor gels. a) FTIR of the gels reveal carboxylic acid features after the successful completion of succinic anhydride and removal of these features once conjugated to NTA-Cu<sup>2+</sup>. Colorimetric assay (TNBSA – insert) tracking the amine content also shows the importance of high capacity buffer when coupling succinic anhydride to amine. b) Gels created with varying concentrations of suspended chitosan wrapped SWNT, optimum concentration found to be ~ 50mg/ml by OD. c) Poroelastic indentation with AFM to

measure the pore size, shear modulus ( $G$ ), poisson ratio ( $\nu$ ), and Young's modulus ( $E$ ). d) Consistency of gels confirmed with stress controlled rheometer which also gives a measure of Young's modulus.

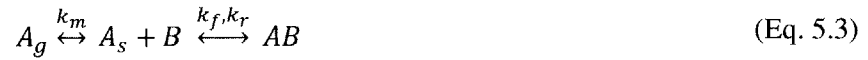
The performance of the portable detector was first evaluated with antibody sensing. Our new methodology presented in this work generalizes and makes portable the approach using his-tagged Protein A<sup>10</sup> for detection of IgG. Note that the response curves (Fig 5.5a) resemble 'sensograms' obtained from surface plasmon resonance<sup>16</sup> (SPR - Biacore®) and biolayer interferometry (BLI - Forte Bio Octet®)<sup>17</sup> as well as the ensemble curves previously reported from our larger detector setups for SWNT sensors<sup>10</sup>. Our new methodology deconvolves the response into six discreet events: baseline, sensor protein loading, wash, unloading of sensor protein, analyte loading, and small molecule control (Fig 5.5a). The small molecule control in this case was the addition of imidazole which complexes with all divalent cations ( $\text{Cu}^{2+}$ ) that may not be accessible to the larger sensor protein.<sup>10</sup> By analyzing the slope ( $dR/dt$ ) of the Protein A loading curve and IgG loading curve (Fig 5.5b), one can determine the forward and reverse kinetic rate constants ( $k_f$ ,  $k_r$ ) and a lumped mass transfer coefficient ( $k_m$ ) using a bimolecular surface absorption model (next page).



**Figure 5.5** – Portable detector measurements of IgG and kinetic characterization. a) Representative response curve ('sensogram') from portable detector showing six recorded events (colored frames): baseline, His-tag Protein A loading, wash, unloading of Protein A, IgG loading, and Imidazole control. b) Plots of  $d\text{Response}/dt$  vs Response used to determine  $k_f$ ,  $k_r$ , and  $k_m$  of Protein A loading and IgG. c) Kinetic rates and mass transfer coefficients for Protein A and IgG are moderately conserved over many different batches of gels (averages included in the plots). d)

Extent of Protein A modulation ((S1-B1)/B1 in Fig 5.5a) is needed for good IgG modulation ((S2-B2)/B2) as denoted by IgG points (o) below the dotted line, however the percent modulation varies (all experiments here at 100ug/ml of IgG). Conversely imidazole modulation is independent of Protein A modulation, denoted by (x) points.

The kinetic model is as follows: assume an analyte in the gel phase ( $A_g$ ) travels to the SWNT surface subject to a lumped mass transfer coefficient ( $k_m$  – assume same in and out of gel). The analyte on the SWNT surface ( $A_s$ ) then complexes with binding sites (B) subject to a forward and reverse kinetic rate ( $k_f, k_r$  – see Eq. 5.3). The rate of analyte-binding site (AB) formation (Eq. 5.4) is the response rate (R). The number of binding sites is conserved (Eq. 5.5). The flux of species in and out of the sensor surface is conserved (Eq. 5.6). Solving for  $A_s$  using Eq. 5.5 and 5.6 yields Eq. 5.7. Similar assumptions to SPR kinetic analysis are made (Eq. 5.8) namely bulk concentration [C] does not change and the response [R] is proportional to the bound species [AB]. Solve for  $k_m$  where adsorption is dominated by mass transfer, the horizontal regime of  $dR/dt$  vs. R (Eq. 5.9). Solve for  $k_f$  and  $k_r$  where adsorption is dominated by ligand kinetics, linear regime of  $dR/dt$  vs. R (Eq. 5.10).



$$\frac{dAB}{dt} = k_f[A_s][B] - k_r[AB] \quad (\text{Eq. 5.4})$$

$$[B] = [AB_{max}] - [AB] \quad (\text{Eq. 5.5})$$

$$k_m[A_g] - k_m[A_s] = k_f[A_s][B] - k_r[AB] \quad (\text{Eq. 5.6})$$

$$[A_s] = \frac{k_m[A_g] + k_r[AB]}{k_m + k_f([AB_{max}] - [AB])} \quad (\text{Eq. 5.7})$$

$$[A_g] = C, [AB] = R, [AB_{max}] = R_{max} \quad (\text{Eq. 5.8})$$

$$\text{if } k_m \ll k_f[B] \text{ then } \frac{dR}{dt} = k_m[C] \quad (\text{Eq. 5.9})$$

$$\text{if } k_m \gg k_f[B] \text{ then } \frac{dR}{dt} = k_f[C](R_{max} - [R]) - k_r[R] \quad (\text{Eq. 5.10})$$

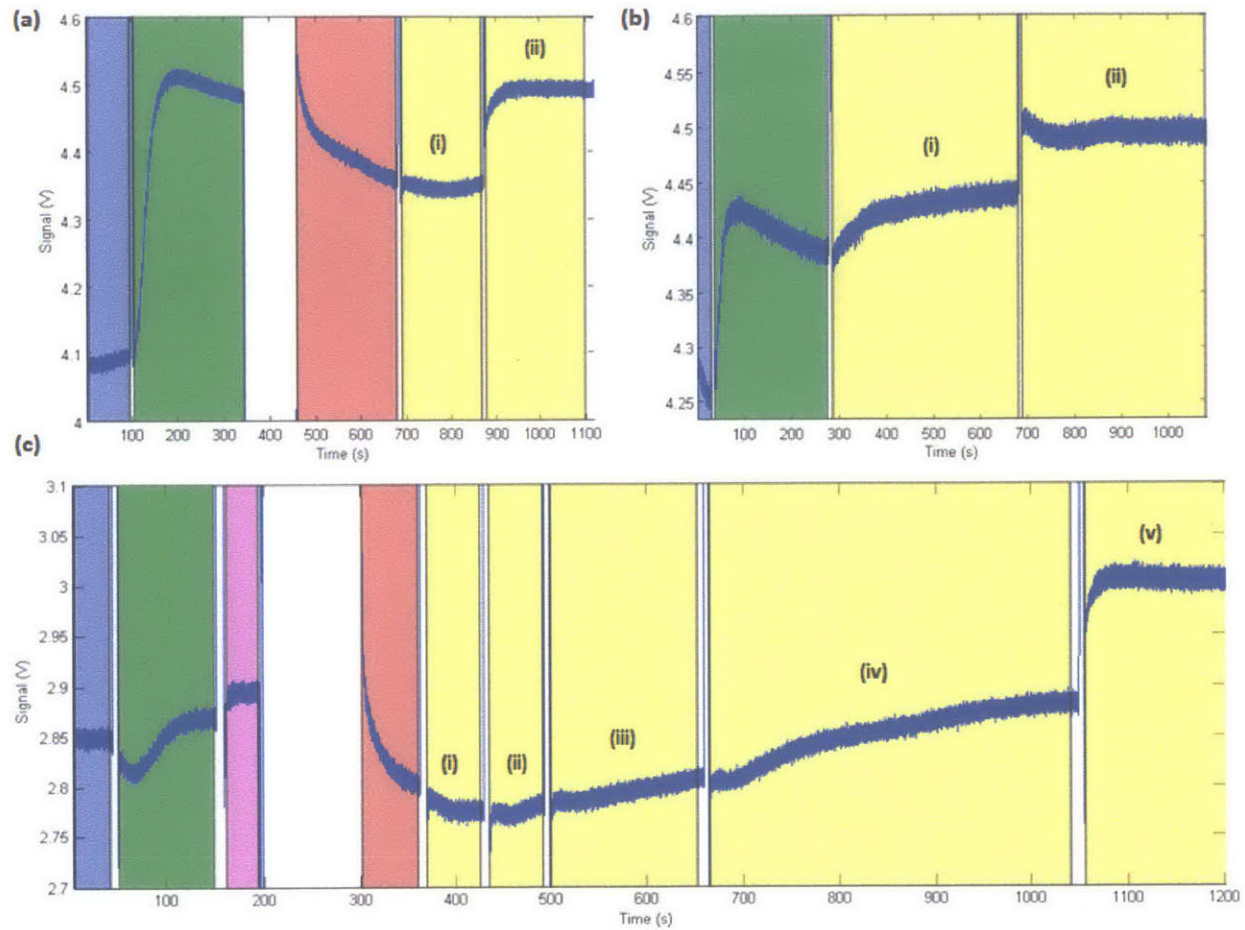
After examining hundreds of such traces over many different fabrication batches, we see encouraging consistency in the kinetic rate constants (Fig 5.5c) and the average rates predict a  $K_D$  of 250 and 10 nM for Protein A histag to  $\text{Cu}^{2+}$  and IgG to Protein A. If one assumes the diffusion of a large protein through a membrane is approximately  $10^{-10} \text{ cm}^2/\text{s}$  the average mass transfer coefficients give a

length scale (or approximate penetration depth) of 1000 and 160 nm for Protein A and IgG respectively. These parameters are expected as Protein A to IgG should be a strong interaction ( $K_D < \mu\text{M}$ ) and the 150kDa IgG should have a shallower penetration than the 43kDa Protein A.

Such portable instrumentation for the detection of IgG has applications in bio-manufacturing and specifically upstream process development, production control, and product quality testing<sup>10</sup>. The methodology and instrumentation can be utilized to perform simple, binary tests (determining presence/absence) very efficiently. This was demonstrated with many different types of commercial IgG supernatants in our Cambridge, MA lab (Fig 5.6) and on site in Basel Switzerland (Fig 5.7 and movie [here](#)<sup>b</sup>). Binary tests can also be run without a wash step if the analyte concentration is greater than the sensor protein (Fig 5.6b). Obtaining a robust, universal calibration curve for all fabricated batches of gels to determine analyte concentration from response is difficult due to the variance in response extent (Fig 5.5d). All sensor gels that load well with protein A respond to the IgG analyte but to varying modulation percentages (points below dotted line in Fig 5.5d). Conversely, the imidazole control loads independent of the success of the Protein A modulation (points above and below the line), but also with varying degrees of 'turn on' response. Automated gel fabrication and structure-function studies are currently being pursued to optimize gel response and make them more consistent than the current hand-fabricated gels. Another approach to calibration for concentration readings could be using the response of a single gel. The slope of the Protein A loading curve can be compared to the slope of the analyte additions to give an approximate read of concentration (Fig 5.6c). We note that such a universal calibration does not exist for many widely utilized methods such as ELISA, SPR, and BLI measurements which require explicit calibration as a part of the method.

---

<sup>b</sup> <http://www.youtube.com/watch?v=jbzZ4A1Lyn0>



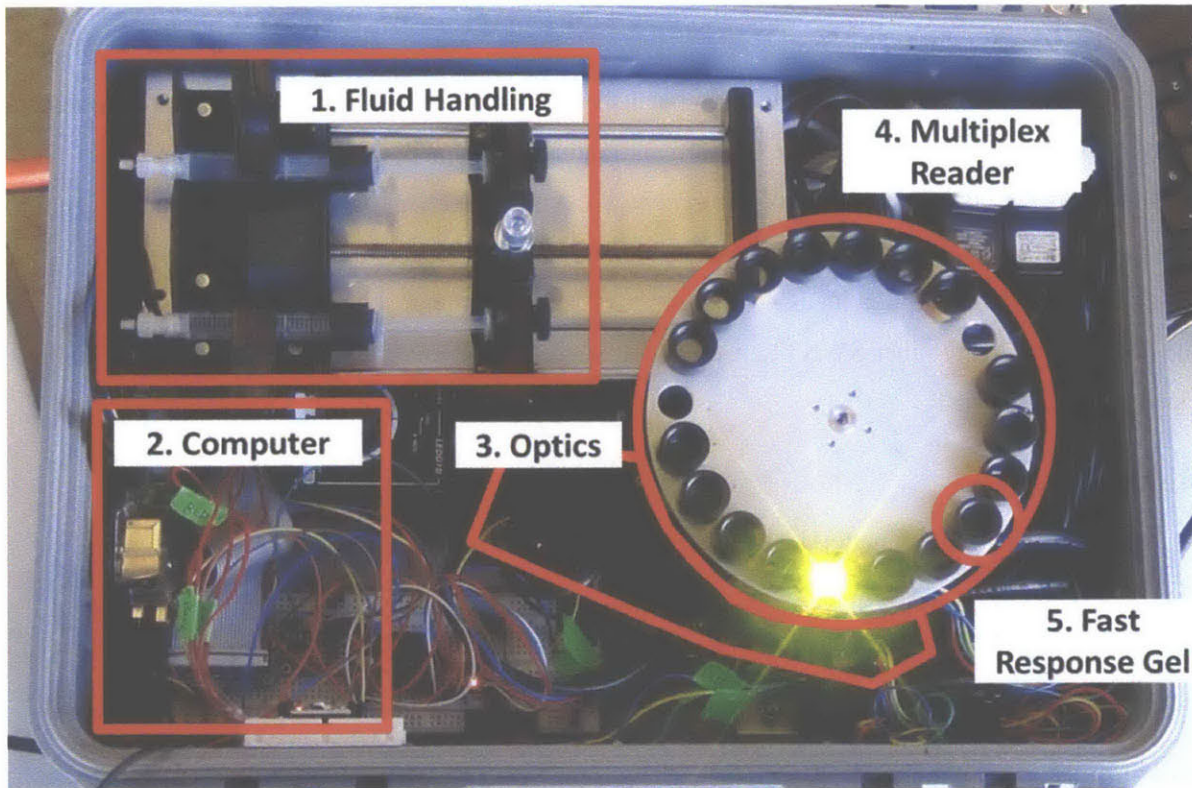
**Figure 5.6** – Other informative IgG detection traces. a) Example of clear binary sensing. After protein A loading, a control commercial supernatant with only host cell proteins gives no response (i) whereas the supernatant from the same strain producing IgG has a clear loading (ii). Raw supernatant was diluted at 10,000x in PBS. b) Example of a no wash experiment, where the Protein A (i) is loaded (at 100ug/ml) and followed by a higher concentration of IgG (ii) (at 230 ug/ml). c) Concentration dependence can be observed within a single trace, protein A loading followed by 10, 100, 1000, and 10000 ng/ml (i-iv) and imidazole control (v) at 15mM. The sensor starts responding in the 100ng/ml range.



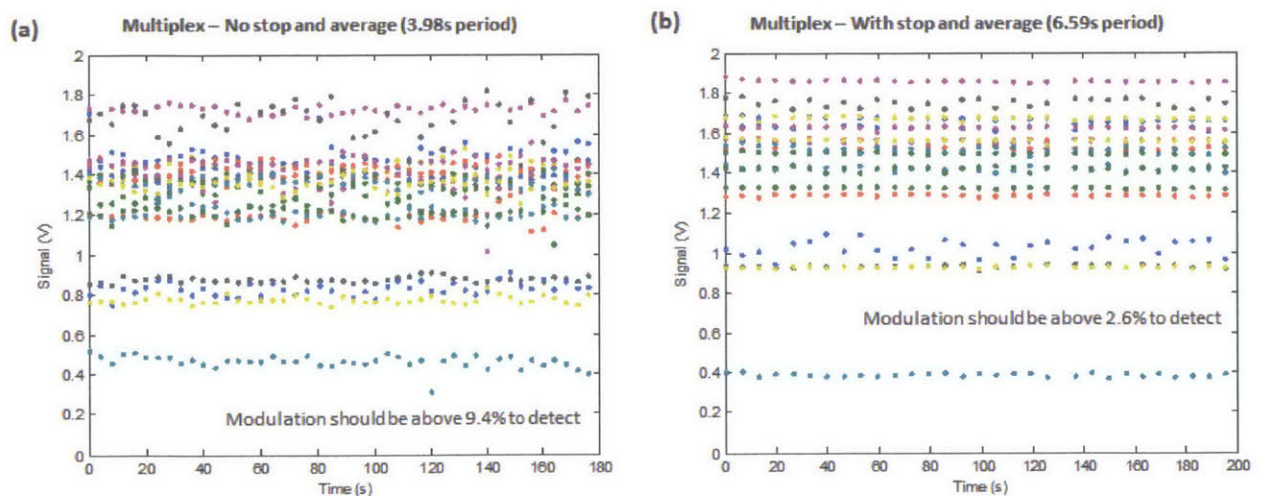
**Figure 5.7** – Demonstrating field portability of detector by transporting sensor gels and two prototypes from Cambridge, MA lab to Novartis Campus in Basel, Switzerland. Demonstrated detection of IgG type molecules (fusion proteins, IgG in raw extract, IgG in clarified extract) in three different labs.

One great advantage to these label-free optical sensors versus other techniques (SPR, BLI) is their ease of multiplexing. The response is emitted from the SWNT sensor so there is no need of physical coupling or waveguide insertion to transduce the signal. To demonstrate the multiplex capabilities we installed a single portable detector in a hand-held briefcase with samples held on a rotating plate (Fig 5.8). The number of samples is limited by the number of gel spots one can print on the circumference of the rotating disc (up to 100 tested on a 150mm disc). Each revolution yields a data point to all of the loading curves. Obviously there is a trade-off on rotation velocity and signal stability. We found that signal responses of greater than 2.6% modulation could be multiplexed with a rotational period of 6.6 seconds, thus yielding 20-100 loading curves in less than two minutes (Fig 5.9). This would be sufficient to do rapid testing of SWNT sensors that have a high percent modulation, such as dopamine sensors in solution phase<sup>6</sup> (Fig 5.2).





**Figure 5.8** – Multiplexed portable reader in fully-integrated briefcase. Multiplexing is achieved by spinning sensors patterned on the circumference of a rotating holder (20-100 samples). The case also contains integrated fluid handling, on-board control and analysis (Raspberry Pi), optics (Fig 5.1a).



**Figure 5.9** – Trade-off in multiplex rotation speed and signal stability. a) At high speeds (<5s period) the stepper motor is less efficient in positioning the gel module at the same position during each revolution, as shown by traces of the 20 SWNT gels. b) At lower speeds (>6.5s) per period the repositioning is more efficient and modulation events greater than 2.6% can easily be detected (based on 3x the standard deviation of the noise introduced by rotation).

## 5.4 Conclusions

Optical SWNT based sensors have increasingly been shown as a promising platform for label-free sensing with applications in defense<sup>18</sup>, healthcare<sup>13</sup>, and manufacturing<sup>10</sup>. One dominant hurdle to their implementation is the large capital cost of equipment to excite, collect, and gather their response signal. Herein we are the first to demonstrate a truly portable and inexpensive method for optical SWNT sensor detection. We have demonstrated its use with antibody<sup>10</sup> and dopamine<sup>6</sup> sensing but it can enable portable detection of many other SWNT sensors already established in the literature and currently being developed by other groups.

## 5.5 References

- 1 Kruss, S. *et al.* Carbon nanotubes as optical biomedical sensors. *Advanced Drug Delivery Reviews* **65**, 1933-1950, doi:<http://dx.doi.org/10.1016/j.addr.2013.07.015> (2013).
- 2 McNicholas, T. P. *et al.* Structure and Function of Glucose Binding Protein-Single Walled Carbon Nanotube Complexes. *Small* **8**, 3510-3516, doi:10.1002/sml.201200649 (2012).
- 3 Yum, K. *et al.* Boronic Acid Library for Selective, Reversible Near-Infrared Fluorescence Quenching of Surfactant Suspended Single-Walled Carbon Nanotubes in Response to Glucose. *Acs Nano* **6**, 819-830, doi:10.1021/nn204323f (2012).
- 4 Ahn, J. H. *et al.* Label-Free, Single Protein Detection on a Near-Infrared Fluorescent Single-Walled Carbon Nanotube/Protein Microarray Fabricated by Cell-Free Synthesis. *Nano Letters* **11**, 2743-2752, doi:10.1021/nl201033d (2011).
- 5 Reuel, N. F. *et al.* Transduction of Glycan-Lectin Binding Using Near-Infrared Fluorescent Single-Walled Carbon Nanotubes for Glycan Profiling. *Journal of the American Chemical Society* **133**, 17923-17933, doi:10.1021/ja2074938 (2011).
- 6 Kruss, S. & al., e. Neurotransmitter Detection Using Corona Phase Molecular Recognition on Fluorescent Single-Walled Carbon Nanotube Sensors. *JACS (in press)* (2013).
- 7 Iverson, N. M. *et al.* In vivo biosensing via tissue-localizable near-infrared-fluorescent single-walled carbon nanotubes. *Nature Nanotechnology* **8**, 873-880, doi:10.1038/nnano.2013.222 (2013).
- 8 Zhang, J. *et al.* Molecular recognition using corona phase complexes made of synthetic polymers adsorbed on carbon nanotubes. *Nat Nano* **8**, 959-968, doi:10.1038/nnano.2013.236 <http://www.nature.com/nnano/journal/v8/n12/abs/nnano.2013.236.html#supplementary-information> (2013).
- 9 Fan, X. D. *et al.* Sensitive optical biosensors for unlabeled targets: A review. *Analytica Chimica Acta* **620**, 8-26, doi:10.1016/j.aca.2008.05.022 (2008).
- 10 Reuel, N. F. *et al.* Emergent Properties of Nanosensor Arrays: Applications for Monitoring IgG Affinity Distributions, Weakly Affined Hypermannosylation, and Colony Selection for Biomanufacturing. *Acs Nano* **7**, 7472-7482, doi:10.1021/nn403215e (2013).
- 11 Wang, L. A., Wei, L., Chen, Y. A. & Jiang, R. R. Specific and reversible immobilization of NADH oxidase on functionalized carbon nanotubes. *Journal of Biotechnology* **150**, 57-63, doi:10.1016/j.jbiotec.2010.07.005 (2010).
- 12 Bachilo, S. M. *et al.* Structure-assigned optical spectra of single-walled carbon nanotubes. *Science* **298**, 2361-2366, doi:10.1126/science.1078727 (2002).
- 13 Zhang, J. *et al.* A Rapid, Direct, Quantitative, and Label-Free Detector of Cardiac Biomarker Troponin T Using Near-Infrared Fluorescent Single-Walled Carbon Nanotube Sensors. *Adv Healthc Mater*, doi:10.1002/adhm.201300033 (2013).

- 14 Chan, E. P., Hu, Y. H., Johnson, P. M., Suo, Z. G. & Stafford, C. M. Spherical indentation testing of poroelastic relaxations in thin hydrogel layers. *Soft Matter* **8**, 1492-1498, doi:10.1039/c1sm06514a (2012).
- 15 Kalcioğlu, Z. I., Mahmoodian, R., Hu, Y. H., Suo, Z. G. & Van Vliet, K. J. From macro- to microscale poroelastic characterization of polymeric hydrogels via indentation. *Soft Matter* **8**, 3393-3398, doi:10.1039/c2sm06825g (2012).
- 16 Karlsson, R., Michaelsson, A. & Mattsson, L. KINETIC-ANALYSIS OF MONOCLONAL ANTIBODY-ANTIGEN INTERACTIONS WITH A NEW BIOSENSOR BASED ANALYTICAL SYSTEM. *Journal of Immunological Methods* **145**, 229-240, doi:10.1016/0022-1759(91)90331-9 (1991).
- 17 Abdiche, Y., Malashock, D., Pinkerton, A. & Pons, J. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Analytical Biochemistry* **377**, 209-217, doi:10.1016/j.ab.2008.03.035 (2008).
- 18 Heller, D. A. *et al.* Peptide secondary structure modulates single-walled carbon nanotube fluorescence as a chaperone sensor for nitroaromatics. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 8544-8549, doi:10.1073/pnas.1005512108 (2011).



## 6. Sensor Signal State Evaluator Program (NoRSE)

Some of the work, text and figures presented in this chapter are reprinted or adapted from “NoRSE: noise reduction and state evaluator for high-frequency single event traces,” by Nigel F. Reuel et al.

Bioinformatics (2012) 28 (2): 296-297. doi: 10.1093/bioinformatics/btr632 (Reproduced under thesis use allowance by Oxford Journals © 2012)

### 6.1 Summary

NoRSE was developed in the course of this thesis work to analyze high-frequency data sets collected from multi-state, dynamic experiments, such as molecular adsorption and desorption onto carbon nanotubes – such as those presented in the single nanotube analysis of Chapter 3. As technology improves sampling frequency, these stochastic data sets become increasingly large with faster dynamic events. More efficient algorithms are needed to accurately locate the unique states in each time trace. NoRSE adapts and optimizes a previously published noise reduction algorithm (Chung et al., 1991) and uses a custom peak flagging routine to rapidly identify unique event states. The algorithm is explained using experimental data from our lab and its fitting accuracy and efficiency are then shown with a generalized model of stochastic data sets. The algorithm is compared to another recently published state finding algorithm and is found to be 27 times faster and more accurate over 55% of the generalized experimental space. NoRSE is written as an M-file for Matlab (Appendix 6.1).

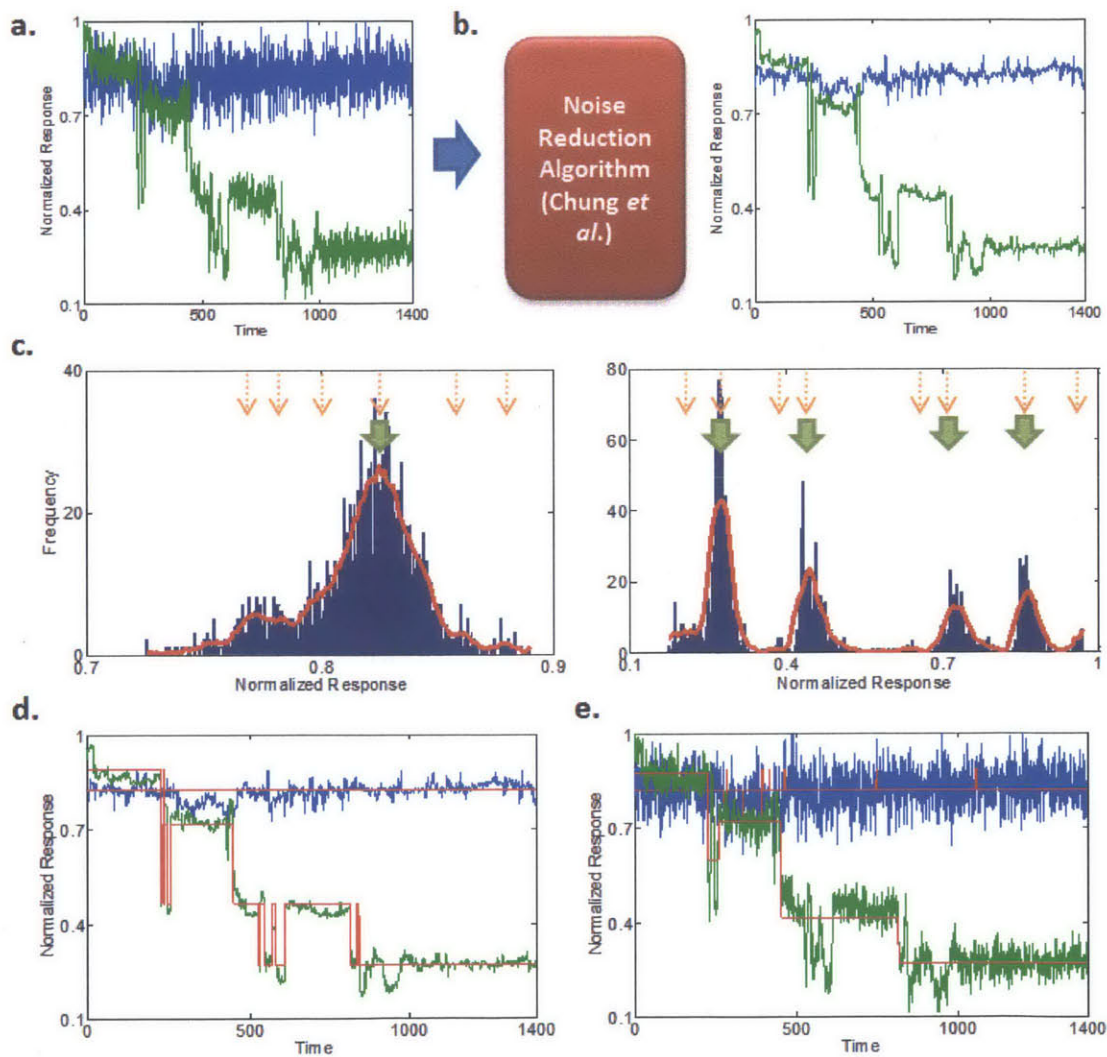
### 6.2 Motivation

As stochastic biological studies are performed at increasingly smaller length and time scales, the analysis of large, noisy data-sets is becoming an increasingly common problem. In our lab, these sets consist of time traces gathered from monitoring the fluorescence of single-walled-carbon nanotube (SWNT)-based sensors. As target molecules bind to a single sensor, the fluorescence is quenched in a step-wise manner (Cognet et al., 2007), resulting in unique fluorescent states that correspond to the number of bound sites along the length of the nanotube (Jin et al., 2008). Although NoRSE was created with our fluorescence data sets in mind, the algorithm can be readily applied to other, analogous biological experiments that contain event states. This is demonstrated by an error analysis of generalized stochastic traces as shown below.

### 6.3 Algorithm Explanation

In this chapter, data from our SWNT-based, protein-protein sensors are analyzed. Excited SWNT fluoresce due to their unique band-gap structure (Bachilo et al., 2002) and can be quenched by specific

molecules or chelating groups (Jin et al., 2008). At a single-molecule level, the quenching is exhibited in a stepwise manner as each of the limited, exciton excursion distances of the SWNT are occupied (Cognet et al., 2007). Our 900 nm sensor exhibits a maximum of ten binding events, or ten step levels. The fluorescence of each SWNT is monitored for 3000 time steps and NoRSE efficiently resolves the bound states. NoRSE imports the data traces  $X_1(t), X_2(t), \dots, X_n(t)$  and normalizes them by the maximum intensity such that  $\{X_i: 0 < X_i < 1\}$  (Fig 6.1a – top trace is a control and the bottom trace exhibits binding events).



**Figure 6.1** – Explanation and evaluation of NoRSE algorithm. A) Control (blue) and signal (green) traces are normalized and (B) noise-reduced by Chung et al. algorithm. C) All point histogram peaks of each trace are identified as unique event states to which the trace is fit (D). E) Corresponding SFA fit of these traces.

$X_i(t)$  is then noise-reduced (Fig 6.1b) by the Chung and Kennedy algorithm (Chung et al., 1991). Their forward-backward, non-linear filtering technique was designed to preserve high-frequency step

events surrounded by background noise. In brief, the algorithm searches forward and backwards from each data point using a bank of window sizes to weigh the probability of a sharp step occurring. Their algorithm contains three parameters ( $p$ - which sets the sharpness of state transitions,  $N$ - which specifies a set of forward and backward windows, and  $M$ - which sets how much data the  $N$  windows are run against) which were optimized in the original work for patch-clamp experiments (Chung et al., 1991) and subsequently for protein folding experiments (Haran 2004). We used Monte Carlo simulations of generic, noisy, high-frequency event traces to further explore this parameter optimization. Generally, we found that an increasing value of  $p$  and  $N$  and a decreasing value of  $M$  improved the noise reduction algorithm's ability to reconstruct the original trace, but the improvement was marginal. Thus the program is set to run with  $p$ ,  $N$ , and  $M$  values of 40, [4 8 16 32], and 10 respectively for all types of experimental traces.

After noise-reduction of  $X_i(t)$ , all points histograms with 200 bins are generated for  $X_i(t)$  (Fig 6.1c). The distinct groups of histogram peaks represent event states but vary in height and width depending on their frequency and uniqueness respectively. A detailed explanation of this peak flagging routine is given below, but briefly all potential peaks are flagged and then logic is used to determine which peaks are unique and which should be combined with neighboring peaks. The result is a vector of significant peak bin numbers, which are then translated back to the corresponding signal levels of  $X_i(t)$ . These are the unique state levels for the trace. Finally, each time point of the noise reduced trace of  $X_i(t)$  is compared to the possible state levels and the one that fits closest is assigned. Qualitatively, the fit looks good for our experimental data (Fig 6.1d-e), but we were interested in a quantitative method of determining the error as well as finding its applicability to other types of data that contain step events.

### 6.3.1 Peak Finding Subroutine

The distinct peaks of the all-point histogram for each signal trace reveal the unique event states in the given trace. Determining the peaks by eye is a simple process, but creating a routine that is general enough for all peak types took some creative code. On the first pass, the algorithm flags each possible peak; that is for each bin in the histogram it evaluates if the bin prior and the bin following have a frequency count lower than the evaluation bin. If they do, this bin is flagged as a possible peak.

The next step is to analyze peaks in a pair-wise fashion. By analyzing the valley distances between two peaks, each peak can be sorted into three categories: 1) a unique and significant peak, 2) an insignificant peak, or 3) significant but a candidate for averaging with the next peak. A significant valley distance is defined as one that is greater than 50 percent of the average frequency response. For example, a trace with 1000 time points, that is binned with 200 bins in an all-point histogram and then smoothed with a running average window of 5, would have an average response of 1 frequency count per bin ( $=1000/200/5$ ). So if the differences in averaged-frequency count from the top of the peak to the bottom

of the valleys (right and left side of the peak) are greater than 50 percent of this average response, it is considered as significant and unique. If both valley distances are smaller than the average response, the peak is labeled as insignificant. If only one valley distance (right-side or left-side) is greater than this cut-off, further logic is used to determine if the peak should be merged with adjacent peaks or ignored.

Peaks are merged if their bins are separated by a small distance. This cut off distance is specified as 2.5 percent of the total bin count (5 bins for a 200 bin all point histogram). The two bins are placed in a temporary vector of candidate for peak averaging and the next peak is evaluated to see if it too should be combined. When a significant valley distance is found, the candidate peaks for merging are averaged and the valley distances of this new combined peak are evaluated. If the frequency count is less than the cut-off specified above, the combined peak is ignored. This works especially well in eliminating supposed peaks that are closely grouped together in a slow-rising side of another significant peak, or as a valley between two other significant peaks. The fidelity of this peak finding algorithm over a wide range of signal types can be assessed in the following general model for error analysis.

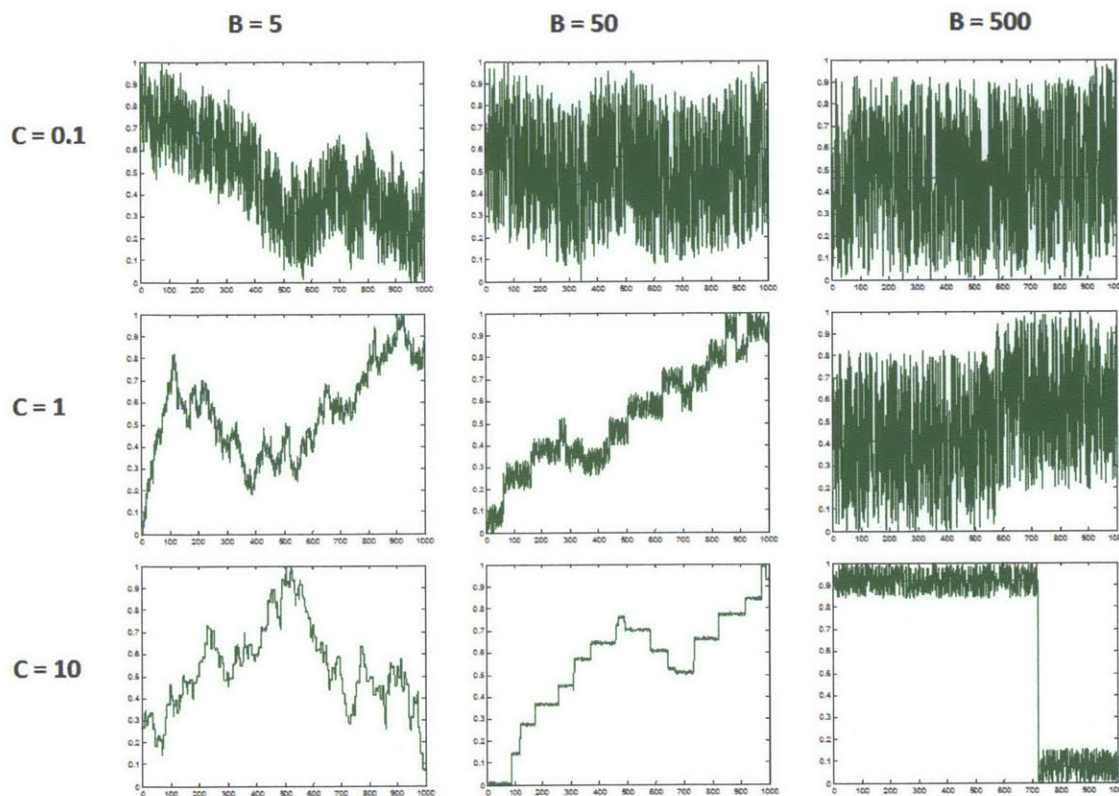
#### **6.4 Gauging Performance of Algorithm with a Generalized Model**

There are many types of data sets that are stochastic in nature and contain imbedded event states. The biological data sets highlighted in this article include protein-protein interactions, protein-folding FRET experiments, and patch-clamp data sets. These are just a few of the many possible data sets. To evaluate the error rate of the NoRSE algorithm on these and other types of data sets, we have created a model with three general parameters that roughly describe most of this experimental space:  $A = (\text{Tendency upwards movement} / \text{Tendency downwards movement})$ ,  $B = (\text{Average event time} / \text{Signal sampling time})$ , and  $C = (\text{Average event transition size} / \text{Average noise level})$ . Parameter 'A' can be thought of as a quasi-dissociation constant. A large value of 'B' represents an experiment in which the sampling time is much smaller than the event time, and thus step events are more easily resolved. Conversely, a small 'B' value is in the regime of fast dynamic experiments where the sampling time is comparatively small. A small value of 'C' represents data sets that are highly obstructed by noise.

Many sample traces were generated (using random, Gaussian distributed values) around the trace parameter specified above. Limits to these generalized parameters, in terms of plausible experimental traces, were set as follows: 0.1 to 10 for A, 5 to 500 for B, and 0.1 to 10 for C. Values deviating from 1 (equilibrium) for parameter A are symmetric (<1 representing traces that move in a downwards fashion, >1 upwards) and has less of an effect on the fitting fidelity than the other two parameters (it will simply increase the number of unique states and increase computation time). Most experimental data sets are interested in the fluctuating, equilibrium state, thus a value of  $A = 1$  was used in this analysis. However,

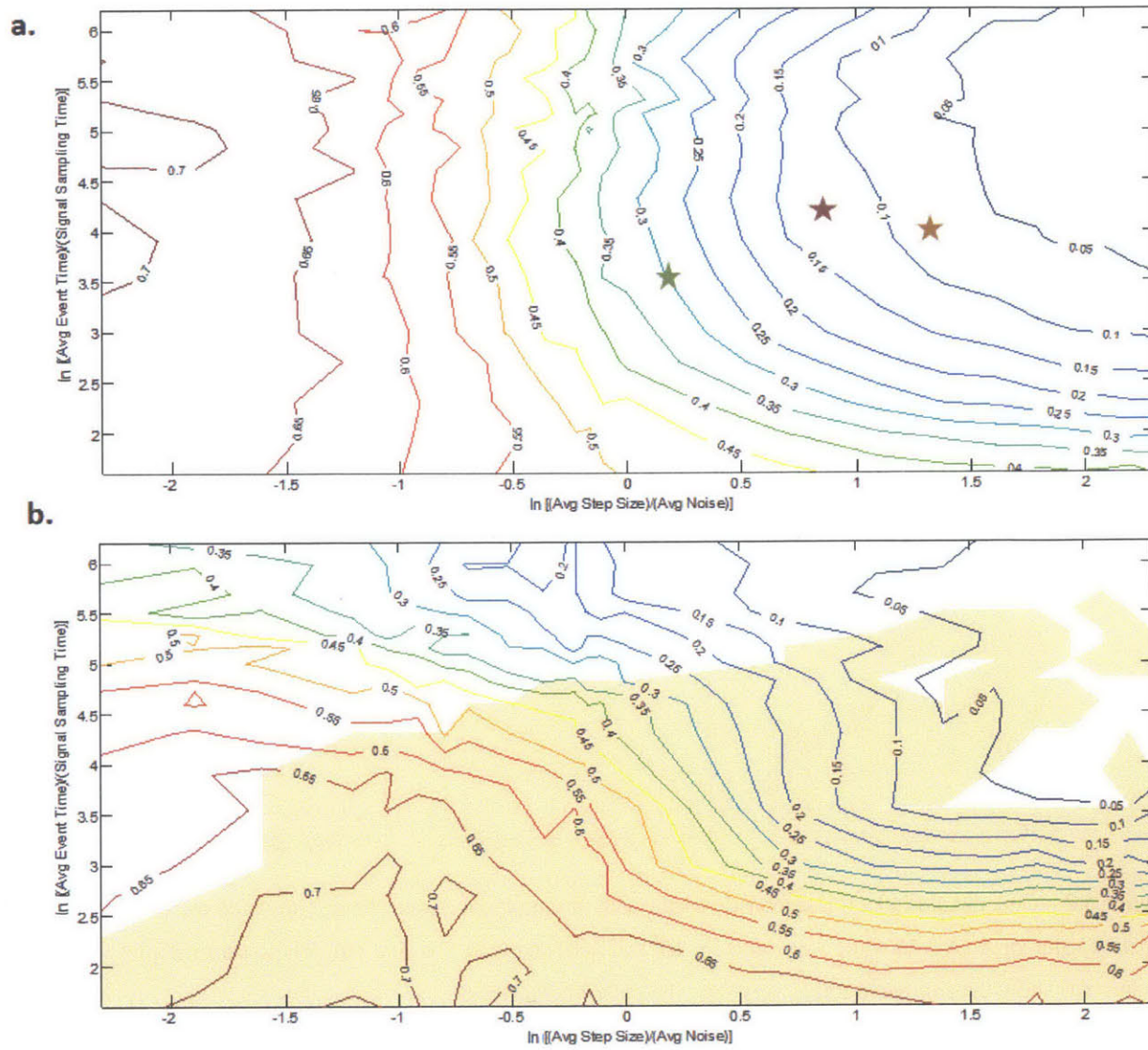


our code for this trace generator can be simply adjusted to vary this parameter as well. Example simulated traces for  $A = 1$ ,  $B = [5 \ 50 \ 500]$ , and  $C = [0.1 \ 1 \ 10]$  are provided (Fig 6.2).



**Figure 6.2** – Generalized model of stochastic traces with unique event states. These simulated traces are at a constant value of  $A$  ( $A = 1$ ) while varying  $B$  (5, 10, and 500) and  $C$  (0.1, 1, and 10). These traces depict qualitatively how the simulated traces vary over the experimental space.

Many types of single-event experiments can be cast on this generalized experimental space (Figure 6.3a). As the parameter ‘ $A$ ’ has less of an effect on the fitting-fidelity, it was held at the equilibrium level of 1, while the other parameters ‘ $B$ ’ and ‘ $C$ ’ were varied over the ranges of 5 to 500 and 0.1 to 10 respectively. For each parameter combination, 100 traces were generated at 1000 time steps each. The noisy trace was sent to NoRSE and the resulting fit was compared to the known generated trace. Each time step was analyzed to see if it was within 2% of the real value, if not, it was marked as an error. Finally the average error percent for each  $B$ - $C$  combination was reported (Fig 6.3a). The contour plot maps the specific regions where the NoRSE algorithm is most accurate; it begins to have errors greater than 10% at  $B$  values  $< 20$  and  $C$  values  $< 2.7$ .



**Figure 6.3** - a) The fit error percentage of NorSE over a general experimental space (Green Star = general region for patch clamp experiments, Red = FRET Protein folding, and Orange = SWNT protein-protein sensors). b) Error percentage of SFA algorithm. Shaded area is the experimental space in which NorSE outperforms SFA.

To compare NorSE’s computational efficiency and accuracy, an identical error analysis was performed on a popular “State-Fitting Algorithm” (SFA) used for similar noisy, dynamic traces (Kerssemakers et al., 2006) (Fig 6.1e). SFA uses a series of Chi-squared best fit analyses as the number of steps (S) is iteratively increased and compared to corresponding “anti-fits.” The resulting error contour map (Fig 6.3b) indicates regions in which NorSE outperforms SFA (shaded region – code provided in Appendix 6.2). The physical run time of NorSE was also 27 times faster. By scaling the algorithms and taking the approximate ratio of SFA to NorSE (see below), we find that NorSE is on the order of  $[S^2/(K(N+3))]$  more efficient than SFA, where S is the number of SFA fit steps and K and N are parameters

of the noise reduction algorithm. The two algorithms were performed on the protein-protein traces and the average run time was 25 times faster for NoRSE (32 predicted from approximate scaling).

### 6.4.1 Scaling comparison of NoRSE and SFA Algorithms

The two algorithms involve many steps which complicate the general scaling estimation. However, by analyzing the loops of each step a general picture of the required number of computations can be obtained. This was done for both the SFA algorithm and the NoRSE code. The ratio of the two was then created to determine how computationally efficient the NoRSE program is in comparison to the existing SFA code.

Variables with common values:

N = 1000	Number of traces
t = 3000	Number of time points
S = 32	Number of attempted fit states
k = 4	Number of comparison windows in noise reduction algorithm
nBin = 200	Number of histogram bins
M = 5	Parameter in noise reduction algorithm

Approximate scaling:

$$SFA = N \cdot t \cdot S + N \cdot S^2 \cdot t + N \cdot S$$

$$NoRSE = 3 \cdot N \cdot t \cdot k + N \cdot t \cdot k \cdot M + 2 \cdot N \cdot t + 2 \cdot N \cdot nBin + N \cdot S$$

Approximate ratio:

$$\frac{SFA}{NoRSE} = \frac{N \cdot t \cdot S + N \cdot S^2 \cdot t + N \cdot S}{3 \cdot N \cdot t \cdot k + N \cdot t \cdot k \cdot M + 2 \cdot N \cdot t + 2 \cdot N \cdot nBin + N \cdot S}$$

Simplified:

$$\frac{SFA}{NoRSE} = \frac{S(t + S \cdot t + 1)}{2 \cdot nBin + S + 2 \cdot t + 3 \cdot k \cdot t + k \cdot m \cdot t}$$

Keeping most significant values:

$$\frac{SFA}{NoRSE} = \frac{S^2}{k \cdot (M + 3)}$$

Using the common values for the parameters stated above, the full approximate ratio yields 30.93 and the simplified scaling ratio yields 32. Thus, for any given set of parameters NoRSE is approximately  $S^2$  more computationally efficient than the established state finding program, SFA (where S is the maximum number of states that each program attempts to fit to the data).

## 6.5 References

Bachilo, S.M. et al. (2002) Structure-assigned optical spectra of single-walled carbon nanotubes. *Science*, 298, 2361-2366.

Chung, S.H. and Kennedy, R.A. (1991) Forward-backward non-linear filtering technique for extracting small biological signals from noise. *J. Neurosci. Meth.*, 40, 71-86.

Cognet, L. et al. (2007) Stepwise Quenching of Exciton Fluorescence in Carbon Nanotubes by Single-Molecule Reactions. *Science*, 316, 1465-1468.

Haran, G. (2004) Noise reduction in single-molecule fluorescence trajectories of folding proteins. *Chem. Phys.* 307, 137-145.

Jin, H. et al. (2008) Stochastic analysis of stepwise fluorescence quenching reactions on single-walled carbon nanotubes: single molecule sensors. *Nano Lett.*, 8, 4299-4304.

Kerssemakers, J.W.J., et al. (2006) Assembly dynamics of microtubules at molecular resolution. *Nature*, 442, 709-712.

## Appendix 6.1

function NoRSE

```
% Coded by Nigel F. Reuel 4.01.2011
% Massachusetts Institute of Technology
% Department of Chemical Engineering
%
% ----- Step 1: Read in the Traces (can normalize here if desired) -----
%
% Data Columns = Trace number , Data Rows = Time steps
%
RTD = csvread('Data.csv');
%Determine the number of time steps (Ns) and number of transducers (Nt):
[Ns,Nt]= size(RTD);
RTD_n = RTD;
% Create empty matrix to receive the noise reduced traces:
M = zeros(Ns,Nt);
%
% ----- Step 2: Denoise each of the Traces -----
```

```

%
%The denoising algorithm loses the first and last time steps, so set a new
%time window:
time = Ns - 2;
%Create an empty matrix to receive the denoised trace data:
T_denoise = zeros(time,Nt);
for i = 1:Nt
    NT = RTD_n(:,i);
    IC = feval(@NoiseReduc,NT);
    T_denoise(:,i) = IC(:,1);
end
% csvwrite(['T_denoise',int2str(iii),'.csv'],T_denoise);
%
%----- Step 3: Construct APH for each trace -----
%
%Number of bins for the histogram construction?
nbins = 200;
%Create empty receiving matrices for histogram counts (nM) and the bin
%centers (xM)
nM = zeros(Nt,nbins);
xM = zeros(Nt,nbins);
for i = 1:Nt
    Y = T_denoise(:,i);
    [n,xout] = hist(Y,nbins);
    nM(i,:) = n(1,:);
    xM(i,:) = xout(1,:);
end
%xlswrite('nM',nM);
%xlswrite('xM',xM);
%
%----- Step 4: Find peaks for each APH histogram -----
%
%Use PeakFinder Function to find the Average Peak Locations (APL) - these
%translate to the bin number
APL = feval(@PeakFinder,nM);
[Nt,Nlev] = size(APL);
%
%Translate the APL bin numbers back to the normalized intensity level to
%determine the stochastic step levels (SLL)
SLL = zeros(Nt,Nlev);
for i = 1:Nt
    % Very rarely, the peak finder will flag no bins as peaks, in this case
    % we will use a general gaussian fit of the denoised data to find a
    % mean step value.
    if sum(APL(i,:)) == 0
        Ncurves = 1;
        X = T_denoise(i,:);
        [u,~,~,~] = feval(@fit_mix_gaussian,X,Ncurves);
        SLL(i,1) = u;
    else
        for j = 1:Nlev

```

```

        if APL(i,j) ~= 0
            N_bin = APL(i,j);
            SLL(i,j) = xM(i,N_bin);
        end
    end
end
end
end
%
%----- Step 5 Determine the NoRSE Fit Trace -----
%
for ii = 1:Nt
    RealT = T_denoise(:,ii);
    count = 0;
    for m = 1:Nlev
        if SLL(ii,m) > 0
            count = count + 1;
        end
    end
    for i = 1:time
        compare = zeros(count,1);
        for j = 1:count
            compare(j,1) = abs(RealT(i,1)-SLL(ii,j));
        end
        [C,I] = min(compare);
        M(i+1,ii,1) = SLL(ii,I);
    end
    % Because the noise reduction program takes away the first and last time
    % point, estimate their values as those immediately following and
    % proceeding respectively
    M(1,ii,1) = M(2,ii,1);
    M(time+2,ii,1) = M(time+1,ii,1);
end
% Record the noise-reduced traces in a csv format:
csvwrite('NRData.csv',M)
end
%
%
%
% <-----Functions embedded in the Trace Analyzer ----->
%
function IC = NoiseReduc(NT)
%
% Coded by Nigel Reuel on 8.31.2010
%
% Adaptation of Chung and Kennedy (1991) forward-backward non-linear
% filtering technique to reduce the noise of our
% fluorescent signal in an attempt to resolve the single molecule events
% amidst the noise.
%
% Input: Takes supplied noisy trace (NT)

```

```

%Output: Returns cleaned trace without the first and last data point
%
time = length(NT);
%
%Set algorithm parameters:
N = [4 8 16 32];    %Sampling window length
K = length(N);      %Number of sampling windows
M = 10;             %Weighting parameter length
P = 40;             %Parameter that effects sharpness of steps
%
% ----- The Forward-Backward non-linear Algorithm -----
%
%First - Run each time step and store the average forward and backward
%predictors for each of the sampling windows
%
I_avg_f = zeros(time,K);
I_avg_b = zeros(time,K);
for g = 1:time
    %Run for each sampling window
    for k = 1:K
        %Solve for the average forward predictor (w/ some
        %logic to help solve the beginning of the trace
        window = N(k);
        if g == 1
            I_avg_f(g,k) = NT(1,1);
        elseif g - window - 1 < 0
            I_avg_f(g,k) = sum(NT(1:g-1,1))/g;
        else
            epoint = g - window;
            spoint = g - 1;
            I_avg_f(g,k) = sum(NT(epoint:spoint,1))/window;
        end
        %Now do the same for the backward predictor
        if g == time
            I_avg_b(g,k) = NT(g,1);
        elseif g + window > time
            sw = time - g;
            I_avg_b(g,k) = sum(NT(g+1:time,1))/sw;
        else
            epoint = g + window;
            spoint = g + 1;
            I_avg_b(g,k) = sum(NT(spoint:epoint,1))/window;
        end
    end
end
%
%Second - Solve for non-normalized forward and backward weights:
f = zeros(time,K);
b = zeros(time,K);
for i = 1:time
    for k = 1:K

```

```

Mstore_f = zeros(M,1);
Mstore_b = zeros(M,1);
for j = 0:M-1
    t_f = i - j;
    t_b = i + j;
    if t_f < 1
        Mstore_f(j+1,1) = (NT(i,1) - I_avg_f(i,k))^2;
    else
        Mstore_f(j+1,1) = (NT(t_f,1) - I_avg_f(t_f,k))^2;
    end
    if t_b > time
        Mstore_b(j+1,1) = (NT(i,1) - I_avg_b(i,k))^2;
    else
        Mstore_b(j+1,1) = (NT(t_b,1) - I_avg_b(t_b,k))^2;
    end
end
f(i,k) = sum(Mstore_f)^(-P);
b(i,k) = sum(Mstore_b)^(-P);
end
end
% Third - Solve for vector of normalization factors for the weights
C = zeros(time,1);
for i = 1:time
    Kstore = zeros(K,1);
    for k = 1:K
        Kstore(k,1) = f(i,k) + b(i,k);
    end
    C(i,1) = 1/sum(Kstore);
end
% Fourth and final step - Put all parameters together to solve for the
% intensities
Iclean = zeros(time,1);
for i = 1:time
    TempSum = zeros(K,1);
    for k = 1:K
        TempSum(k,1) = f(i,k)*C(i,1)*I_avg_f(i,k) + b(i,k)*C(i,1)*I_avg_b(i,k);
    end
    Iclean(i,1) = sum(TempSum);
end
IC = Iclean(2:time-1,1);
return
end

```

```

function PeakLocAvg = PeakFinder(nM)
% Coded by Nigel Reuel on 9.15.2010 updated in March 2011 by NFR
% This function finds the peaks of a histogram using a modified running
% average algorithm + peak flagging routine
%
[nT,nB] = size(nM);
%
%Size of window for running average algorithm (increasing this causes more

```



```

%of a rough approximation of peaks)
SW = 3;
% Centered running average algorithm to help find peaks:
%
T_ra = zeros(nT,nB);
for j = 1:nT
    for i = 1:nB
        if i < (SW + 1)
            T_ra(j,i) = sum(nM(j,1:i+SW))/(i+SW);
        elseif i >= (SW+1) && i <= (nB-SW)
            T_ra(j,i) = sum(nM(j,i-SW:i+SW))/(SW*2+1);
        else
            T_ra(j,i) = sum(nM(j,i:nB))/(nB-i+1);
        end
    end
end
end
%xlswrite('T_ra',T_ra);
%
%-----Peak finding algorithm-----
%
% Base function on physical example of climbing up and down hills. Drop
% flags on supposed peaks. Eliminate "peaks" that are molehills. Combine
% peaks that are close neighbors.
Flags = zeros(nT,nB,6);
% Layer 1 = Peak Flag
% Layer 2 = Right Valley Bin#
% Layer 3 = Left Valley Bin#
% Layer 4 = Peak magnitude (Histogram Count)
% Layer 5 = Right Valley magnitude (Histogram Count)
% Layer 6 = Left Valley magnitude (Histogram Count)
% Molehill level (number of responses that are not significant - avg!)
MoleL = (5/SW)*.5; % = #responses/#bins/SW = Avg. Response, no significance)
PeakLocAvg = zeros(nT,2);
for i = 1:nT
    for j = 1:nB
        if T_ra(i,j) > MoleL
            % Logic to find significant peaks at endpoints (and the R/L
            % valleys)
            if j == 1 && T_ra(i,j) > T_ra(i,j+1)
                Flags(i,j,1) = 1;
                Flags(i,j,4) = T_ra(i,j);
                Pval = T_ra(i,j);
                Pnext = T_ra(i,j+1);
                index = j;
                % Find right hand valley:
                while Pval > Pnext
                    index = index + 1;
                    Pval = T_ra(i,index);
                    Pnext = T_ra(i,index+1);
                end
                Flags(i,j,2) = index;
            end
        end
    end
end

```

```

    Flags(i,j,5) = T_ra(i,index);
elseif j == 1 && T_ra(i,j) <= T_ra(i,j+1)
    Flags(i,j,1) = 0;
elseif j == nB && T_ra(i,j) > T_ra(i,j-1)
    Flags(i,j,1) = 1;
    Flags(i,j,4) = T_ra(i,j);
    Pval = T_ra(i,j);
    Pprior = T_ra(i,j-1);
    index = j;
    % Find left hand valley:
    while Pval > Pprior
        index = index - 1;
        Pval = T_ra(i,index);
        Pprior = T_ra(i,index-1);
    end
    Flags(i,j,3) = index;
    Flags(i,j,6) = T_ra(i,index);
elseif j == nB && T_ra(i,j) <= T_ra(i,j-1)
    Flags(i,j,1) = 0;
    % Logic to find peaks at all midpoints (and their R/L valleys)
elseif T_ra(i,j) >= T_ra(i,j-1) && T_ra(i,j) >= T_ra(i,j+1)
    Flags(i,j,1) = 1;
    Flags(i,j,4) = T_ra(i,j);
    % Find left hand valley:
    Pval = T_ra(i,j);
    Pprior = T_ra(i,j-1);
    index = j;
    if j > 2
        Switch = 1;
    else
        Switch = 0;
        index = 1;
    end
    while Pval > Pprior && Switch == 1;
        index = index - 1;
        Pval = T_ra(i,index);
        Pprior = T_ra(i,index-1);
        if index == 2
            Switch = 0;
            index = 1;
        end
    end
    end
    Flags(i,j,3) = index;
    Flags(i,j,6) = T_ra(i,index);
    % Find right hand valley:
    Pval = T_ra(i,j);
    Pnext = T_ra(i,j+1);
    index = j;
    if j < nB - 1
        Switch = 1;
    else

```



```

VD2 = abs(Flags(i,PBN2,6) - Flags(i,PBN2,4));
%
if VD1 <= SigD && VD2 <= SigD
    % Combine
    PeakCode(1,j) = 2;
    PeakCode(2,j+1) = 2;
elseif VD1 >= SigD && VD2 <= SigD
    % Left Peak sig, right peak ignore
    PeakCode(1,j) = 1;
    PeakCode(2,j+1) = 3;
elseif VD1 <= SigD && VD2 >= SigD
    % Left peak ignore, right peak significant
    PeakCode(1,j) = 3;
    PeakCode(2,j+1) = 1;
elseif VD1 >= SigD && VD2 >= SigD
    % Both peaks are significant
    PeakCode(1,j) = 1;
    PeakCode(2,j+1) = 1;
end
end
% Logic for the peak endpoints
Left = abs(Flags(i,PeakVec(1,1),4) - Flags(i,PeakVec(1,1),6));
if Left >= SigD
    PeakCode(2,1) = 1;
else
    PeakCode(2,1) = 3;
end
Right = abs(Flags(i,PeakVec(1,Npeaks),4) - Flags(i,PeakVec(1,Npeaks),5));
if Right >= SigD
    PeakCode(1,Npeaks) = 1;
else
    PeakCode(1,Npeaks) = 3;
end
% Now determine what to do with the peaks. Look at combinations
% reading left to right. Put significant peaks into a vector.
SPV = 0;
j = 1;
count = 1;
while j <= Npeaks
    C1 = PeakCode(1,j);
    C2 = PeakCode(2,j);
    if C1 == 1 && C2 == 1
        % This is a bonafide unique peak. Put it in the vector.
        SPV(1,count) = PeakVec(1,j);
        count = count + 1;
        j = j+1;
    elseif C1 == 2
        % This peak will be merged with other peak(s)
        count2 = 2;
        Merge = PeakVec(1,j);
        while C1 == 2

```

```

    j = j + 1;
    C1 = PeakCode(1,j);
    Merge(1,count2) = PeakVec(1,j);
    count2 = count2 + 1;
end
% Insert logic here to eliminate valleys and slow rises (see
% what is happening at the significant endpoints!)
End1 = Flags(i,Merge(1,1),4) - Flags(i,Merge(1,1),6);
End2 = Flags(i,Merge(1,end),4) - Flags(i,Merge(1,end),5);
Distance = Merge(1,end) - Merge(1,1);
if End1 >= SigD && End2 >= SigD
    SPV(1,count) = round(mean(Merge));
    count = count + 1;
    j = j + 1;
elseif End1 >= SigD && End2 <= (SigD)*(-1) && Distance >= SigBinD
    SPV(1,count) = round(mean(Merge));
    count = count + 1;
    j = j + 1;
elseif End1 <= SigD*(-1) && End2 >= (SigD) && Distance >= SigBinD
    SPV(1,count) = round(mean(Merge));
    count = count + 1;
    j = j + 1;
end
else % Represents 1/3, 3/1, and 3/3 combos which are all ignored!!
    j = j + 1;
end
end
Nsigpeaks = length(SPV);
for j = 1:Nsigpeaks
    PeakLocAvg(i,j) = SPV(1,j);
end
%{
% Plot the PeakLocAvg to see how the algorithm performs)
S = Nsigpeaks;
FF = zeros(1,nB);
for k = 1:S
    Temp = PeakLocAvg(i,k);
    FF(1,Temp) = 1;
end
X = (1:200)';
Y1 = FF(1,:)';
Y2 = T_ra(i,:);
plot(X,Y1,X,Y2)
stop = 1;
%}
end

return
end

```

```

function [u,sig,t,iter] = fit_mix_gaussian( X,M )
%
% Gaussian Fit Function by Ohad Gal (c) 2003
% Download at:
% http://www.mathworks.com/matlabcentral/fileexchange/4222-a-collection-of-fitting-functions
% on 3.21.2011
%
% fit_mix_gaussian - fit parameters for a mixed-gaussian distribution using EM algorithm
%
% format: [u,sig,t,iter] = fit_mix_gaussian( X,M )
%
% input:  X - input samples, Nx1 vector
%         M - number of gaussians which are assumed to compose the distribution
%
% output: u - fitted mean for each gaussian
%         sig - fitted standard deviation for each gaussian
%         t - probability of each gaussian in the complete distribution
%         iter- number of iterations done by the function
%
% initialize and initial guesses
N      = length( X );
Z      = ones(N,M) * 1/M;          % indicators vector
P      = zeros(N,M);              % probabilities vector for each sample and each model
t      = ones(1,M) * 1/M;        % distribution of the gaussian models in the samples
u      = linspace(min(X),max(X),M); % mean vector
sig2   = ones(1,M) * var(X) / sqrt(M); % variance vector
C      = 1/sqrt(2*pi);           % just a constant
Ic     = ones(N,1);              % - enable a row replication by the * operator
Ir     = ones(1,M);              % - enable a column replication by the * operator
Q      = zeros(N,M);             % user variable to determine when we have converged to a steady
solution
thresh = 1e-3;
step   = N;
last_step = inf;
iter   = 0;
min_iter = 10;

% main convergence loop, assume gaussians are 1D
while ( (( abs((step/last_step)-1) > thresh) & (step>(N*eps)) ) | (iter<min_iter) )

    % E step
    % =====
    Q = Z;
    P = C ./ (Ic*sqrt(sig2)) .* exp( -(X*Ir - Ic*u).^2 ./ (2*Ic*sig2) );
    for m = 1:M
        Z(:,m) = (P(:,m)*t(m))./(P*t(:));
    end

    % estimate convergence step size and update iteration number
    prog_text = sprintf(repmat( '\b',1,(iter>0)*12+ceil(log10(iter+1)) ));
    iter      = iter + 1;
end

```

```

last_step = step * (1 + eps) + eps;
step      = sum(sum(abs(Q-Z)));
fprintf( '%s%d iterations\n',prog_text,iter );

% M step
% =====
Zm        = sum(Z);          % sum each column
Zm(find(Zm==0)) = eps;      % avoid division by zero
u         = (X')*Z ./ Zm;
sig2      = sum(((X*Ir - Ic*u).^2).*Z) ./ Zm;
t         = Zm/N;
end

sig  = sqrt( sig2 );

return
end

```

## Appendix 6.2

```

function NoRSE_SFA_Comparison
% Finished by Nigel F. Reuel on 3.31.2011
% Massachusetts Institute of Technology
% Department of Chemical Engineering
%
% This code uses the generalized model of stochastic data sets to generate
% traces with embeded event states. It is currently set to compare NoRSE
% with the SFA algorithm. For each parameter set (A,B,C) it does the
% following:
%
% 1) Generates 100 traces
% 2) Fits traces with NoRSE
% 3) Fits traces with SFA
% 4) Computes average error for each algorithm
% 5) Reports the average times to run the 100 traces through NoRSE and SFA
%
% The errors and run times are then recorded in a *.csv file for further
% analysis.
%
% The program goes as follows:
%
% 1) Specify the trace generator parameters
%
A = 1;
B = [5 7 10 15 20 35 50 75 100 125 150 175 200 250 300 400 500];
C = [0.1 .15 .2 .3 .35 .4 .45 .5 .6 .7 .8 .85 .9 1 2 3 4 5 6 7 8 9 10];

% Number of traces:
nt = 100;
% Number of steps:

```

```

ns = 1000;
% Answer matrices:
NorseQual = zeros(length(B),length(C));
FSQual = zeros(length(B),length(C));
TimeN = zeros(length(B),length(C));
TimeFS = zeros(length(B),length(C));
%
for i = 1:length(B)
    for j = 1:length(C)
        b = B(i);
        c = C(j);
        disp(['Analyzer started for condition B = ',num2str(b),' and C = ',num2str(c)])
        % 2) Generate Traces
        M = feval(@GenTrace,A,b,c,nt,ns);
        % 3) Generate Fits
        [M2,T] = feval(@TraceFitter,M(:,2));
        % 4) Compare Fits
        % First you must specify the tolerance of fit:
        TF = 0.02; %<--- Remember this is on a scale of 0 to 1
        % Reset the count of bad NORSE points and bad FS points:
        BadN = 0;
        BadFS = 0;
        for k = 1:nt
            for l = 1:ns
                if abs(M2(1,k,1)-M(1,k,1))> TF
                    BadN = BadN + 1;
                end
                if abs(M2(1,k,2)-M(1,k,1))> TF
                    BadFS = BadFS + 1;
                end
            end
        end
        % Now compute the percent error:
        %disp('End of analysis. Run time and results below.')
        TotalSteps = nt*ns;
        PerBN = BadN/TotalSteps;
        PerBFS = BadFS/TotalSteps;
        NorseQual(i,j) = PerBN;
        FSQual(i,j) = PerBFS;
        TimeN(i,j) = T(1,1);
        TimeFS(i,j) = T(1,2);
    end
end
% These are the output files:

csvwrite('BadFitPercentN.csv',NorseQual)
csvwrite('BadFitPercentFS.csv',FSQual)
csvwrite('TimeN.csv',TimeN)
csvwrite('TimeFS.csv',TimeFS)

```



```

return
end

function [M,T] = TraceFitter(RTD)
% This function fits the generated traces with both the NoRSE program and
% the fit states program.
% ----- Step 1: Read in and the Traces -----
%
%Determine the number of time steps (Ns) and number of transducers (Nt):
[Ns,Nt]= size(RTD);
RTD_n = RTD;
%
M = zeros(Ns,Nt,4); % First level Norse fit, second level FS
T = [0 0]; % 1 = NoRSE Time, 2 = FS time

%##### NoRSE ALGORITHM #####
%
% ----- Step 2: Denoise each of the Traces -----
%
tic
%The denoising algorithm loses the first and last time steps, so set a new
%time window:
time = Ns - 2;
%Create an empty matrix to recieve the denoised trace data:
T_denoise = zeros(time,Nt);
for i = 1:Nt
    NT = RTD_n(:,i);
    IC = feval(@NoiseReduc,NT);
    T_denoise(:,i) = IC(:,1);
end
% csvwrite(['T_denoise',int2str(iii),'.csv'],T_denoise);
%
%----- Step 3: Construct APH for each trace -----
%
%Number of bins for the histogram construction?
nbins = 200;
%Create empty recieving matrices for histogram counts (nM) and the bin
%centers (xM)
nM = zeros(Nt,nbins);
xM = zeros(Nt,nbins);
for i = 1:Nt
    Y = T_denoise(:,i);
    [n,xout] = hist(Y,nbins);
    nM(i,:) = n(1,:);
    xM(i,:) = xout(1,:);
end
%xlswrite('nM',nM);
%xlswrite('xM',xM);
%
%----- Step 4: Find peaks for each APH histogram -----
%
```

```

%Use PeakFinder Function to find the Average Peak Locations (APL) - these
%translate to the bin number
APL = feval(@PeakFinder,nM);
[Nt,Nlev] = size(APL);
%
%Translate the APL bin numbers back to the normalized intensity level to
%determine the stochastic step levels (SLL)
SLL = zeros(Nt,Nlev);
for i = 1:Nt
    % Very rarely, the peak finder will flag no bins as peaks, in this case
    % we will use a general gaussian fit of the denoised data to find a
    % mean step value.
    if sum(APL(i,:)) == 0
        Ncurves = 1;
        X = T_denoise(i,:);
        [u,~,~,~] = feval(@fit_mix_gaussian,X,Ncurves);
        SLL(i,1) = u;
    else
        for j = 1:Nlev
            if APL(i,j) ~= 0
                N_bin = APL(i,j);
                SLL(i,j) = xM(i,N_bin);
            end
        end
    end
end
end
%
%----- Step 5 Determine the NoRSE Fit Trace -----
%
for ii = 1:Nt
    RealT = T_denoise(:,ii);
    count = 0;
    for m = 1:Nlev
        if SLL(ii,m) > 0
            count = count + 1;
        end
    end
    for i = 1:time
        compare = zeros(count,1);
        for j = 1:count
            compare(j,1) = abs(RealT(i,1)-SLL(ii,j));
        end
        [C,I] = min(compare);
        M(i+1,ii,1) = SLL(ii,I);
    end
    % Because the noise reduction program takes away the first and last time
    % point, estimate their values as those immediately following and
    % proceeding respectively
    M(1,ii,1) = M(2,ii,1);
    M(time+2,ii,1) = M(time+1,ii,1);
end
end

```

```

%
T(1,1) = toc;
%
%
% #####
%----- FitStates Analyzer ----- (Added 1.15.2011 by NFR)
% #####
%
%
tic
forward_trans = 0;
NonNormTraces = RTD;
[num_rows, num_cols] = size(NonNormTraces);
Ftime = 1:1:num_rows;
% Make empty matrix to receive the traces and their fits
count = 0;
GoodTrace2 = 0;
for i = 1:num_cols
    Tracemax = max(NonNormTraces(:,i));
    Tracemin = min(NonNormTraces(:,i));
    data3 = 1000*(NonNormTraces(:,i) - Tracemin)/(Tracemax - Tracemin);
    data2 = 1000 - data3;
    MaxNumofStates = 30;
    maxN = 30;
    [NumofStates, BestFitTraces] = fitStates(maxN, Ftime, data2, data3, MaxNumofStates);
    M(:,i,2) = BestFitTraces; % Best fit trace from Fit States
end
T(1,2) = toc;
return
end
%
%
% <-----Functions embeded in the Trace Analyzer ----->
%
function IC = NoiseReduc(NT)
%
% Coded by Nigel Reuel on 8.31.2010
%
% Adaptation of Chung and Kennedy (1991) forward-backward non-linear
% filtering technique to reduce the noise of our
% fluorescent signal in an attempt to resolve the single molecule events
% amidst the noise.
%
% Input: Takes supplied noisy trace (NT)
% Output: Returns cleaned trace without the first and last data point
%
time = length(NT);
%
% Set algorithm parameters:
N = [4 8 16 32]; %Sampling window length
K = length(N); %Number of sampling windows

```

```

M = 10;           %Weighting parameter length
P = 40;          %Parameter that effects sharpness of steps
%
% ----- The Forward-Backward non-linear Algorithm -----
%
%First - Run each time step and store the average forward and backward
%predictors for each of the sampling windows
%
I_avg_f = zeros(time,K);
I_avg_b = zeros(time,K);
for g = 1:time
    %Run for each sampling window
    for k = 1:K
        %Solve for the average forward predictor (w/ some
        %logic to help solve the beginning of the trace
        window = N(k);
        if g == 1
            I_avg_f(g,k) = NT(1,1);
        elseif g - window - 1 < 0
            I_avg_f(g,k) = sum(NT(1:g-1,1))/g;
        else
            epoint = g - window;
            spoint = g - 1;
            I_avg_f(g,k) = sum(NT(epoint:spoint,1))/window;
        end
        %Now do the same for the backward predictor
        if g == time
            I_avg_b(g,k) = NT(g,1);
        elseif g + window > time
            sw = time - g;
            I_avg_b(g,k) = sum(NT(g+1:time,1))/sw;
        else
            epoint = g + window;
            spoint = g + 1;
            I_avg_b(g,k) = sum(NT(spoint:epoint,1))/window;
        end
    end
end
end
%
%Second - Solve for non-normalized forward and backward weights:
f = zeros(time,K);
b = zeros(time,K);
for i = 1:time
    for k = 1:K
        Mstore_f = zeros(M,1);
        Mstore_b = zeros(M,1);
        for j = 0:M-1
            t_f = i - j;
            t_b = i + j;
            if t_f < 1
                Mstore_f(j+1,1) = (NT(i,1) - I_avg_f(i,k))^2;
            end
        end
    end
end

```

```

else
    Mstore_f(j+1,1) = (NT(t_f,1) - I_avg_f(t_f,k))^2;
end
if t_b > time
    Mstore_b(j+1,1) = (NT(i,1) - I_avg_b(i,k))^2;
else
    Mstore_b(j+1,1) = (NT(t_b,1) - I_avg_b(t_b,k))^2;
end
end
f(i,k) = sum(Mstore_f)^(-P);
b(i,k) = sum(Mstore_b)^(-P);
end
end
% Third - Solve for vector of normalization factors for the weights
C = zeros(time,1);
for i = 1:time
    Kstore = zeros(K,1);
    for k = 1:K
        Kstore(k,1) = f(i,k) + b(i,k);
    end
    C(i,1) = 1/sum(Kstore);
end
% Fourth and final step - Put all parameters together to solve for the
% intensities
Iclean = zeros(time,1);
for i = 1:time
    TempSum = zeros(K,1);
    for k = 1:K
        TempSum(k,1) = f(i,k)*C(i,1)*I_avg_f(i,k) + b(i,k)*C(i,1)*I_avg_b(i,k);
    end
    Iclean(i,1) = sum(TempSum);
end
IC = Iclean(2:time-1,1);
return
end

```

```

function PeakLocAvg = PeakFinder(nM)
% Coded by Nigel Reuel on 9.15.2010 updated in March 2011 by NFR
% This function finds the peaks of a histogram using a modified running
% average algorithm + peak flagging routine
%
[nT,nB] = size(nM);
%
%Size of window for running average algorithm (increasing this causes more
%of a rough approximation of peaks)
SW = 3;
% Centered running average algorithm to help find peaks:
%
T_ra = zeros(nT,nB);
for j = 1:nT
    for i = 1:nB

```

```

    if i < (SW + 1)
        T_ra(j,i) = sum(nM(j,1:i+SW))/(i+SW);
    elseif i >= (SW+1) && i <= (nB-SW)
        T_ra(j,i) = sum(nM(j,i-SW:i+SW))/(SW*2+1);
    else
        T_ra(j,i) = sum(nM(j,i:nB))/(nB-i+1);
    end
end
end
%xlswrite('T_ra',T_ra);
%
%-----Peak finding algorithm-----
%
% Base function on physical example of climbing up and down hills. Drop
% flags on supposed peaks. Eliminate "peaks" that are molehills. Combine
% peaks that are close neighbors.
Flags = zeros(nT,nB,6);
% Layer 1 = Peak Flag
% Layer 2 = Right Valley Bin#
% Layer 3 = Left Valley Bin#
% Layer 4 = Peak magnitude (Histogram Count)
% Layer 5 = Right Valley magnitude (Histogram Count)
% Layer 6 = Left Valley magnitude (Histogram Count)
% Molehill level (number of responses that are not significant - avg!)
MoleL = (5/SW)*.5; % = #responses/#bins/SW = Avg. Response, no significance)
PeakLocAvg = zeros(nT,2);
for i = 1:nT
    for j = 1:nB
        if T_ra(i,j) > MoleL
            % Logic to find significant peaks at endpoints (and the R/L
            % valleys)
            if j == 1 && T_ra(i,j) > T_ra(i,j+1)
                Flags(i,j,1) = 1;
                Flags(i,j,4) = T_ra(i,j);
                Pval = T_ra(i,j);
                Pnext = T_ra(i,j+1);
                index = j;
                % Find right hand valley:
                while Pval > Pnext
                    index = index + 1;
                    Pval = T_ra(i,index);
                    Pnext = T_ra(i,index+1);
                end
                Flags(i,j,2) = index;
                Flags(i,j,5) = T_ra(i,index);
            elseif j == 1 && T_ra(i,j) <= T_ra(i,j+1)
                Flags(i,j,1) = 0;
            elseif j == nB && T_ra(i,j) > T_ra(i,j-1)
                Flags(i,j,1) = 1;
                Flags(i,j,4) = T_ra(i,j);
                Pval = T_ra(i,j);
            end
        end
    end
end

```

```

Pprior = T_ra(i,j-1);
index = j;
% Find left hand valley:
while Pval > Pprior
    index = index - 1;
    Pval = T_ra(i,index);
    Pprior = T_ra(i,index-1);
end
Flags(i,j,3) = index;
Flags(i,j,6) = T_ra(i,index);
elseif j == nB && T_ra(i,j) <= T_ra(i,j-1)
    Flags(i,j,1) = 0;
% Logic to find peaks at all midpoints (and their R/L valleys)
elseif T_ra(i,j) >= T_ra(i,j-1) && T_ra(i,j) >= T_ra(i,j+1)
    Flags(i,j,1) = 1;
    Flags(i,j,4) = T_ra(i,j);
    % Find left hand valley:
    Pval = T_ra(i,j);
    Pprior = T_ra(i,j-1);
    index = j;
    if j > 2
        Switch = 1;
    else
        Switch = 0;
        index = 1;
    end
    while Pval > Pprior && Switch == 1;
        index = index - 1;
        Pval = T_ra(i,index);
        Pprior = T_ra(i,index-1);
        if index == 2
            Switch = 0;
            index = 1;
        end
    end
    Flags(i,j,3) = index;
    Flags(i,j,6) = T_ra(i,index);
    % Find right hand valley:
    Pval = T_ra(i,j);
    Pnext = T_ra(i,j+1);
    index = j;
    if j < nB - 1
        Switch = 1;
    else
        Switch = 0;
        index = nB;
    end
    while Pval > Pnext && Switch == 1;
        index = index + 1;
        Pval = T_ra(i,index);
        Pnext = T_ra(i,index+1);

```





```

    % Left Peak sig, right peak ignore
    PeakCode(1,j) = 1;
    PeakCode(2,j+1) = 3;
elseif VD1 <= SigD && VD2 >= SigD
    % Left peak ignore, right peak significant
    PeakCode(1,j) = 3;
    PeakCode(2,j+1) = 1;
elseif VD1 >= SigD && VD2 >= SigD
    % Both peaks are significant
    PeakCode(1,j) = 1;
    PeakCode(2,j+1) = 1;
end
end
% Logic for the peak endpoints
Left = abs(Flags(i,PeakVec(1,1),4) - Flags(i,PeakVec(1,1),6));
if Left >= SigD
    PeakCode(2,1) = 1;
else
    PeakCode(2,1) = 3;
end
Right = abs(Flags(i,PeakVec(1,Npeaks),4) - Flags(i,PeakVec(1,Npeaks),5));
if Right >= SigD
    PeakCode(1,Npeaks) = 1;
else
    PeakCode(1,Npeaks) = 3;
end
% Now determine what to do with the peaks. Look at combinations
% reading left to right. Put significant peaks into a vector.
SPV = 0;
j = 1;
count = 1;
while j <= Npeaks
    C1 = PeakCode(1,j);
    C2 = PeakCode(2,j);
    if C1 == 1 && C2 == 1
        % This is a bonafide unique peak. Put it in the vector.
        SPV(1,count) = PeakVec(1,j);
        count = count + 1;
        j = j+1;
    elseif C1 == 2
        % This peak will be merged with other peak(s)
        count2 = 2;
        Merge = PeakVec(1,j);
        while C1 == 2
            j = j + 1;
            C1 = PeakCode(1,j);
            Merge(1,count2) = PeakVec(1,j);
            count2 = count2 + 1;
        end
        % Insert logic here to eliminate valleys and slow rises (see
        % what is happening at the significant endpoints!)
    end
end

```

```

End1 = Flags(i,Merge(1,1),4) - Flags(i,Merge(1,1),6);
End2 = Flags(i,Merge(1,end),4) - Flags(i,Merge(1,end),5);
Distance = Merge(1,end) - Merge(1,1);
if End1 >= SigD && End2 >= SigD
    SPV(1,count) = round(mean(Merge));
    count = count + 1;
    j = j + 1;
elseif End1 >= SigD && End2 <= (SigD)*(-1) && Distance >= SigBinD
    SPV(1,count) = round(mean(Merge));
    count = count + 1;
    j = j + 1;
elseif End1 <= SigD*(-1) && End2 >= (SigD) && Distance >= SigBinD
    SPV(1,count) = round(mean(Merge));
    count = count + 1;
    j = j + 1;
end
else % Represents 1/3, 3/1, and 3/3 combos which are all ignored!!
    j = j + 1;
end
end
end
Nsigpeaks = length(SPV);
for j = 1:Nsigpeaks
    PeakLocAvg(i,j) = SPV(1,j);
end
%{
% Plot the PeakLocAvg to see how the algorithm performs)
S = Nsigpeaks;
FF = zeros(1,nB);
for k = 1:S
    Temp = PeakLocAvg(i,k);
    FF(1,Temp) = 1;
end
X = (1:200)';
Y1 = FF(1,:);
Y2 = T_ra(i,:);
plot(X,Y1,X,Y2)
stop = 1;
%}
end

return
end

%% ----- FIT STATES -----
function [NumofStates, BestFitTraces] = fitStates(maxN, time_vector, data2, data3, MaxNumofStates)
%This function determines the number of states for each trace

%INPUTS:
% maxN      maximum number of transitions for each traces
% time_vector  vector of times corresponding to each frame

```

```

% data2    second column of normalized data in *.dat files
% data3    third column of normalized data in *.dat files

%OUTPUTS
% NumofStates    row vector containing number of states for each trace
% BestFitTraces  matrix showing the current states occupied by each
%                trace for each frame (crude approximation)

A{1} = time_vector';
[row, cols] = size(data2); % # rows = # timeframes, # cols = # SWNT
for q = 1:1:cols % Loop through each SWNT trace
    %redefine variable to match formatting used in StatesFinder
    A{2} = data2(:,q);
    A{3} = data3(:,q);

    time=A{1};
    data=A{3}./1000;
    timeinc=time(2,1)-time(1,1);
    fulldata=[A{1} A{2} A{3}];

    fitdata=data;
    antifit=data;

    resultArray=zeros(maxN,3); %first column the beginning index. second column is the length. third
    colum is the step fit.
    resultArray(1,1)=1;
    resultArray(1,2)=length(data);
    resultArray(1,3)=mean(data);
    fitdata(:)=resultArray(1,3);
    chisquaredFit=zeros(maxN,1);
    chisquaredAntifit=zeros(maxN,1);
    chisquaredFit(1)=sum((data-fitdata).^2,1);
    SS=zeros(maxN,1);
    SS(1)=1;

    flag=0;
    resultArray0 = resultArray;
    for i=1:(maxN-1),
        para=0;
        for j=1:i,
            temp=resultArray(j,1);
            temp2=resultArray(j,2);
            B=data(temp:(temp2+temp-1));
            [StepSize,StepLocation,left,right]=TJstepFinder(B);
            temp3=StepSize*(temp2)^0.5;
            if temp3 > para %optimal num. of transitions (maximize temp3)
                para=temp3;
                newBegin1=temp;
                newLength1=StepLocation-1;
                newValue1=left;
                newBegin2=temp+StepLocation-1;

```

```

        newLength2=temp2-newLength1;
        newValue2=right;
        bestJ=j;
    end
end

resultArray(bestJ,1)=newBegin1;
resultArray(bestJ,2)=newLength1;
resultArray(bestJ,3)=newValue1;
resultArray(i+1,1)=newBegin2;
resultArray(i+1,2)=newLength2;
resultArray(i+1,3)=newValue2;
currentFitArray=sort_on_key(resultArray(1:i+1,:),1);
antiFitArray=zeros(i+2,3);
antiFitArray(1,1)=1;
antiFitArray(1,2)=length(data);
for j=1:(i+1),
    temp=currentFitArray(j,1);
    temp2=currentFitArray(j,2);
    if (temp2 == 1)
        flag = 1;
        break;
    end
    B=data(temp:(temp2+temp-1));
    [StepSize,StepLocation,left,right]=TJstepFinder(B);
    antiFitArray(j+1,1)=temp+StepLocation-1;
end
if (flag == 1)
    break;
end
for j=1:i+1,
    antiFitArray(j,2)=antiFitArray(j+1,1)-antiFitArray(j,1);
    antiFitArray(j,3)=mean(data(antiFitArray(j,1):(antiFitArray(j,2)+antiFitArray(j,1)-1)));
end
antiFitArray(i+2,2)=length(data)-antiFitArray(i+2,1)+1;
antiFitArray(i+2,3)=mean(data(antiFitArray(i+2,1):end));

for j=1:i+1,
    fitdata(resultArray(j,1):(resultArray(j,2)+resultArray(j,1)-1))=resultArray(j,3);
end
for j=1:i+2,
    antifit(antiFitArray(j,1):antiFitArray(j,2)+antiFitArray(j,1)-1)=antiFitArray(j,3);
end

chisquaredFit(i+1)=sum((data-fitdata).^2,1);
chisquaredAntifit(i+1)=sum((data-antifit).^2,1);
SS(i+1)=chisquaredAntifit(i+1)/chisquaredFit(i+1);

resultArray0 = resultArray;
end

```

```

resultArray = resultArray0;

[C,I]=max(SS);

bestfitnumber=I;
resultArray(1,1)=1;
resultArray(1,2)=length(data);
resultArray(1,3)=mean(data);
fitdata(:)=resultArray(1,3);
chisquaredFit=zeros(maxN,1);
chisquaredFit(1)=sum((data-fitdata).^2,1);
for i=1:(bestfitnumber),
    para=0;
    for j=1:i,
        temp=resultArray(j,1);
        temp2=resultArray(j,2);
        B=data(temp:(temp2+temp-1));
        [StepSize,StepLocation,left,right]=TJstepFinder(B);
        temp3=StepSize*(temp2)^0.5;
        if temp3 > para,
            para=temp3;
            newBegin1=temp;
            newLength1=StepLocation-1;
            newValue1=left;
            newBegin2=temp+StepLocation-1;
            newLength2=temp2-newLength1;
            newValue2=right;
            bestJ=j;
        end
    end
    resultArray(bestJ,1)=newBegin1;
    resultArray(bestJ,2)=newLength1;
    resultArray(bestJ,3)=newValue1;
    resultArray(i+1,1)=newBegin2;
    resultArray(i+1,2)=newLength2;
    resultArray(i+1,3)=newValue2;

end

for j=1:bestfitnumber+1,
    fitdata(resultArray(j,1):(resultArray(j,2)+resultArray(j,1)-1))=resultArray(j,3);
end
finalresult=sort_on_key(resultArray((1:bestfitnumber+1),:),1);
finalresult(:,1)=(finalresult(:,1)-1)*timeinc+time(1);
finalresult(:,2)=finalresult(:,2)*timeinc;

if MaxNumofStates == 10 %if maximum number of states is 10
    statesnum=length(unique(round(10*finalresult(:,3))));
    rounded_fit = round(fitdata*10)/10;
    fitdata_average = zeros(1,length(fitdata));

```

```

for i = 1:1:length(rounded_fit)
    if rounded_fit(i) ~= -1
        sameState_coord = find(rounded_fit==rounded_fit(i));
        total_sum = 0;
        for j = 1:1:length(sameState_coord)
            total_sum = total_sum+fitdata(sameState_coord(j));
            rounded_fit(sameState_coord(j)) = -1;
        end
        average_state = total_sum/(length(sameState_coord));
        for k = 1:1:length(sameState_coord)
            fitdata_average(sameState_coord(k)) = average_state;
        end
    end
end
else %if maximum number of states is something other than 10
    %from the fit, group states with similar values according to the
    %maximum number of states as specified by the user. First, create a
    %vector of states between 0 and 1 consisting of number of states
    %specified by the user. Then find the difference between each of these
    %states and the fitdata value, and whichever state i closest to the
    %fitdata value is determined to be the state.
    possible_states = 0:(1/MaxNumofStates):1;
    rounded_fit = zeros(1, length(fitdata));
    for i = 1:1:length(fitdata)
        difference = abs(fitdata(i)-possible_states);
        [dum, state_coord] = min(difference); %find the state closest to the fit data
        rounded_fit(i) = possible_states(state_coord);
    end
    %Now start to average all the states that were grouped into having the
    %same rounded state
    fitdata_average = zeros(1,length(fitdata));
    for i = 1:1:length(rounded_fit)
        if rounded_fit(i) ~= -1
            sameState_coord = find(rounded_fit==rounded_fit(i));
            total_sum = 0;
            for j = 1:1:length(sameState_coord)
                total_sum = total_sum+fitdata(sameState_coord(j));
                rounded_fit(sameState_coord(j)) = -1;
            end
            average_state = total_sum/(length(sameState_coord));
            for k = 1:1:length(sameState_coord)
                fitdata_average(sameState_coord(k)) = average_state;
            end
        end
    end
    end
    statesnum = length(unique(fitdata_average));
end
output(q,:) = [statesnum, fitdata_average];
end

NumofStates = output(:,1);

```

```

BestFitTraces = output(:, 2:end);

return;
end

%% ----- TJ STEP FINDER -----
function [StepSize,StepLocation,left,right]=TJstepFinder(B)

%This function is called upon by the fitStates function to locate steps in
%the traces.

global C; %this is what we send to TJflatFit

lengthB=length(B);

%Initial conditions
fitResult=B;
tempfit=B;
StepSize=0;
bestchisquared=10^10;
time=(1:lengthB);
StepLocation=1;
C=zeros(1,2);
C(:,1)=(1:1)';
C(:,2)=B(1:1);
A1=mean(C(:,2));
tempfit(1:1)=A1;
C=zeros((lengthB-1),2);
C(:,1)=((1+1):lengthB)';
C(:,2)=B((1+1):lengthB);
A2=mean(C(:,2));
StepSize=abs(A2-A1);
left=A1;
right=A2;

for i=1:(lengthB-1),
    C=zeros(i,2);
    C(:,1)=(1:i)';
    C(:,2)=B(1:i);
    A1=mean(C(:,2));%fmins('TJflatFit',mean(C(:,2)));
    tempfit(1:i)=A1;
    C=zeros((lengthB-i),2);
    C(:,1)=((i+1):lengthB)';
    C(:,2)=B((i+1):lengthB);
    A2=mean(C(:,2));%fmins('TJflatFit',mean(C(:,2)));
    tempfit((i+1):lengthB)=A2;
    chisquared=sum((tempfit-B).^2,1);
    if chisquared < bestchisquared,
        bestchisquared=chisquared;
        fitResult=tempfit;
        StepSize=abs(A2-A1);
    end
end

```

```

    StepLocation=i+1;
    left=A1;
    right=A2;
end
end

```

```

return;
end

```

```

%% -----SORT ON KEY -----

```

```

function sorteer=sort_on_key(rij,key)

```

```

%This function is called upon by the fitStates function to locate steps in
%the traces. It reads a (index,props) array and sorts along the index with one of the
%props as sort key

```

```

size=length(rij(:,1));
sorteer=0*rij;
buf=rij;
for i=1:size
    [g,h]=min(buf(:,key));
    sorteer(i,:)=buf(h,:);
    buf(h,key)=max(buf(:,key))+1;
end

```

```

return;
end

```

```

function M = GenTrace(A,B,C,numTr,Ttotal)

```

```

% Coded by Nigel Reuel on 3.3.2011
% This function generates general step traces according to three
% parameters which encompass all types of step data:
% A - tendency up / tendency down
% B - avg event time / signal sampling time
% C - avg step size / avg noise
% It generates a user-specified number of traces with these three
% parameters and returns a 3D matrix of the traces (first level - real
% response, 2nd level - noise). To aid in comparing traces, all traces are
% normalized on a scale of 0 to 1.

```

```

%
%

```

```

Tdown = 1;
Tup = A;
Tprob = Tdown + Tup;
%SigSampT = 1; %< --- Not used, but coded in my index numbers
AvgEventT = B;
AvgNoise = 1;
AvgStep = C;
%

```



```

% Set total simulation time and number of traces: - Now function input
% Ttotal = 1000;
% Initialize solution matrix:
M = zeros(Ttotal,numTr,2);
% Generate the traces
for i = 1:numTr
    t = 1;
    % Signal start (arbitrary due to normalization):
    S = 100;
    % Track signal changes:
    SigChange = [0 0];
    num = 1;
    while t < Ttotal
        SigChange(num,1) = t;
        % 1) Determine step size:
        SS = rand()*2*AvgStep;
        % 2) Determine the direction:
        rand1 = rand();
        check1 = rand1*Tprob;
        if check1 < Tdown
            S = S - SS;
        else
            S = S + SS;
        end
        SigChange(num,2) = S;
        num = num + 1;
        % 3) Determine the next event time:
        tstep = rand()*2*AvgEventT;
        t = t + tstep;
    end
    % Turn signal change matrix into real trace and trace w/ noise
    Real_Tr = zeros(Ttotal,1);
    Noise_Tr = zeros(Ttotal,1);
    Intervals = round(SigChange(:,1));
    Intervals(num,1) = Ttotal;
    for j = 1:num-1
        for k = Intervals(j,1):Intervals(j+1,1)
            Real_Tr(k,1) = SigChange(j,2);
            NoiseL = (rand()*2-1)*AvgNoise;
            Noise_Tr(k,1) = SigChange(j,2) + NoiseL;
        end
    end
    % Normalize according to max and min levels in traces
    Max(1,1) = max(Real_Tr);
    Max(2,1) = max(Noise_Tr);
    Min(1,1) = min(Real_Tr);
    Min(2,1) = min(Noise_Tr);
    MaxN = max(Max);
    MinN = min(Min);
    SigDiff = MaxN-MinN;
    Real_n = (Real_Tr-MinN)./SigDiff;

```

```

Noise_n = (Noise_Tr-MinN)./SigDiff;
M(:,i,1) = Real_n;
M(:,i,2) = Noise_n;
% Graphical check of traces:
%{
X = 1:1000;
X = X';
plot(X,Real_n,X,Noise_n);
%}
end
return
end

function [u,sig,t,iter] = fit_mix_gaussian( X,M )
%
% Gaussian Fit Function by Ohad Gal (c) 2003
% Download at:
% http://www.mathworks.com/matlabcentral/fileexchange/4222-a-collection-of-fitting-functions
% on 3.21.2011
%
% fit_mix_gaussian - fit parameters for a mixed-gaussian distribution using EM algorithm
%
% format: [u,sig,t,iter] = fit_mix_gaussian( X,M )
%
% input: X - input samples, Nx1 vector
%        M - number of gaussians which are assumed to compose the distribution
%
% output: u - fitted mean for each gaussian
%         sig - fitted standard deviation for each gaussian
%         t - probability of each gaussian in the complete distribution
%         iter- number of iterations done by the function
%
% initialize and initial guesses
N = length( X );
Z = ones(N,M) * 1/M; % indicators vector
P = zeros(N,M); % probabilities vector for each sample and each model
t = ones(1,M) * 1/M; % distribution of the gaussian models in the samples
u = linspace(min(X),max(X),M); % mean vector
sig2 = ones(1,M) * var(X) / sqrt(M); % variance vector
C = 1/sqrt(2*pi); % just a constant
Ic = ones(N,1); % - enable a row replication by the * operator
Ir = ones(1,M); % - enable a column replication by the * operator
Q = zeros(N,M); % user variable to determine when we have converged to a steady
solution
thresh = 1e-3;
step = N;
last_step = inf;
iter = 0;
min_iter = 10;

% main convergence loop, assume gaussians are 1D

```

```

while ( (( abs((step/last_step)-1) > thresh) & (step>(N*eps)) ) | (iter<min_iter) )

% E step
% =====
Q = Z;
P = C ./ (Ic*sqrt(sig2)) .* exp( -(X*Ir - Ic*u).^2 ./ (2*Ic*sig2) );
for m = 1:M
    Z(:,m) = (P(:,m)*t(m))./(P*t(:));
end

% estimate convergence step size and update iteration number
prog_text = sprintf(repmat( '\b',1,(iter>0)*12+ceil(log10(iter+1)) ));
iter      = iter + 1;
last_step = step * (1 + eps) + eps;
step      = sum(sum(abs(Q-Z)));
fprintf( '%s%d iterations\n',prog_text,iter );

% M step
% =====
Zm      = sum(Z);           % sum each column
Zm(find(Zm==0)) = eps;     % avoid division by zero
u       = (X')*Z ./ Zm;
sig2    = sum(((X*Ir - Ic*u).^2).*Z) ./ Zm;
t       = Zm/N;
end

sig     = sqrt( sig2 );

return
end

```



## 7. 3D Tracking of Carbon Nanotubes within Living Cells

Some of the work, text and figures presented in this chapter are reprinted or adapted from “Three-Dimensional Tracking of Carbon Nanotubes within Living Cells,” Nigel F. Reuel, Aurélie Dupont, Olivier Thouvenin, Don C. Lamb, and Michael S. Strano. *ACS Nano* **2012** 6 (6), 5420-5428 (reprinted under thesis use allowance from the American Chemical Society © 2012).

### 7.1 Background and Motivation

Semiconducting single-walled carbon nanotubes (SWNTs), which fluoresce in the near infrared (nIR) due to their unique band-gap structure,<sup>1</sup> can be engineered into biosensor constructs by non-covalent wrappings that are 1) selective to target analytes (such as glucose,<sup>2,3</sup> NO,<sup>4</sup> proteins,<sup>5</sup> and glycans<sup>6</sup>) and 2) modulate the SWNT fluorescent signal when a binding event occurs.<sup>7</sup> However, much remains to be understood regarding SWNT interaction with biological cells and tissues. In particular, their mechanisms of endocytosis and trafficking, aggregation and exocytosis.<sup>8</sup> Past work in the Strano group utilized two-dimensional tracking of SWNTs using nIR fluorescent microscopy to elucidate the mechanistic steps of endo- and exo-cytosis in NIH3T3 cells.<sup>9</sup> Subsequently, we derived and validated a model that described the size-dependent uptake of SWNTs *via* endocytosis.<sup>10</sup> However, 2D tracking was limited in the spatial and temporal resolution as the particles drifted in the z-direction out of the focal volume or out of the field of view in the x-y direction. Feedback-based, three-dimensional nanoparticle tracking can overcome these limitations by changing the field of view or focal volume depth in real time to follow the motion of the particle. This allows one to follow SWNTs in the focus of the microscope over longer periods of time and one has the possibility to extract additional information such as corral volumes (rather than corral surfaces or regions found in 2D tracking) or determine the local viscosity experienced by the nanoparticles. As the SWNT remains in the focus of the tracking microscope, the fluorescence intensity can also be used during the trajectory to provide orientation information of anisotropic SWNT particles. Understanding the rotational behavior and corresponding signal fluctuations at varying viscosities is necessary to utilize the fluorescence intensity as a readout of the SWNT sensor.<sup>11</sup> Thus, not only is the pathway of the particle mapped through the cell but also the production or consumption of a specific analyte (protein, sugar, or small molecule) can be measured at different locations within a cell.

Here, we demonstrate 3D tracking of single SWNTs with an orbital tracking microscope,<sup>12-14</sup> and extend this technique to SWNTs localized within living HeLa cells. Mathematical modeling and analysis of the experimental *in vitro* trajectories shows the effect of solution viscosity on SWNT rotational motion

and confirms the hypothesis that SWNT rotation causes signal fluctuations.<sup>15</sup> At high viscosities (250-1000 cP), the rotation is slow enough to use the rotational diffusion coefficient for accurate SWNT length calculations<sup>5,15-17</sup> whereas, at lower viscosities, the translational coefficient is preferred. After quantifying the translational and rotational characteristics of SWNTs as a function of viscosity, we used them to probe the viscosity of a HeLa cell. The internalized SWNTs demonstrated normal diffusion, active transport, and confined diffusion. The 3D motion of the SWNTs was used to calculate the active transport velocities and corral volumes; the relationship between SWNT rotation and translation determined *in vitro* was used to predict local apparent viscosities and SWNT lengths. Finally, we show how the rotational diffusion constant can be used to evaluate the level of background noise of a SWNT sensor and what level of quenching or fluorescence enhancement would be necessary for a statistically significant binding event to be distinguished from fluorescence fluctuations due to rotation.

## 7.2 Methods

### 7.2.1 Preparation of Chitosan Wrapped SWNT Suspension

The SWNT were prepared as previously published.<sup>6</sup> Briefly, 3 mg of SWNT (Southwest Nanotechnologies – (6,5) chirality > 70%) were added to a chitosan solution. The chitosan was made soluble by protonation in an acidic solution (add 1 vol% acetic acid). The resulting mixture was sonicated for 40 min at 10W amplitude with a probe tip sonicator (Cole Parmer, Model CV18). The black suspension was centrifuged at 16,000g for 2 hours using a table top centrifuge (Labnet Inc) and the supernatant was collected. The centrifugation process was repeated two more times to eliminate any SWNT aggregates. The supernatant can be stored at room temperature in a sealed container for more than 3 months without any aggregation. The final concentration of our SWNT sample was 15 mg/L as found by optical absorbance at 990nm using the calibration value for (6,5) SWNTs of 13 mg/L for an absorption of 1 OD at 990 nm.<sup>18</sup>

### 7.2.2 Tracking of SWNT Particles

First, the optimal concentration of SWNT must be determined for single particle tracking. This was done by immobilizing SWNT in a moisture curable resin (MY-133MC from MY Polymers Ltd) at varying SWNT concentrations and imaging the sample to see whether single particles were resolvable. At concentrations greater than 0.5% vol, the sample emitted large photon counts above 950 nm, but single particles could not be detected (bright background). At a concentration of 0.5% vol, we were able to visualize single SWNT as fluorescent diffraction-limited spots and ‘track’ them by moving the stage

holding the immobilized SWNTs in a set 2D sinusoidal and 3D box pattern as has been discussed previously.<sup>12</sup> Seven 100  $\mu\text{l}$  SWNT-glycerol solutions were created by adding approximately 40, 50, 60, 75, 85, 95, and 99.5  $\mu\text{l}$  of glycerol to 0.5  $\mu\text{l}$  of SWNT stock. Particle-filtered (Millipore) pure water was added to make up the remaining volume. The solutions were bath sonicated for 15 minutes in a warm bath ( $T \approx 30^\circ\text{C}$ ) to ensure complete mixing of the viscous solutions and then allowed to cool to room temperature.

After calibrating the system (to determine pixel size and ensure tracking accuracy), SWNT tracking in 3D was accomplished by entrapping them in a sample chamber and acquiring tracks using the Globals for Images (Sim-FCS) software from the Gratton group.<sup>19</sup> Sample chambers were constructed from thin (No. 1) coverslips separated by 300 $\mu\text{m}$  of Parafilm (two film layers) on both sides of a small channel between the two coverslips. The SWNT solutions were drawn into the chambers *via* capillary flow and the input and output of the chamber were sealed by adding melted paraffin at the ends of the channel. The small sample chambers allowed the SWNT to freely diffuse in all 3 dimensions without wall interactions and preserved the sample for multiple days of imaging. The chambers were placed on the water-immersed objective (Nikon water immersion objective, NA 1.20 63x) and the middle of the fluid chamber was determined by observing the reflected excitation beam at the surface of the top and bottom coverslip and positioning the objective between these points. The 561nm laser (50 mW Cobolt Jive) was set to full power (0.4 mW at the sample) to observe the SWNT. The photon counts fluctuated during a tracking event but on average were 3 times greater than the background count. This signal-to-noise ratio could be drastically improved by utilizing nIR detectors.

Data analysis was then done with a few custom Matlab programs (all available in Appendix 1-4). The text log output from the SimFCS software was parsed to extract the X, Y, and Z coordinates as well as the photon counts from the upper and lower detectors. As the traces contained some dead time before and after the real tracking events or jumping between different tracked particles, another automated program was used to analyze the data and find significant tracking events (Code in Appendix 7.1). Specified criteria were longer tracks (greater than 100 steps) that had small internal step lengths (total squared displacement less than  $3 \times 10^4 \text{ nm}^2$  between steps or average of 100 nm in each direction). Finally, another program was used to analyze each of these significant traces with the following steps: 1) calculate MSD plots from the position and rotational data, 2) fit the first 10% of the time steps in the MSD plot to determine the translational and rotational diffusion coefficients, and 3) calculate lengths using the equations detailed by Marshall *et al.*<sup>20</sup> A distribution of the lengths was then created from the individual measurements.

### **7.2.3 Tracking of SWNT Particles in Live HeLa Cells**

Live HeLa cells were seeded at a density of  $1 \times 10^4$  cells/well into 8 well chamber slides (IBIDI). The cell culture media was replaced prior to experiment by warm DPBS buffer and the sample was maintained at  $37^\circ\text{C}$  on the microscope stage. A target cell was visualized using the widefield channel of the microscope. First, the center of the cell was approximated by bringing the intracellular compartments in sharp focus and then a stack of five images below and five above the center point were acquired at a step resolution of 200 nm using an external z-axis control. This stack of images was later used to reconstruct the cell volume (*via* the darker cell walls) and see if the overlaid SWNT tracks were within the cell volume. SWNT were added from the stock solution to the media (10  $\mu\text{L}$ , final concentration 2.5% vol) without disturbing the cell in focus. The SWNTs diffuse too quickly in aqueous solution to be tracked with the current setup. After 30 minutes, the SWNT are taken up by the HeLa cells and the motion of the SWNTs can then be followed. After overlaying the acquired tracks, 17 were found within the cell volume along with one extra-cellular SWNT resting on or near the plasma membrane.

#### 7.2.4 AFM Imaging of SWNT

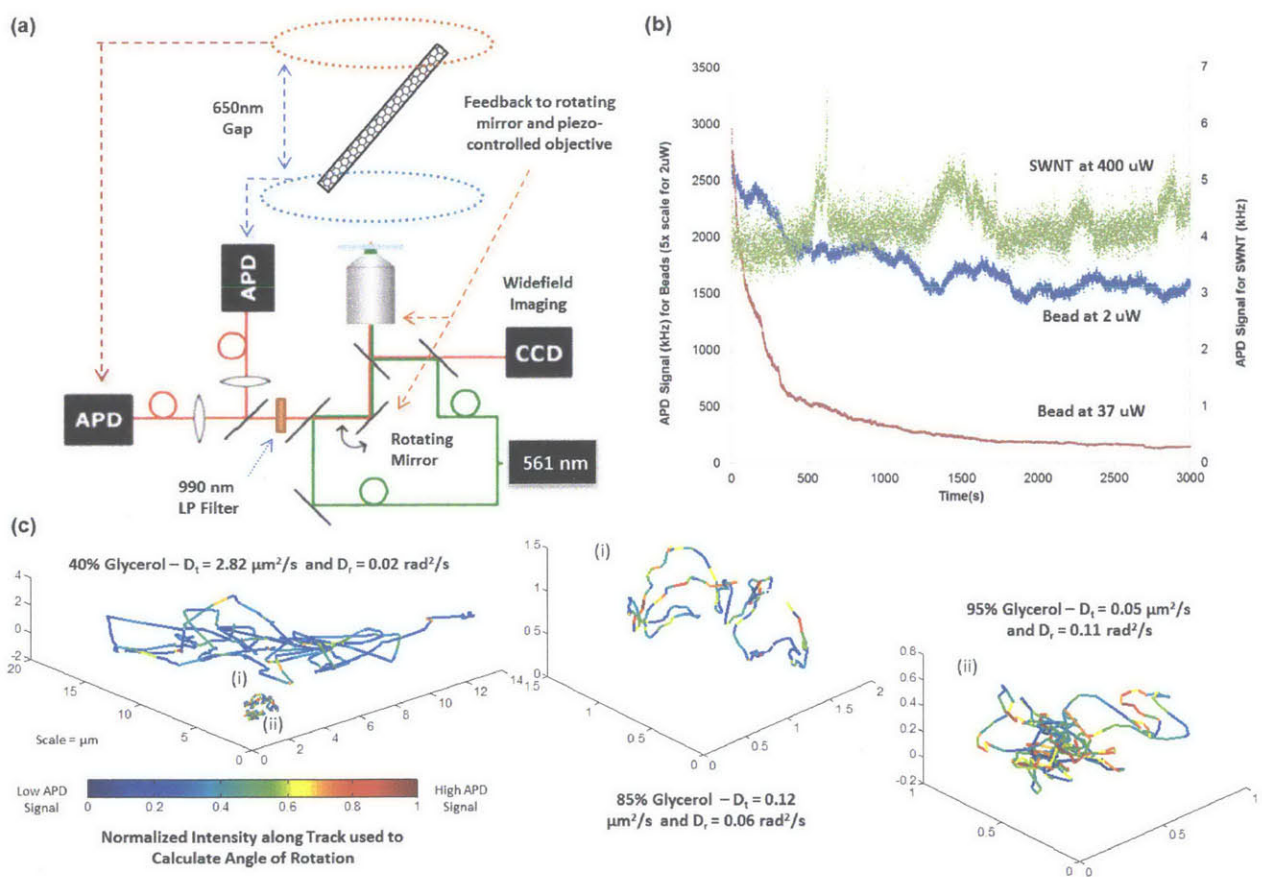
For AFM measurements, the stock solution of chitosan SWNT was diluted to 10x and deposited on a clean silicon wafer. After allowing for physisorption for 30 min, the chip is washed with water three times and then dried with nitrogen. The SWNTs on the chip are then imaged with AFM (Asylum Research) using a silicon nitride tip (Olympus Micro Cantilevers – resonant frequency of 70 kHz, spring constant of 2 N/m) in tapping mode at 700mW drive set point and scan angle of  $90^\circ$ . A scan area of 5  $\mu\text{m}$  square and scan frequency of 0.5 Hz was found to visualize SWNT well. The SWNT lengths were measured by the multi-point length determining tool of Gwyddion software (freely available online).

### 7.3 Results and Discussion

An orbital tracking system (with established spatial resolutions of 7-15 nm and temporal resolution of 32 ms)<sup>12</sup> was used to visualize chitosan wrapped SWNT.<sup>6</sup> The orbital microscope utilizes an orbiting excitation beam and a real-time feedback control system (Figure 7.1a) to keep the particle in focus: 1) deviations from the orbit's center give rise to an intensity modulation from which the particle's position can be inferred and the orbit location is adjusted to follow the particle *via* the piezo-mirror; 2) deviations in the z-plane are detected *via* the difference in photon counts between the two, off-focus detection planes and the microscope responds by a piezo-controlled objective. The microscope was originally designed for use with visible fluorescent probes<sup>12,13</sup> and was therefore modified for nIR detection by exchanging the emission filter with a 950 nm longpass filter in the emission path to the silicon avalanche photo diodes (APD) (Figure 7.1a). The APDs have a low quantum efficiency in the nIR but the sensitivity is still sufficient for single particle tracking. The (6,5) SWNT chirality was targeted as



its excitation resonance matched the 561 nm laser source and its emission (990 nm)<sup>1</sup> could be detected on the APD. The large Stokes shift of the SWNT fluorescence is advantageous for tracking in live cells as the absorption of the nIR emission from the SWNTs is low ('Tissue Transparent Window') and there will be minimal background autofluorescence.<sup>21</sup> Our ability to detect and track SWNT with orbital tracking was tested using immobilized particles in cured polymer, which were first located and then tracked while being moved in known test patterns by a piezo stage. It was found that the stock solution of SWNT (15 mg/L) must be diluted to 0.5 vol% to detect and track individual particles. Three-dimensional single particle tracking was then performed on SWNT in solution using the Globals for Images (Sim-FCS) software.<sup>19</sup> SWNTs exhibited their characteristic high photostability even at maximum laser power during tracking, whereas the fluorescent beads used for calibration quickly bleached at a rate dependent on laser power (Figure 7.1b).



**Figure 7.1** - Three dimensional tracking of SWNT. a) Schematic diagram of the orbital tracking microscope with a 990 nm longpass filter placed in the fluorescence beam path. b) Fluorescence intensity of a SWNT and fluorescent beads for comparison at varying laser power (quantum efficiency of silicon detectors at 990 nm is less than 10%, thus, the SWNT signal is comparatively low but photostable). c) Representative SWNT tracks at approximately 40, 85, and 95 vol% glycerol (19.3, 6.9, and 8.9 s tracks respectively) with accompanying translational (2.82, 0.12, 0.05

$\mu\text{m}^2/\text{s}$  respectively) and rotational diffusion coefficients (0.03, 0.06 and 0.11  $\text{rad}^2/\text{s}$  respectively). The color of the track denotes the signal intensity to show the fluctuations during observed tracking (normalized scale).

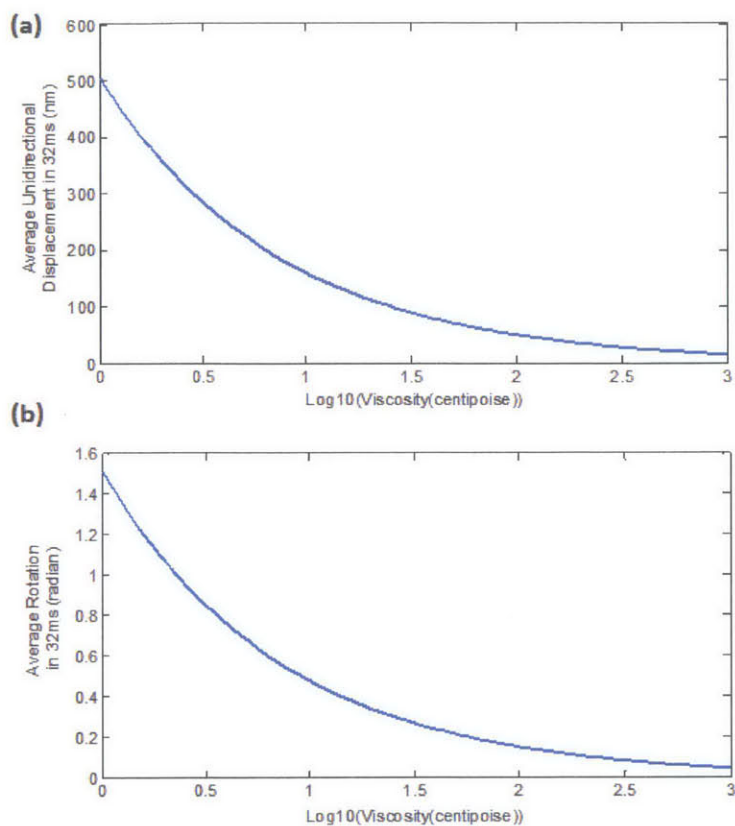
As observed also in previous 2D tracking work,<sup>15</sup> SWNT diffusion in pure water is too rapid to allow single particle tracking in aqueous solution with our orbital tracking microscope with 32 ms per orbit (on average 500 nm in each direction per 32ms orbit – Fig 7.2). Hence, we increased the viscosity with glycerol and measured the free diffusion of SWNTs in seven different concentrations of glycerol (approximately 40, 50, 60, 75, 85, 95, and 99.5 vol% in water, Table 7.1) (Figure 1c). As pipetting was only approximate and the hygroscopic properties of glycerol make the high glycerol concentrations unreliable, we experimentally determined the viscosity of the different solutions (Table 7.1). We used 190nm diameter fluorescent beads as a calibration and extracted the viscosity from the diffusion coefficient determined from a mean square displacement (MSD) analysis assuming the Stokes-Einstein relationship:<sup>22</sup>

$$D = \frac{k_B \cdot T}{6 \cdot \pi \cdot \eta \cdot r} \quad \text{Eq.7.1}$$

where  $D$  is the translational diffusion coefficient,  $k_B$  is the Boltzmann constant,  $T$  is temperature,  $\eta$  is viscosity of the solvent and  $r$  is the radius of the spherical particle. During tracking, the fluorescence intensity of the SWNTs fluctuated (Figure 7.1b) significantly more than the control beads or when the SWNTs were placed in cured polymer. These fluctuations were attributed to the rotation of the SWNTs in solution as hypothesized by Tysbouski *et al.*<sup>15</sup> We also tested this hypothesis by simulating the effect of SWNT rotation in a random walk model (Appendix 7.2) and analyzing the simulation results with the same analysis protocol used for the acquired experimental tracks.

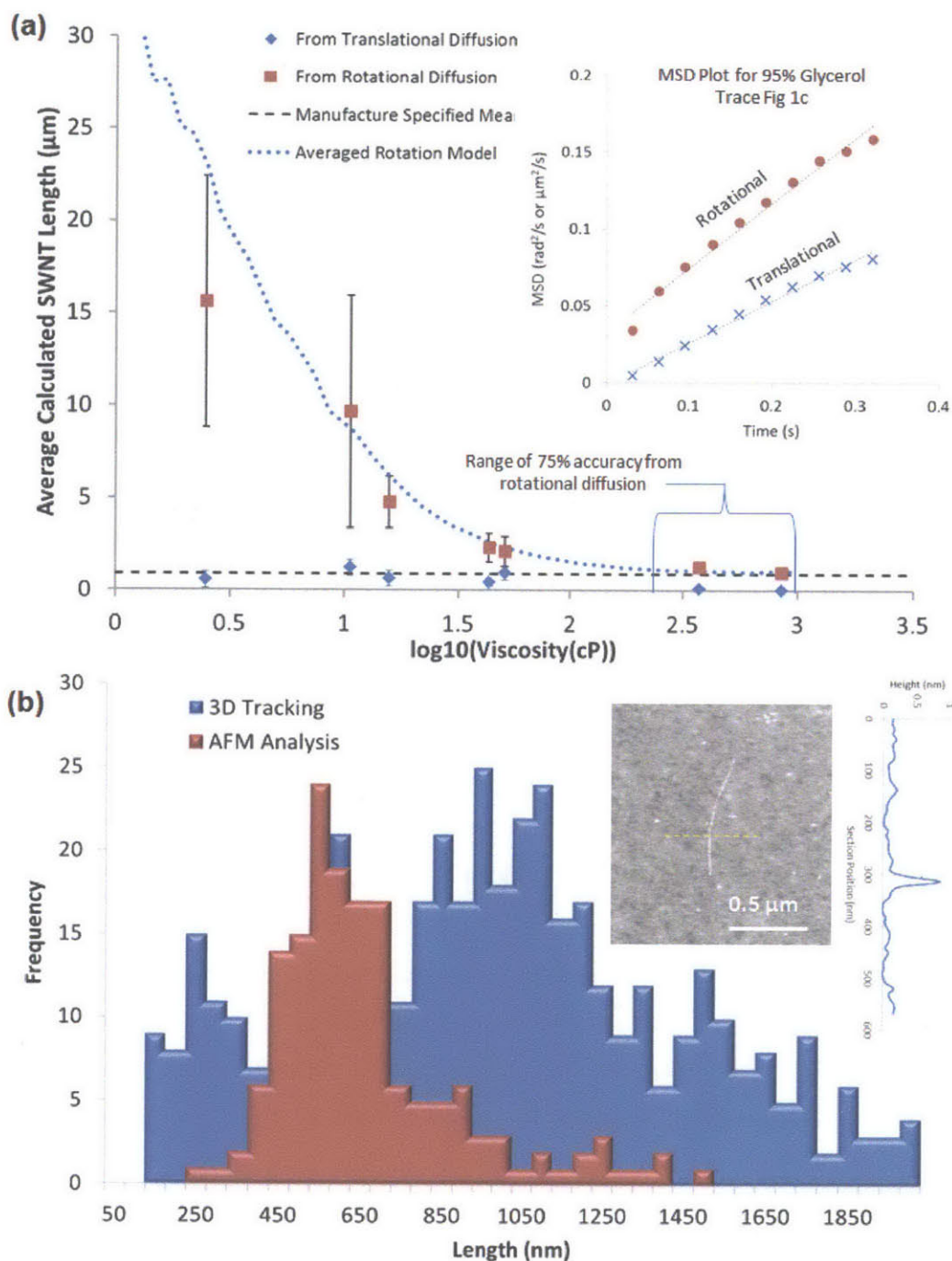
**Table 7.1** – Calculated viscosities for seven samples with different concentrations of glycerol. Averaged viscosity values come from 8-10 experimental tracks.

Sample #	Glycerol Content (vol %)	Calculated Viscosity from Bead MSD (cp) and 95% confidence interval
1	40%	2.5 ± 0.7
2	50%	10.6 ± 4.2
3	60%	15.7 ± 10.8
4	75%	43.7 ± 5
5	85%	51.4 ± 9
6	95%	374 ± 110
7	99.5%	856 ± 200



**Figure 7.2** - Plots were generated from equations 7.2 and 7.3 for a 900 nm nanotube at viscosities from 1 to 1000 centipoise. They illustrate how the unidirectional displacement (a) and degree of rotation (b) change with relation to viscosity.

The length of the nanotube is an unknown parameter that needs to be determined, either through translation diffusion or rotational diffusion assuming that SWNT rotation causes fluctuations in the fluorescence signal. Calculating nanotube length from simulated and experimentally acquired trajectories and intensity data is straightforward. For translation diffusion MSD plots ( $\mu\text{m}^2$  vs.  $\Delta t$  – Figure 7.3a insert) are created from the 3D trajectories recorded by the orbital tracking microscope and the diffusion coefficient is determined from the slope of MSD curve<sup>20</sup> (code in Appendix).



**Figure 7.3 - Measured length distribution.** a) Effect of viscosity on length predictions from rotational diffusion (red) and translational diffusion (blue). An averaged rotation model is overlaid (···) as well as the manufacture specified length of 900 nm (---). Brackets denote the range of viscosities from random walk model where the rotational diffusion coefficient yields length calculations within 25% of the true length. Insert shows typical MSD plot used to calculate the SWNT length. b) Histogram of SWNT lengths calculated from 3D tracking (blue) and AFM surface measurements (red). A representative AFM scan is shown in insert along with a height profile of the single 0.8 nm diameter SWNT deposited on silicon.

For determination of rotational diffusion, the intensity values were first scaled to geometrical angles of rotation (ranging from 0 to  $\pi/2$ ) by assuming a maximum signal when the nanotube's length is parallel to the linear polarization of the impinging excitation (perpendicular to direction of the light - angle of 0) and a minimum when the nanotube is perpendicular to the polarization ( $\pi/2$ ),<sup>23</sup> with a  $\cos^2(\theta)$  proportion holding between these two endpoints.<sup>23,24</sup> For this assumption to hold, the nanotube must always be centered in the excitation pathway, which is a good approximation as the orbital tracking microscope continually recenters the excitation beam around the particle. This simple approximation also ignores the decrease in linear polarization of the excitation beam due to dichroics or the use of high numerical aperture objectives. However, although the approximation is simple, the comparison of calculated lengths to the measured and known physical lengths is very good. After this scaling is made, the MSD of the angle ( $\text{rad}^2$  vs.  $\Delta t$  – Figure 7.3a insert) is calculated and the rotational diffusion coefficients determined from the slope of the MSD curve<sup>20</sup> (code in Appendix 7.3). The rotational diffusion of nanorods can also be determined using fluorescence correlation spectroscopy from fluctuations induced by rotation of the SWNT in the polarized excitation beam.<sup>24</sup> In fact, orbital tracking should allow direct observation of the rotational diffusion in FCS without being affected by translation diffusion as the excitation beam is tracking the SWNT. However, the trajectories measured in this work were too short to perform a quantitative correlation analysis. Thus, we continued with our angle scaling approximation to find the rotational diffusion constant and used the equations outlined by Marshall *et al.*<sup>20</sup> without corrections for sidewall interactions to calculate the nanotube's length. This was done for both the translational (Eq.6.2) and rotational (Eq.6.3) diffusion coefficients (code in Appendix 7.3):

$$D_{trans} = \frac{k_B \cdot T \cdot \left[ 3 \cdot \ln\left(\frac{L}{d}\right) + 2 \cdot B + C \right]}{8 \cdot \pi \cdot \eta \cdot L} \quad \text{Eq.7.2}$$

$$D_{rot} = \frac{k_B \cdot T \cdot 3 \left[ \ln\left(\frac{L}{d}\right) + A \right]}{\pi \cdot \eta \cdot L^3} \quad \text{Eq.7.3}$$

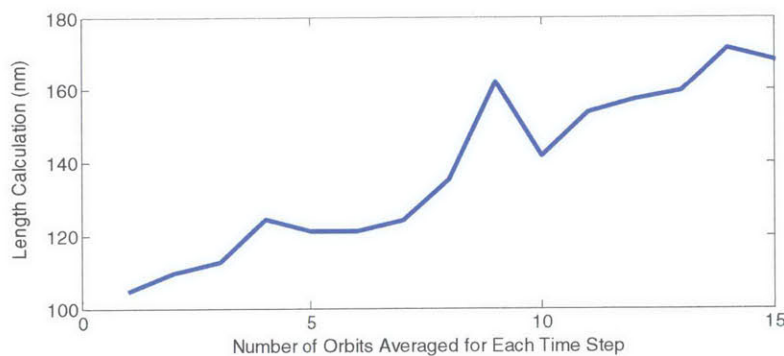
where  $D_{trans}$  and  $D_{rot}$  are the translational and rotational diffusion coefficients respectively,  $k_B$  is the Boltzmann constant,  $T$  is temperature,  $\eta$  is viscosity,  $L$  is the length of the nanotube, and  $d$  is the diameter of the nanotube (held constant at 0.8 nm). The parameters  $A$ ,  $B$ , and  $C$  are geometrical correction factors, which for an infinitely long cylinder are -0.447, -0.114, and 0.886 respectively.<sup>25</sup>

The rotational diffusion coefficient is preferred to calculate SWNT length due to its higher sensitivity (function of  $L^3$  and not  $L$  – Eq. 7.2 and 7.3). However, this is only possible at viscosities where the rotation is slow enough for the orbit to accurately map changes in rotation between each orbit, and thus depends on the viscosity, size, orbital speed etc. To determine at what viscosity the length

calculation accuracy from rotational diffusion begins to break down for the current 32 ms minimum orbit time, a random walk model was created in Matlab (Appendix 7.2). A 900 nm long nanotube was simulated to take 10,000 steps at 1 ms per step dictated by translational and rotational diffusion coefficients. One thousand traces were generated for each of the 50 viscosities (ranging from 1-1000 cP). The 'experimental traces' were simulated by averaging the positions and rotational signals of the particle for every 32 ms period. The resulting 'experimental traces' were then analyzed to determine the diffusion coefficients (from MSD plots) and the coefficients were then used with Eq. 7.2 and Eq. 7.3 to determine the limits of length calculation from rotational diffusion. It was found that above 250 cP the length predictions are within 25% of the true length value when the rotational coefficient is used. At viscosities below this point, the rotational diffusion should not be used and the translational diffusion coefficient should be taken for length predictions.

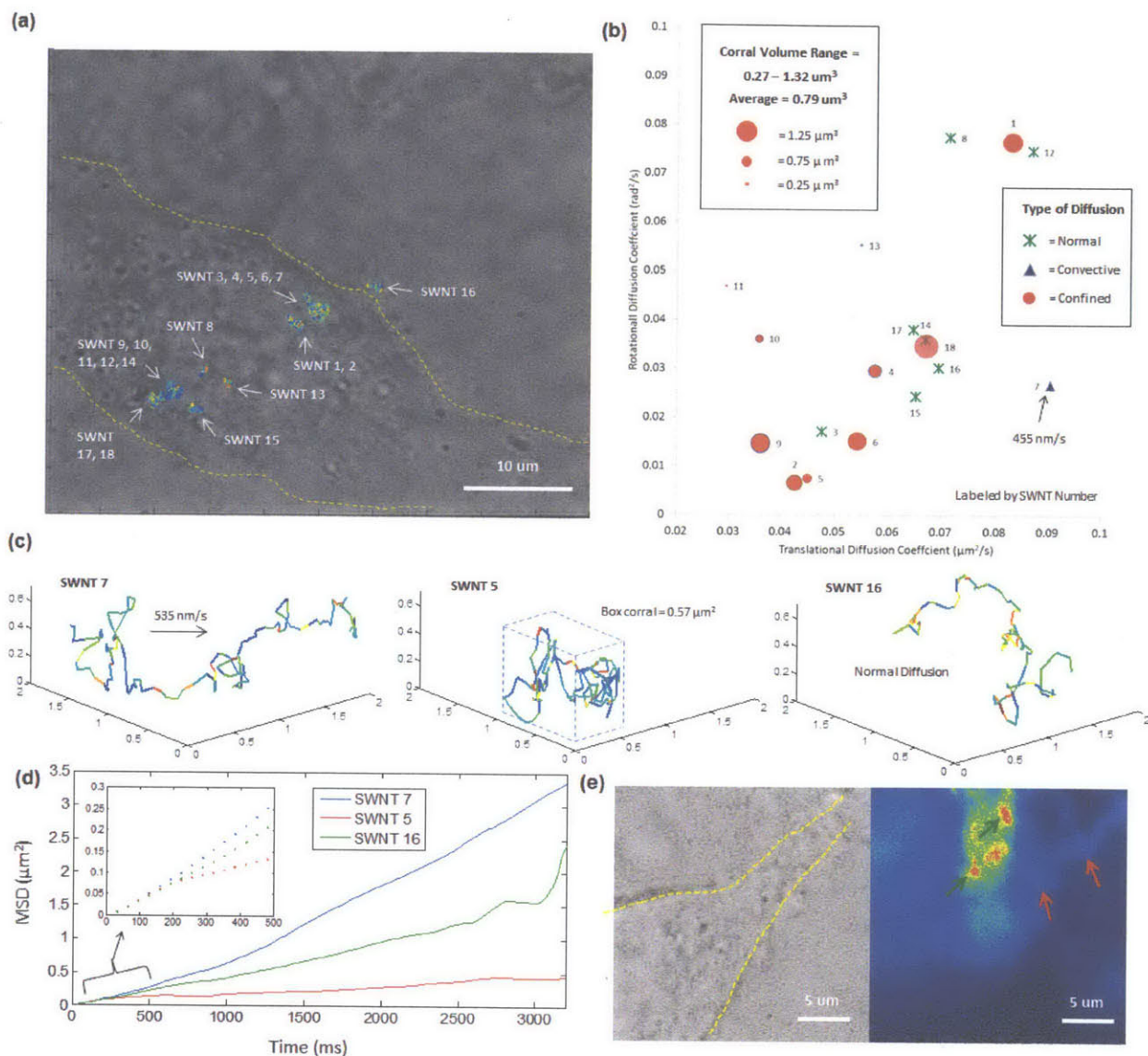
This viscosity limitation is seen in our experimental data when we plot both length predictions *versus* viscosity (Figure 7.3a). When the model is plotted on the experimental data (dotted line Figure 7.3a) we see a good fit further establishing the assumption that we can correlate signal fluctuations to SWNT rotation. When we create a distribution of lengths (Figure 7.3b) calculated from the tracks (using translational diffusion coefficients for the 40, 50, 60, 75, and 85 glycerol vol % samples and rotational for the 95 and 99.5 vol %), we find that, of the 515 successful SWNT tracks, 476 had length predictions within the expected range of 0.5 to 2  $\mu\text{m}$  provided by the manufacturer. The remaining lengths (2-8  $\mu\text{m}$ ) were assigned to larger aggregates of SWNT and excluded from further analysis. Interestingly, the length distribution found by single particle tracking more closely fits the manufacturer's range and mode (400-2000 nm and 900 nm respectively) than measurements with the same SWNT solution on a surface *via* AFM (Figure 7.3b). It is known that surfaces that irreversibly absorb particles will disproportionately display the smaller end of the distribution, as these species diffuse the fastest to the surface from bulk and capture the available area before larger particles reach the surface.<sup>26-28</sup> The experimental tracks also exhibit less accurate length predictions from translational diffusion at higher viscosities (Figure 7.3a). As the SWNT slows down in a more viscous solution, macroscopic errors begin to dominate over the microscopic movements. These errors can come from tracking inaccuracies (7-10 nm in each direction), changes in the location of the fluorescence along the length of the ~900 nm SWNT, or errors associated with microscope drift. All of these movements cause an overestimate of the actual diffusion coefficient, particularly at high viscosity. Averaging multiple orbit steps during post-processing to lengthen the SWNT movement between time steps improved the results but was still not sufficient to overcome the macroscopic errors (below). To calculate accurate translational diffusion coefficients at such high viscosities, the measurement has to be optimized with great care given to the stability of the system with respect to the diffusing step sizes.

To try and overcome the macroscopic errors (drift, tracking error, and movement along the SWNT axis) we studied the effect of averaging multiple orbit periods in order to obtain a longer time-step allowing the SWNT to diffuse further thus overcoming the experimental errors. This was done for the 95 and 99.5 vol% glycerol traces. The following plot (Fig 7.4) is the result of averaging the length predictions from translational diffusion coefficient at 95 vol% and seeing how the average length calculation changes dependent on the averaging window (number of orbits that are combined). The averaged orbits do ‘improve’ the predictions – but there is still a significant discrepancy between the determined length of the SWNT from translational diffusion and the average 900 nm value determined from rotational diffusion. This suggests that other sources of error such as microscope drift are dominant at short times in viscous media.



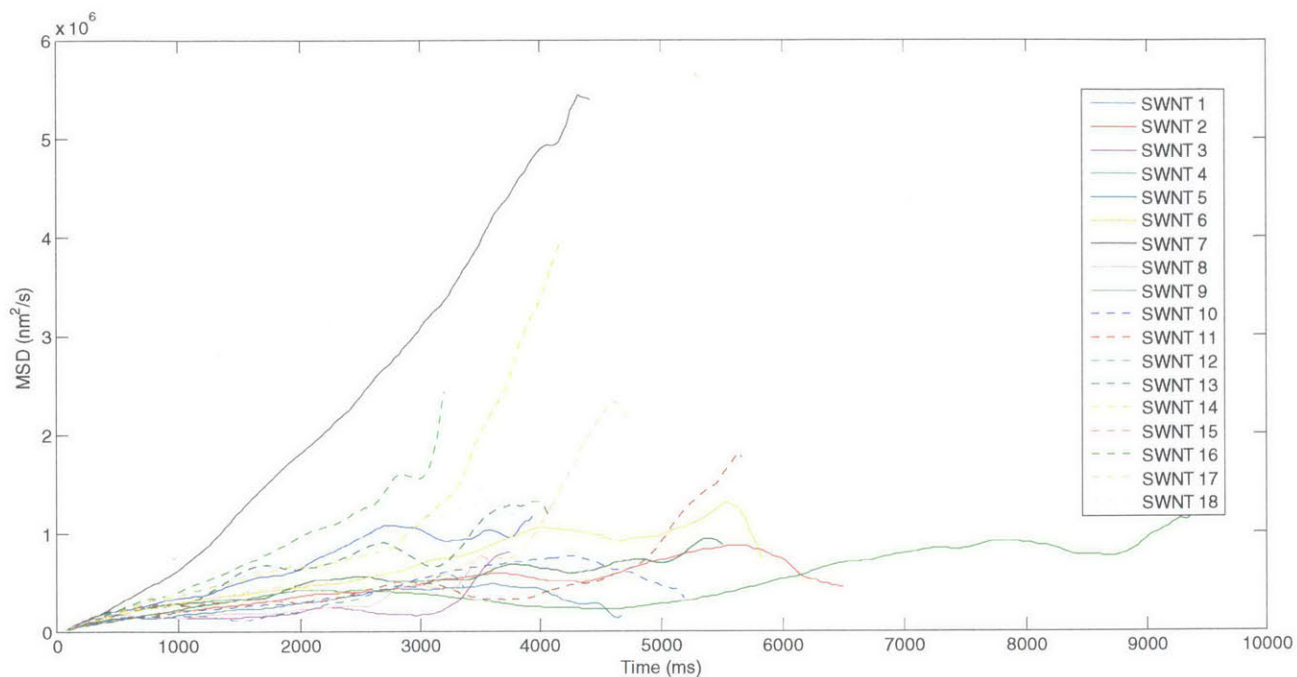
**Figure 7.4** - Effect of averaging the orbits to improve length predictions from high viscosity data. Here all of the 95 vol % glycerol traces are used to calculate lengths from translational diffusion coefficients. The predictions generally improve upon averaged orbits, but the effect is small. If one is concerned with accurate translational coefficients at such a high velocity, they should focus on minimizing the experimental errors introduced by the setup.

We then demonstrated 3D tracking of SWNT in living HeLa cells and used the above relations found *in vitro* to determine local environment conditions of the cell such as apparent viscosity and regions of SWNT confinement *versus* active transport. The microscope setup (Figure 7.1a) allows for both tracking and widefield imaging; thus the acquired 3D traces can be mapped into a three-dimensional image of the cell volume during post processing (code available in Appendix 7.4). Of the 50 traces we acquired during cell experiments, 17 tracks were found to reside within a single HeLa cell and an additional track was observed in the vicinity of the exterior cell surface (Figure 7.5a – the other tracks were found outside cell walls). The track lengths (3.2-14.4 s) were shorter and more difficult to obtain than *in vitro* (3.2 to 90 s) due to increased signal-to-noise ratio within the cell. By analyzing the MSD plots of these tracks (Fig 7.5d, Fig 7.6, Fig 7.7), we could deduce the type of diffusion (normal, convective or confined<sup>8,12,29</sup>) as well as the translational and rotational diffusion constants (Figure 7.5b and Table 7.2).

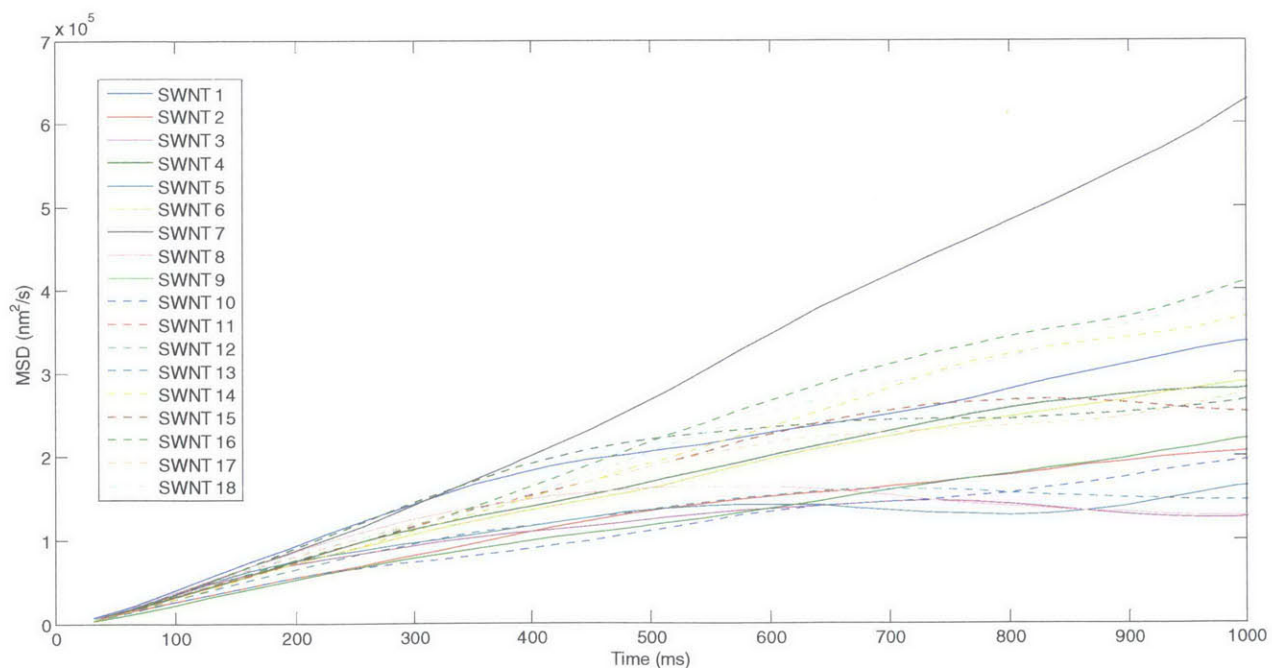


**Figure 7.5** - 3D SWNT tracks in HeLa cell. a) Overlaid 2D location of 18 SWNT tracks in relation to HeLa cell (outlined in yellow). b) Master plot of the translational, rotational diffusion coefficients, types of diffusion, and corral volumes or active transport velocities for 18 measured trajectories. Ten of the tracks demonstrated corralled diffusion with an average volume of  $0.79 \mu\text{m}^3$ . (c) Representative 3D plots of three SWNT tracks demonstrating the three types of diffusion and accompanying MSD plots (d). The color of the 3D tracks denote the level of intensity to show the fluctuations during tracking (same normalized scale as Figure 1). Internalized SWNT exhibit normal, convective, and corralled diffusion. e) After 1.5 hours, internalized SWNT have clearly formed large aggregates within (red arrows) and surrounding the cell (white arrows). Left image is from the widefield channel (yellow lines denote approximate cell boundaries) and the right is an integrated (20s) photon count from the SWNT emission (arbitrary color scale denoting the fluorescence intensity).





**Figure 7.6** – Full trace MSD plots for translational motion of 18 SWNT tracks within HeLa cell. Diffusion coefficients are calculated from the first 10% of each trace not from the full trace.

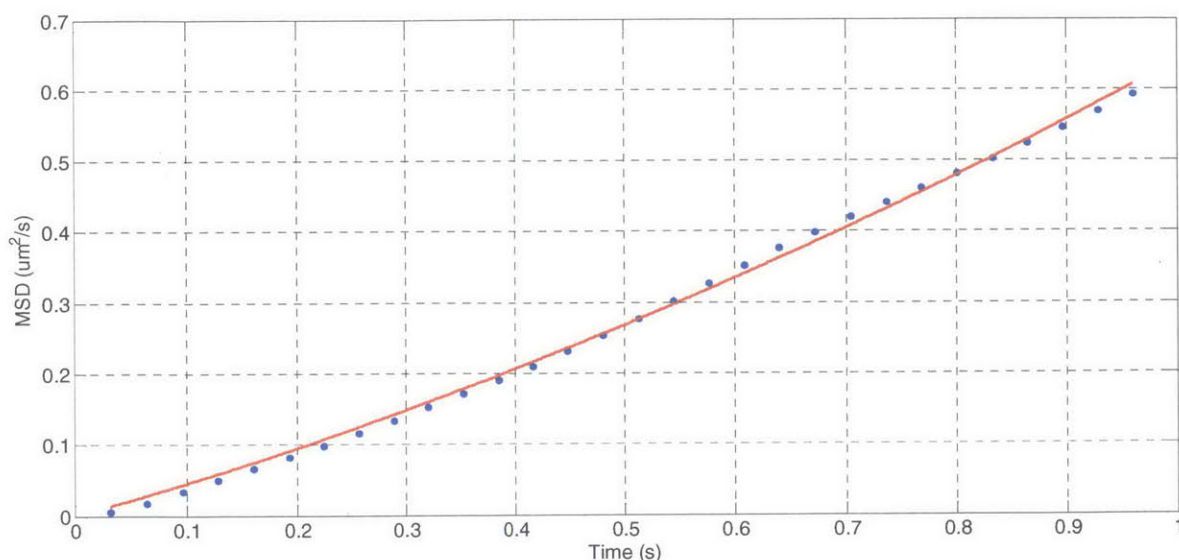


**Figure 7.7** – Close up MSD plots for translational motion of 18 SWNT tracks within HeLa cell

**Table 7.2** – Calculated results from the 3D SWNT Tracks in HeLa Cell

SWNT #	Local Viscosity (cP)	Probable SWNT Length (nm)	$D_T$ ( $\mu\text{m}^2/\text{s}$ )	$D_R$ ( $\text{rad}^2/\text{s}$ )	Diffusion Type	Corral Volume ( $\mu\text{m}^3$ )	Convective Velocity (nm/s)	Avg. rot during 32ms (rad)	% ( $I_{\text{max}}-I_{\text{min}}$ ) for Significant Event Signal
1	78	513	0.083	0.077	Confined	1.07	--	0.070	13%
2	72	1204	0.043	0.007	Confined	0.94	--	0.020	4%
3	85	880	0.048	0.017	Normal	--	--	0.033	6%
4	83	725	0.058	0.030	Confined	0.80	--	0.043	8%
5	72	1147	0.045	0.008	Confined	0.57	--	0.022	4%
6	73	899	0.054	0.015	Confined	1.02	--	0.031	6%
7	54	706	0.090	0.027	Active Transport	--	455	0.041	8%
8	89	521	0.071	0.078	Normal	--	--	0.070	13%
9	105	947	0.036	0.015	Confined	1.12	--	0.031	6%
10	137	704	0.036	0.036	Confined	0.52	--	0.048	9%
11	179	650	0.029	0.047	Confined	0.27	--	0.055	10%
12	74	514	0.087	0.075	Normal	--	--	0.069	13%
13	104	594	0.055	0.055	Confined	0.27	--	0.060	11%
14	77	668	0.067	0.036	Normal	--	--	0.048	9%
15	71	758	0.065	0.024	Normal	--	--	0.039	8%
16	71	705	0.069	0.030	Normal	--	--	0.044	8%
17	80	660	0.065	0.038	Normal	--	--	0.049	9%
18	76	675	0.067	0.035	Confined	1.32	--	0.047	9%

Ten of the tracks were found to be confined and the confinement volumes were approximated with rectangular corral volumes ranging from 0.27 to 1.32  $\mu\text{m}^3$ . One track (Figure 7.5c) demonstrated convective diffusion with an active transport velocity of 455 nm/s (422-490 for 95% confidence with a  $R^2$  0.998, Fig 7.8), slightly larger than those found by 2D tracking of quantum dots in HeLa cells.<sup>30</sup> This discrepancy is likely due to the added contribution of the velocity along the third dimension that is not available in 2D measurements. Our measured transport velocity falls within the range observed for dynein<sup>31</sup> (centripetal and centrifugal directions were  $691 \pm 233$  nm/s and  $676 \pm 214$  nm/s respectively) and other kinesin motors<sup>32</sup> ( $570 \pm 20$  nm/s) within HeLa cells; these are the cargo-carrying proteins that march along structural microtubules. It is also at the upper end of the range of velocities measured for actin-mediated transport (200-500 nm/s).<sup>33</sup>



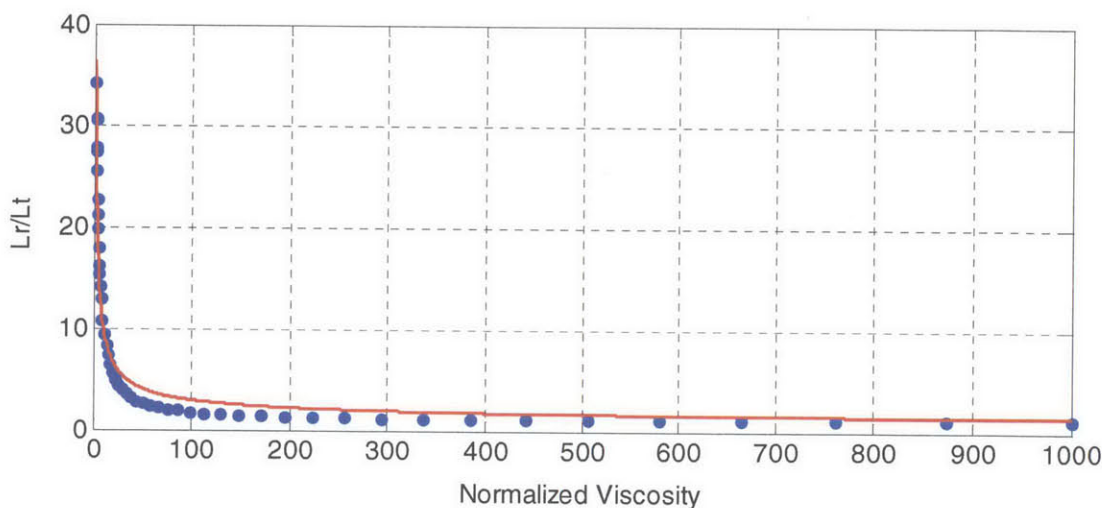
**Figure 7.8** – Fit to determine active transport velocity – blue dots are the MSD data and red line is the fit to  $MSD = 6 \cdot D \cdot t + (v \cdot t)^2$ . Velocity was found to be 455 nm/s (422-490 for 95% confidence interval) with an  $R^2$  fit of 0.998

For the SWNT undergoing random motion within the cell, we could use the translational and rotational data of the tracked SWNT to determine both the probable SWNT length and apparent local viscosity within the cell. This was done by first fitting the random walk model to create an expression for the ratio of length calculations *versus* solution viscosity. The form of the fit function is predicted by taking the approximate ratio of Eq. 3 to Eq. 2 and inserting the inverse relationship of diffusion constants to viscosity (Eq.7.1). By neglecting the many constants and focusing on the length values, we get a function form of:

$$\frac{L_R}{L_T} = A \cdot \left(\frac{\eta}{\eta_o}\right)^{-\frac{2}{3}} + 1 \quad \text{Eq. 7.4}$$

where  $L_R$  and  $L_T$  are the rotational and translational length calculations,  $\eta$  is viscosity of the solution,  $\eta_o$  is viscosity of water, and  $A$  is a fit coefficient. The model predictions are fit well ( $R^2$  value of 0.985) by this function form with a value of 41.31 for  $A$  (Fig 7.9). The same code used to calculate SWNT length *in vitro* is then used with the exception that the viscosity is now treated as an unknown parameter. The program was modified to sweep through viscosities in the range of 1-300 cP and compare the ratio of the resulting length calculations ( $L_R/L_T$ ) to Eq. 4 until the two are in agreement. In this manner, the SWNT length and local viscosities can be determined (Table 1). The calculated internal viscosities (54-179 cP) are greater than the recent simulation results of Kalwarczyk *et al.*<sup>34</sup> who estimated the macroscopic cytoplasmic viscosity of HeLa cells to be 44cP based on estimated length scales inserted in their

rheological model. The higher measured viscosity is likely due to the effects of cellular crowding on macromolecular transport<sup>35</sup> that might not be fully captured in their model. This effect is exaggerated with SWNT which have a length dimension two orders of magnitude greater than large proteins. The viscosity of Track 16 located near the cell membrane (71 cP) indicates that this track is indeed interacting with the cell surface and not freely diffusing in the media. This would be expected as diffusion of SWNTs in pure water was too fast to image with our current setup. Full endosomal aggregation of SWNT occurred roughly 1.5 hours after the stock solution of SWNT was added to the cell medium (2.5% vol SWNT). Several SWNT formed large fluorescent aggregates within the cell (Figure 7.5e), similar to agglomerates in endosomes observed previously by our laboratory<sup>9,10</sup> and others.<sup>36-38</sup>



**Figure 7.9** – The random walk model data (blue dots) with a fit equation (red line – Eq 7.4) used to deduce local viscosity and SWNT length of intracellular tracks.

With the ultimate goal of utilizing the SWNT as internal sensors, we estimate the quality of SWNT sensors that will need to be engineered to clearly transduce fast binding events from rotational diffusion (occurring on the time scale of the orbit or faster). From the measured radial diffusion coefficient, we calculated the average rotation during our 32 ms orbit (Table 7.2). By evaluating the ratio of this rotation amount to the full range of motion ( $\pi/2$ ), we can estimate the amount of ‘noise’ that will be present in a given track. Using a common sensor heuristic that the signal event must be three times that of the background noise, we can estimate the signal size that would be necessary to report a clear binding event – represented as a percentage of the fluorescence range presented upon SWNT rotation or  $I_{\max}-I_{\min}$  (Table 7.2). For the 18 tracks, we see that the binding event would need to quench or increase the nanotube fluorescent range by 4 to 13%. This requirement can be lessened by increasing the orbital speed where fluorescence fluctuations due to rotation of the SWNT are diminished. For example, if the orbit is

decreased to 3.2 ms, the fluorescence change required for sensor applications would be 1 to 4% and only a change of 1% would be necessary for a 0.32 ms orbit, provided the signal from the SWNT is sufficient for such quick orbits.

## 7.4 Conclusion

Herein we have successfully demonstrated 3D tracking of fluorescent SWNT for the first time, *in vitro* and in live HeLa cells, with over 500 tracks collected from our modified orbital tracking microscope. We have created a random walk model to describe the effect of viscosity on length calculation accuracy due to averaging during a 32 ms orbit of the excitation beam. Our model shows that, for viscosities from 250-1000 cP, the rotational diffusion coefficient is accurate (within 25%), but at low viscosities the translation diffusion coefficient should be used. Our *in vitro* experimental results corroborate the model by plotting the calculated lengths from both translational and rotational diffusion coefficients for the 476 experimental tracks *versus* solution viscosity. Furthermore, the close fit justifies the assumption made that fluctuations in the SWNT intensity during tracking are due to SWNT rotation as previously hypothesized.<sup>15</sup> AFM was also used to characterize the distribution of SWNT lengths. The length distribution was shorter than the distribution determined from 3D tracking, consistent with the property of absorbing surfaces to be covered preferentially with the smaller, faster diffusion particles.

Tracking was then demonstrated in living HeLa cells revealing normal and confined diffusion as well as active transport. By fitting to a random walk model, we created an expression that allowed us to calculate local viscosity conditions within the cell as well as the SWNT length. The intracellular tracks also demonstrated that moderately large response SWNT sensors (4 to 13% of  $I_{\max} - I_{\min}$ ) are needed to overcome the signal noise introduced by rotation. This constraint is relaxed significantly when faster orbital times can be used. Although the demonstrated tracks are short (2-90 s), the photostability of SWNT would allow for collection of longer trajectories in an orbital tracking microscope specifically designed for nIR emission (800 nm-1600 nm). This could be done by installing nIR sensitive detectors with, for example, Germanium or InGaAs elements. SWNT emission can also be improved with a more powerful excitation source (200 mW-1 W). However, this is not recommended as higher laser powers increase the photo-induced toxicity in cells. With improvements in detection, one could better track the location of an individual SWNT sensor from cell internalization to eventual aggregation and also better deconvolve the fluctuating SWNT signal with a faster orbital time. In this study, we have established the contribution of SWNT rotation to signal fluctuation and its dependence on solution viscosity. This will be applied in future studies where the SWNT is utilized as an *in vivo* sensor to monitor binding events along the tracked pathway, such as reversible quenching in the presence of nitrous oxide.<sup>39</sup>

## 7.5 References

- 1 Bachilo, S. M. *et al.* Structure-assigned optical spectra of single-walled carbon nanotubes. *Science* **298**, 2361-2366, doi:10.1126/science.1078727 (2002).
- 2 Barone, P. W. & Strano, M. S. Reversible control of carbon nanotube aggregation for a glucose affinity sensor. *Angewandte Chemie-International Edition* **45**, 8138-8141, doi:10.1002/anie.200603138 (2006).
- 3 Barone, P. W., Parker, R. S. & Strano, M. S. In vivo fluorescence detection of glucose using a single-walled carbon nanotube optical sensor: Design, fluorophore properties, advantages, and disadvantages. *Analytical Chemistry* **77**, 7556-7562, doi:10.1021/ac0511997 (2005).
- 4 Zhang, J. Q. *et al.* Single Molecule Detection of Nitric Oxide Enabled by d(AT)(15) DNA Adsorbed to Near Infrared Fluorescent Single-Walled Carbon Nanotubes. *Journal of the American Chemical Society* **133**, 567-581, doi:10.1021/ja1084942 (2011).
- 5 Ahn, J. H. *et al.* Label-Free, Single Protein Detection on a Near-Infrared Fluorescent Single-Walled Carbon Nanotube/Protein Microarray Fabricated by Cell-Free Synthesis. *Nano Letters* **11**, 2743-2752, doi:10.1021/nl201033d (2011).
- 6 Reuel, N. F. *et al.* Transduction of Glycan-Lectin Binding Using Near-Infrared Fluorescent Single-Walled Carbon Nanotubes for Glycan Profiling. *Journal of the American Chemical Society* **133**, 17923-17933, doi:10.1021/ja2074938 (2011).
- 7 Boghossian, A. A. *et al.* Near-Infrared Fluorescent Sensors based on Single-Walled Carbon Nanotubes for Life Sciences Applications. *Chemsuschem* **4**, 848-863, doi:10.1002/cssc.201100070 (2011).
- 8 Strano, M. S. & Jin, H. Where is it heading? Single-particle tracking of single-walled carbon nanotubes. *Acs Nano* **2**, 1749-1752, doi:10.1021/nn800550u (2008).
- 9 Jin, H., Heller, D. A. & Strano, M. S. Single-particle tracking of endocytosis and exocytosis of single-walled carbon nanotubes in NIH-3T3 cells. *Nano Letters* **8**, 1577-1585, doi:10.1021/nl072969s (2008).
- 10 Jin, H., Heller, D. A., Sharma, R. & Strano, M. S. Size-Dependent Cellular Uptake and Expulsion of Single-Walled Carbon Nanotubes: Single Particle Tracking and a Generic Uptake Model for Nanoparticles. *Acs Nano* **3**, 149-158, doi:10.1021/nn800532m (2009).
- 11 Boghossian, A. A. *et al.* The chemical dynamics of nanosensors capable of single-molecule detection. *Journal of Chemical Physics* **135**, doi:084124  
10.1063/1.3606496 (2011).
- 12 Dupont, A. & Lamb, D. C. Nanoscale three-dimensional single particle tracking. *Nanoscale* **3**, 4532-4541, doi:10.1039/c1nr10989h (2011).
- 13 Katayama, Y. *et al.* Real-Time Nanomicroscopy via Three-Dimensional Single-Particle Tracking. *Chemphyschem* **10**, 2458-2464, doi:10.1002/cphc.200900436 (2009).
- 14 Levi, V., Ruan, Q., Kis-Petikova, K. & Gratton, E. Scanning FCS, a novel method for three-dimensional particle tracking. *Biochem. Soc. Trans.* **31**, 997-1000 (2003).
- 15 Tsyboulski, D. A., Bachilo, S. M., Kolomeisky, A. B. & Weisman, R. B. Translational and rotational dynamics of individual single-walled carbon nanotubes in aqueous suspension. *Acs Nano* **2**, 1770-1776, doi:10.1021/nn800364r (2008).
- 16 Duggal, R. & Pasquali, M. Dynamics of individual single-walled carbon nanotubes in water by real-time visualization. *Physical Review Letters* **96**, doi:246104 10.1103/PhysRevLett.96.246104 (2006).
- 17 Branca, C., Magazu, V. & Mangione, A. Determination of MWNTs length-to-diameter ratio by static and dynamic light scattering. *Diamond and Related Materials* **14**, 846-849, doi:10.1016/j.diamond.2004.10.009 (2005).
- 18 Ming, Z. & Diner, B. A. Solution redox chemistry of carbon nanotubes. *Journal of the American Chemical Society* **126**, 15490-15494, doi:10.1021/ja0457967 (2004).

- 19 Gratton, E. *Globals Software for Spectroscopy and Images*, <<http://www.lfd.uci.edu/globals/>> (2011).
- 20 Marshall, B. D., Davis, V. A., Lee, D. C. & Korgel, B. A. Rotational and translational diffusivities of germanium nanowires. *Rheologica Acta* **48**, 589-596, doi:10.1007/s00397-009-0361-0 (2009).
- 21 Yi, H. *et al.* M13 Phage-Functionalized Single-Walled Carbon Nanotubes As Nanoprobes for Second Near-Infrared Window Fluorescence Imaging of Targeted Tumors. *Nano Letters*, doi:10.1021/nl2031663 (2012).
- 22 Edward, J. T. MOLECULAR VOLUMES AND STOKES-EINSTEIN EQUATION. *J. Chem. Educ.* **47**, 261-& (1970).
- 23 Duesberg, G. S., Loa, I., Burghard, M., Syassen, K. & Roth, S. Polarized Raman spectroscopy on isolated single-wall carbon nanotubes. *Physical Review Letters* **85**, 5436-5439, doi:10.1103/PhysRevLett.85.5436 (2000).
- 24 Tsay, J. M., Doose, S. & Weiss, S. Rotational and translational diffusion of peptide-coated CdSe/CdS/ZnS nanorods studied by fluorescence correlation spectroscopy. *Journal of the American Chemical Society* **128**, 1639-1647, doi:10.1021/ja056162i (2006).
- 25 Broersma, S. VISCOUS FORCE AND TORQUE CONSTANTS FOR A CYLINDER. *Journal of Chemical Physics* **74**, 6989-6990, doi:10.1063/1.441071 (1981).
- 26 Anancharungsuk, W., Taweepreda, W., Wirasate, S., Thonggoom, R. & Tangboriboonrat, P. Reduction of Surface Friction of Natural Rubber Film Coated with PMMA Particle: Effect of Particle Size. *Journal of Applied Polymer Science* **115**, 3680-3686, doi:10.1002/app.31459 (2010).
- 27 Qiu, H. W., Lee, W. Y. & Sukhishvili, S. A. Layer-by-layer self-assembly of ceramic particles for coating complex shape substrates. *Journal of the American Ceramic Society* **89**, 1180-1187, doi:10.1111/j.1551-2916.2005.00847.x (2006).
- 28 Rogers, S. S., Venema, P., Sagis, L. M. C., van der Linden, E. & Donald, A. M. Measuring the length distribution of a fibril system: A flow birefringence technique applied to amyloid fibrils. *Macromolecules* **38**, 2948-2958, doi:10.1021/ma0474224 (2005).
- 29 Bickel, T. A note on confined diffusion. *Physica a-Statistical Mechanics and Its Applications* **377**, 24-32, doi:10.1016/j.physa.2006.11.008 (2007).
- 30 Jiang, X. E. *et al.* Endo- and Exocytosis of Zwitterionic Quantum Dot Nanoparticles by Live HeLa Cells. *Acs Nano* **4**, 6787-6797, doi:10.1021/nn101277w (2010).
- 31 Kobayashi, T. & Murayama, T. Cell Cycle-Dependent Microtubule-Based Dynamic Transport of Cytoplasmic Dynein in Mammalian Cells. *Plos One* **4**, doi:e7827  
10.1371/journal.pone.0007827 (2009).
- 32 Courty, S., Luccardini, C., Bellaiche, Y., Cappello, G. & Dahan, M. Tracking individual kinesin motors in living cells using single quantum-dot imaging. *Nano Letters* **6**, 1491-1495, doi:10.1021/nl060921t (2006).
- 33 Pierobon, P. *et al.* Velocity, Processivity, and Individual Steps of Single Myosin V Molecules in Live Cells. *Biophys. J.* **96**, 4268-4275, doi:10.1016/j.bpj.2009.02.045 (2009).
- 34 Kalwarczyk, T. *et al.* Comparative Analysis of Viscosity of Complex Liquids and Cytoplasm of Mammalian Cells at the Nanoscale. *Nano Letters* **11**, 2157-2163, doi:10.1021/nl2008218 (2011).
- 35 Ando, T. & Skolnick, J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 18457-18462, doi:10.1073/pnas.1011354107 (2010).
- 36 Porter, A. E. *et al.* Uptake of Noncytotoxic Acid-Treated Single-Walled Carbon Nanotubes into the Cytoplasm of Human Macrophage Cells. *Acs Nano* **3**, 1485-1492, doi:10.1021/nn900416z (2009).
- 37 Yehia, H. *et al.* Single-walled carbon nanotube interactions with HeLa cells. *J Nanobiotechnology* **5** (2007).

- 38 Cui, H. F., Vashist, S. K., Al-Rubeaan, K., Luong, J. H. T. & Sheu, F. S. Interfacing Carbon Nanotubes with Living Mammalian Cells and Cytotoxicity Issues. *Chemical Research in Toxicology* **23**, 1131-1147, doi:10.1021/tx1000506 (2010).
- 39 Kim, J. H. *et al.* The rational design of nitric oxide selectivity in single-walled carbon nanotube near-infrared fluorescence sensors for biological detection. *Nature Chemistry* **1**, 473-481, doi:10.1038/nchem.332 (2009).

## Appendix 7.1

Matlab Code for determining significant nanotubes

```
function SWNTTracePrep
% Coded on 7.18.2011 by Nigel F. Reuel
% This program goes through a list of SWNT traces acquired in the day and
% determines the unique SWNT tracking segments.
%
% Criteria for Uniqueness:
SDcut = 3*10^4; % nm^2 - Squared Deviation Cut off
NScut = 100; % Number of steps required to qualify it as a trace
%
% Specify the files
Nfiles = 50;
% Initialize Unique Trace saving:
UTcount = 1;
% Loop through all the files
for iii = 1:Nfiles

    % Use Matlab function to read in the file (without the header):
    [~, M] = feval(@hdrload, ['S',int2str(iii),'.txt']);
    % Initial visualization
    X = M(:,1);
    Y = M(:,2);
    Z = M(:,3);
    %{
    if iii == 7
        plot3(X,Y,Z)
        stope = 5;
    end
    %}
    I1 = M(:,16);
    I2 = M(:,17);
    % Determine the MSD of the trace with step size of 1
    % Scale of the pixels:
    scale = 92; % nm/pixel
    scaleZ = 29; % nm/pixel
    SD = zeros(length(X)-1,2);
    for j = 1:length(X)-1
        delX = ((X(j+1,1)-X(j,1)))*scale;
        delY = ((Y(j+1,1)-Y(j,1)))*scale;
        delZ = ((Z(j+1,1)-Z(j,1)))*scaleZ;
        SD(j,1) = delX^2 + delY^2 + delZ^2;
        if delX^2 + delY^2 + delZ^2 < SDcut;
            SD(j,2) = 1;
        end
    end
end
% Now check for regions that match the length criterion specified
```



```

% above:
count = 1;
while count < length(X)-1
    count2 = 0;
    if SD(count,2) == 0
        count = count + 1;
    else
        while SD(count,2) == 1 && count < length(X)-1
            count = count + 1;
            count2 = count2 + 1;
        end
        if count2 >= NScut;
            % Congrats! You have a trace segment that is considered
            % significant. Save the X,Y,Z,I1,I2 data of this trace.
            Start = count - count2;
            End = count - 1;
            m(:,1) = X(Start:End,1);
            m(:,2) = Y(Start:End,1);
            m(:,3) = Z(Start:End,1);
            m(:,4) = I1(Start:End,1);
            m(:,5) = I2(Start:End,1);
            % Full title to see where the trace comes from:

csvwrite(['SigT_',int2str(UTcount),'_FileNum_',int2str(iii),'_ST_',int2str(St
art),'_ET_',int2str(End),'.csv'],m)
            % Shortened title for the analyzer program:
            csvwrite(['SigT_',int2str(UTcount),'_ForAnalyzer.csv'],m)
            UTcount = UTcount + 1;
            clearvars m;
        end
    end
end
end

% End of outer file reading loop:
end
end

function [header, data] = hdrload(file)

% HDRLOAD Load data from an ASCII file containing a text header.
% [header, data] = HDRLOAD('filename.ext') reads a data file
% called 'filename.ext', which contains a text header. There
% is no default extension; any extensions must be explicitly
% supplied.
%
% The first output, HEADER, is the header information,
% returned as a text array.
% The second output, DATA, is the data matrix. This data
% matrix has the same dimensions as the data in the file, one
% row per line of ASCII data in the file. If the data is not
% regularly spaced (i.e., each line of ASCII data does not
% contain the same number of points), the data is returned as
% a column vector.
%
% Limitations: No line of the text header can begin with

```

```

%      a number.  Only one header and data set will be read,
%      and the header must come before the data.
%
%      See also LOAD, SAVE, SPCONVERT, FSCANF, FPRINTF, STR2MAT.
%      See also the IOFUN directory.

% check number and type of arguments
if nargin < 1
    error('Function requires one input argument');
elseif ~isstr(file)
    error('Input must be a string representing a filename');
end

% Open the file.  If this returns a -1, we did not open the file
% successfully.
fid = fopen(file);
if fid==-1
    error('File not found or permission denied');
end

% Initialize loop variables
% We store the number of lines in the header, and the maximum
% length of any one line in the header.  These are used later
% in assigning the 'header' output variable.
no_lines = 0;
max_line = 0;

% We also store the number of columns in the data we read.  This
% way we can compute the size of the output based on the number
% of columns and the total number of data points.
ncols = 0;

% Finally, we initialize the data to [].
data = [];

% Start processing.
line = fgetl(fid);
if ~isstr(line)
    disp('Warning: file contains no header and no data')
end;
[data, ncols, errmsg, nxtindex] = sscanf(line, '%f');

% One slight problem, pointed out by Peter vanderWal: If the
% first character of the line is 'e', then this will scan as
% 0.00e+00. We can trap this case specifically by using the
% 'next index' output: in the case of a stripped 'e' the next
% index is one, indicating zero characters read.  See the help

```

```

% entry for 'sscanf' for more information on this output
% parameter. We loop through the file one line at a time until
% we find some data. After that point we stop checking for
% header information. This part of the program takes most of the
% processing time, because fgetl is relatively slow (compared to
% fscanf, which we will use later).
while isempty(data) | (nxtindex==1)
    no_lines = no_lines+1;
    max_line = max([max_line, length(line)]);
    % Create unique variable to hold this line of text information.
    % Store the last-read line in this variable.
    eval(['line', num2str(no_lines), '=line;']);
    line = fgetl(fid);
    if ~isstr(line)
        disp('Warning: file contains no data')
        break
    end;
    [data, ncols, errmsg, nxtindex] = sscanf(line, '%f');
end % while

% Now that we have read in the first line of data, we can skip
% the processing that stores header information, and just read
% in the rest of the data.
data = [data; fscanf(fid, '%f')];
fclose(fid);

% Create header output from line information. The number of lines
% and the maximum line length are stored explicitly, and each
% line is stored in a unique variable using the 'eval' statement
% within the loop. Note that, if we knew a priori that the
% headers were 10 lines or less, we could use the STR2MAT
% function and save some work. First, initialize the header to an
% array of spaces.
header = setstr(' '*ones(no_lines, max_line));
for i = 1:no_lines
    varname = ['line' num2str(i)];
    % Note that we only assign this line variable to a subset of
    % this row of the header array. We thus ensure that the matrix
    % sizes in the assignment are equal. We also consider blank
    % header lines using the following IF statement.
    if eval(['length(' varname ')~=0'])
        eval(['header(i, 1:length(' varname ')) = ' varname ';'']);
    end
end % for

% Resize output data, based on the number of columns (as returned
% from the sscanf of the first line of data) and the total number
% of data elements. Since the data was read in row-wise, and
% MATLAB stores data in columnwise format, we have to reverse the
% size arguments and then transpose the data. If we read in
% irregularly spaced data, then the division we are about to do
% will not work. Therefore, we will trap the error with an EVAL
% call; if the reshape fails, we will just return the data as is.

```

```
eval('data = reshape(data, ncols, length(data)/ncols)'';', ' ');
end
```

## Appendix 7.2

### Random Walk Model Matlab Code

```
Coded by Nigel Reuel on 8.3.2011
%
% Number of average of runs
Navg = 10;
%
% ETA to measure
neta = 50;
%ETA = 2;
Log_ETA = linspace(.1,3,neta); % log10(Viscosity) reported in cP
ETA = zeros(1,neta);
for i = 1:neta
    ETA(1,i) = 10^(Log_ETA(1,i));
end
% Lengths to measure
%nL = 100;
%L = linspace(450,2000,nL); % Lengths reported in nm
nL = 1;
L = 900;
% Parameters for trace generation
FullTime = 10; % seconds.
ns = 10000;
ts = FullTime/ns;
nT = 100;
% Experimental setup parameter - orbital average:
Oavg = 0.032; % seconds
%Oavg = ts;
% Initialize a Length matrix for translational and rotational measurements:
%MSE_LT = zeros(nT,neta);
%MSE_LR = zeros(nT,neta);
% Loop through each of the ETA and L combinations to show how well the
% setup will measure the length:
for i = 1:neta
    for k = 1:Navg
        tic
        for j = 1:nL

            eta = ETA(i);
            length = L(j);
            %Generate the 'real' traces
            T = feval(@GenTrace,length,ns,ts,eta,nT);
            % Now determine what the 3D particle tracker sees:
            Texp = feval(@TrackerTrace,T,FullTime,length,nT,ns,Oavg);
            % Calculate the length from the experimental values:
            [Lt Lr Lr_cos] = feval(@SWNTTraceAnal_v2,Texp,nT,Oavg,eta);
            % Now record the mean square error of the measurement at this
length and
            % viscosity:
```

```

MSE_LT(k,1) = ((length-mean(Lt))./length);
MSE_LR(k,1) = ((length-mean(Lr))./length);
MSE_LR2(k,1) = ((length-mean(Lr_cos))./length);

end
MSE_LT_n(i,1) = (mean(MSE_LT));
MSE_LR_n(i,1) = (mean(MSE_LR));
MSE_LR2_n(i,1) = (mean(MSE_LR2));
toc
k

end
end
%
X = ETA';
figure
plot(X,MSE_LT_n)
hold on
plot(X,MSE_LR_n, 'g')
hold off
csvwrite('MSE_LT_n.csv',MSE_LT_n)
csvwrite('MSE_LR_n.csv',MSE_LR_n)
csvwrite('MSE_LR_cos2scale_n.csv',MSE_LR2_n)

end

function Texp = TrackerTrace(T,FullTime,L,nT,ns,Oavg)
% Specify the orbital averaging time of the current setup:
% Oavg = 0.032; % seconds
% Determine what the setup would see during its orbit and introduce error
% associated with length of the nanotube:
Nexpstep = floor(FullTime/Oavg);
StepSpan = floor(ns/Nexpstep);
Texp = zeros(Nexpstep,5,nT);
for i = 1:nT
    for j = 1:Nexpstep
        % What are the averaging indices of T for this time interval:
        sindex = (j-1)*(StepSpan)+1;
        eindex = j*StepSpan;
        % Average the 'real' trace as the experimental setup would and
        % introduce measurement error in the X,Y,Z position due to the
        % nanotube length
        %
        % Let's start by saying the nanotube length induces a random
        % displacement error in one direction (along length of the nanotube)
        % and the magnitude is gaussian around the 1/2 length of the tube.
        %
        % Adjust the X,Y,Z values according to this scheme
        Tadj = T(1:ns,1:3,i);
        %
        for k = 1:ns
            Lhalf = L/2;
            R = (rand()*2-1); % Random number from (-1 to 1) centered around
0
            % assign the direction (1 = x, 2 = y, 3 = z)
            Error = R*Lhalf;
            rd = ceil(rand()*3);

```

```

        Tadj(k,rd) = Tadj(k,rd)+Error;
    end
    %
    %
    % Now solve for the new values
    error Texp(j,1,i) = mean(Tadj(sindex:eindex,1)); % AVG X coordinate +
    error Texp(j,2,i) = mean(Tadj(sindex:eindex,2)); % AVG Y coordinate +
    error Texp(j,3,i) = mean(Tadj(sindex:eindex,3)); % AVG Z coordinate +
    %
    %
    % Texp(j,4,i) = mean(T(sindex:eindex,4,i)); % AVG angle
    % This is the angle value you would get w/ cos2 scaling
    TempTimeAngles = T(sindex:eindex,4,i);
    AnglesCosScale = cos(TempTimeAngles).^2;
    MeanFluor = mean(AnglesCosScale);
    Texp(j,5,i) = acos((MeanFluor)^(1/2));
    % This is the angle value that you would get w/ linear scaling:
    Texp(j,4,i) = (MeanFluor-1)*-pi()/2;

    end
end
% Return the experimental traces to the main function
return
end

function [Lt Lr Lr_cos] = SWNTTraceAnal_v2(Texp,nT,Oavg,eta)
% Coded on 7.18.2011 by Nigel Reuel (altered and inserted in new program on
% 8.4.2011 by Nigel Reuel)
%
% Outer loop refers to trace number...
%
Lt = zeros(nT,1);
Lr = zeros(nT,1);
Lr_cos = zeros(nT,1);
for iii = 1:nT
%
% Assign the X,Y,Z, I1 values
X = Texp(:,1,iii);
Y = Texp(:,2,iii);
Z = Texp(:,3,iii);
I1 = Texp(:,4,iii);
I2 = Texp(:,5,iii);
%
st = 1;
et = length(X);
% Determine the MSD of the trace
Nsteps = (et-st);
MSD = zeros(Nsteps,1);
MSD_rot = zeros(Nsteps,1);
MSD_rot2 = zeros(Nsteps,1);
% Scale of the pixels (already in nm in the case of generated traces):

```

```

scaleX = 1; % nm/pixel
scaleY = 1; % nm/pixel
scaleZ = 1; % nm/pixel
for i = 1:Nsteps
    SD = 0;
    SRD = 0;
    SRD2 = 0;
    for j = 1:et-st+1-i
        delX = ((X(j+i,1)-X(j,1))*scaleX);
        delY = ((Y(j+i,1)-Y(j,1))*scaleY);
        delZ = ((Z(j+i,1)-Z(j,1))*scaleZ);
        delI1 = ((I1(j+i,1)-I1(j,1)));
        delI2 = ((I2(j+i,1)-I2(j,1)));
        SD = SD + delX^2 + delY^2 + delZ^2;
        SRD = SRD + delI1^2;
        SRD2 = SRD2 + delI2^2;
    end
    MSD(i,1) = SD/(et-st+1-i);
    MSD_rot(i,1) = SRD/(et-st+1-i);
    MSD_rot2(i,1) = SRD2/(et-st+1-i);
end
%Portion of the MSD plot to fid
Por = .1;
xp = (1:round(Nsteps*Por))';
yp = MSD(1:round(Nsteps*Por),1);
xpr = (1:round(Nsteps*Por))';
ypr = MSD_rot(1:round(Nsteps*Por),1);
xpr2 = (1:round(Nsteps*Por))';
ypr2 = MSD_rot2(1:round(Nsteps*Por),1);
%plot((1:Nsteps)',MSD)
p = polyfit(xp,yp,1);
p2 = polyfit(xpr,ypr,1);
p3 = polyfit(xpr2,ypr2,1);
%p1 = slope and p2 = intercept
% Must correct for time scale:
ts = Oavg; % seconds per averaging orbit(s)
D = p(1)/ts/6; % nm^2/s <-- For 3D diffusion
Dr = p2(1)/ts/2; % rad^2/s
Dr2 = p3(1)/ts/2; % rad^2/s
%
% Now determine the length of the nanotube from Translation:
d = 0.76; %nm - diameter for (6,5) Nanotubes
Lo = 1000; %nm - initial guess for length of Nanotube
options=optimset('MaxFunEvals',1*10^10,'MaxIter',1*10^10,'TolFun',0.000000000
1);
Lt_solve = fsolve(@SWNTLength,Lo,options,D,d,eta);
%
% Now determine the length of the nanotube from rotation:
options=optimset('MaxFunEvals',1*10^10,'MaxIter',1*10^10,'TolFun',0.000000000
1);
Lr_solve = fsolve(@SWNTLength_rot,Lo,options,Dr,d,eta);
%
% Now determine the length of the nanotube from rotation (with Cos^2
scaling):
options=optimset('MaxFunEvals',1*10^10,'MaxIter',1*10^10,'TolFun',0.000000000
1);
Lr_solve2 = fsolve(@SWNTLength_rot,Lo,options,Dr2,d,eta);

```

```

% Record the solutions:
% Record the solutions:
%
Lt(iii,1) = Lt_solve;
Lr(iii,1) = Lr_solve;
Lr_cos(iii,1) = Lr_solve2;
%
end
return
end

function F = SWNTLength(L,D,d,eta)
% Solve function to determine length of nanotube (all inputs in units of nm):
kB = 1.3806503*10^(-23); %m2 kg s-2 K-1
T = 298; % K
% Drag = 8*pi()*(eta/1000)*(L*10^-9)/(3*log(L/d)+2*(-0.114)+0.886);
F = D - kB*T/(8*pi()*(eta/1000)*(L*10^-9)/(3*log(L/d)+2*(-0.114)+0.886))*(10^9)^2; % Both quantities should have units of m^2/s
return
end

function F = SWNTLength_rot(L,D,d,eta)
% Solve function to determine length of nanotube (all inputs in units of nm):
kB = 1.3806503*10^(-23); %m2 kg s-2 K-1
T = 298; % K
% Drag = 8*pi()*(eta/1000)*(L*10^-9)/(3*log(L/d)+2*(-0.114)+0.886);
F = D - kB*T/(pi()*(eta/1000)*(L^3/(10^9)^3)/(3*(log(L/d) - 0.447))); % Both
quantities should have units of m^2/s
return
end

function Tr = GenTrace(L,ns,ts,eta,nT)
% Coded by Nigel Reuel on 8.2.2011
% This function generates the 3D trajectory of a nanotube based on the
% given input parameters:
% L = length (nm)
% ns = number of time steps (integer)
% ts = time scale (s)
% eta = viscosity (cp)
% nT = number of traces (integer)
%
% Output trace trajectory: T(ns,4,nT) --> c1:X,c2:Y,c3:Z, and c4:current
rotation
% angle
%
% Translational Diffusion
%
kB = 1.3806503*10^(-23); %m2 kg s-2 K-1
T = 298; % K
d = 0.76; % nm
Dt = kB*T/(8*pi()*(eta/1000)*(L*10^-9)/(3*log(L/d)+2*(-0.114)+0.886))*10^18;
% nm^2/s
%
% Rotational Diffusion
%
```



```

Dr = kB*T/(pi()* (eta/1000)*(L^3/(10^9)^3)/(3*(log(L/d) - 0.447))); % rad^2/s
%
% Solve for the MSD (unidirectional) and MSR:
MSD = Dt*ts*2;
MSR = Dr*ts*2;
%
% Initialize the matrix to receive the trajectory coordinates:
%
Tr = zeros(ns,4,nT);
%
for i = 1:nT
    % Initialize the SWNT location at 0,0,0,0
    X = 0;
    Y = 0;
    Z = 0;
    rad = 0;
    for j = 2:ns
        % This was the original way I constructed the X,Y,Z coordinates.
        % From Aug 2011.
        % This is incorrect (I believe) but I will keep it commented in
        % case I want to look at it for reference.
        %{
        % Generate your random numbers with average of 1 (span 0 to 2)
        r(1) = rand()*2;
        r(2) = rand()*2;
        r(3) = rand()*2;
        r(4) = rand()*2;
        % Generate your random -1 or +1 to determine if the distance
        % traveled was positive or negative
        s = zeros(1,4);
        s(1) = round(rand());
        s(2) = round(rand());
        s(3) = round(rand());
        s(4) = round(rand());
        for k = 1:4
            if s(k) == 0
                s(k) = -1;
            end
        end
        % Determine the displacements:
        delX = (MSD)^(1/2)*r(1)*s(1);
        delY = (MSD)^(1/2)*r(2)*s(2);
        delZ = (MSD)^(1/2)*r(3)*s(3);
        delRad = (MSR)^(1/2)*r(4)*s(4);
        % Determine the new locations:
        X = X + delX;
        Y = Y + delY;
        Z = Z + delZ;
        rad = rad + delRad;
        %}
    %
    % This is the new method for determining X,Y,Z coordinates coded on
    % 10.29.2011
    %
    TotalMSD = MSD*3;
    % Determine which portion of Total MSD goes to X,Y,and Z direction
    r(1,1) = rand()*3;

```

```

r(2,1) = rand()*3;
RandN = sort(r);
X_MSD = (RandN(1,1))/3*TotalMSD;
Y_MSD = (RandN(2,1)-RandN(1,1))/3*TotalMSD;
Z_MSD = (3 - RandN(2,1))/3*TotalMSD;
%
% Generate your random -1 or +1 to determine if the distance
% traveled was positive or negative
s = zeros(1,3);
s(1) = round(rand());
s(2) = round(rand());
s(3) = round(rand());
s(4) = round(rand());
for k = 1:4
    if s(k) == 0
        s(k) = -1;
    end
end
end
% Determine the displacements:
delX = (X_MSD)^(1/2)*s(1);
delY = (Y_MSD)^(1/2)*s(2);
delZ = (Z_MSD)^(1/2)*s(3);
delRad = (MSR)^(1/2)*s(4);
% Determine the new locations:
X = X + delX;
Y = Y + delY;
Z = Z + delZ;
rad = rad + delRad;
%
% Record the X,Y,Z, locations and the absolute rotation angle
Tr(j,1,i) = X;
Tr(j,2,i) = Y;
Tr(j,3,i) = Z;
Tr(j,4,i) = abs(rem(rad,pi()/2));
end
end
return
end

```

## Appendix 7.3

Matlab Code for analyzing SWNT length from MSD Plots

```

function SWNTTraceAnal_v2
% Coded on 7.18.2011 by Nigel Reuel
% This code analyzes the significant trace segments (found by the
% 'SWNTTracePrep' program).
%
% Analysis steps:
% 1) Creates MSD plot and fits it linearly
% 2) Determines the translational diffusion coefficient
% 3) Determines the lengths of the nanotubes
% 4) Evaluates the mean intensity
% 5) Evaluates the stdev of intensity
% 6) Saves 3-5 for post analysis

```

```

%
% Look in the data folder and determine the number of significant traces:
Nt = 75;
% Create matrix to save the data:
Data = zeros(Nt,6);
% Loop through the files to do the analysis:
for iii = 1:Nt

M = csvread(['SigT_',int2str(iii),'_ForAnalyzer.csv']);
% Initial visualization
X = M(:,1);
Y = M(:,2);
Z = M(:,3);
I1 = M(:,4);
I2 = M(:,5);
% Normalize the I1 and I2 vectors to be between 0 and pi()/2
I1max = max(I1);
I1min = min(I1);
I2max = max(I2);
I2min = min(I2);
I1n = acos(((I1-I1min)./(I1max-I1min)).^(1/2));
I2n = acos(((I2-I2min)./(I2max-I2min)).^(1/2));
%plot3(X,Y,Z)
st = 1;
et = length(X);
% Determine the MSD of the trace
Nsteps = (et-st);
MSD = zeros(Nsteps,1);
MSD_rot = zeros(Nsteps,1);
% Scale of the pixels:
scaleX = 92; % nm/pixel
scaleY = 92; % nm/pixel
scaleZ = 29; % nm/pixel
for i = 1:Nsteps
    SD = 0;
    SRD = 0;
    for j = 1:et-st+1-i
        delX = ((X(j+i,1)-X(j,1))*scaleX);
        delY = ((Y(j+i,1)-Y(j,1))*scaleY);
        delZ = ((Z(j+i,1)-Z(j,1))*scaleZ);
        delI1 = ((I1n(j+i,1)-I1n(j,1)));
        delI2 = ((I2n(j+i,1)-I2n(j,1)));
        SD = SD + delX^2 + delY^2 + delZ^2;
        SRD = SRD + delI1^2 + delI2^2;
    end
    MSD(i,1) = SD/(et-st+1-i);
    MSD_rot(i,1) = SRD/(et-st+1-i);
end
xp = (1:round(Nsteps*.05))';
yp = MSD(1:round(Nsteps*.05),1);
xpr = (1:round(Nsteps*.05))';
ypr = MSD_rot(1:round(Nsteps*.05),1);
%plot((1:Nsteps)',MSD_rot)
p = polyfit(xp,yp,1);
p2 = polyfit(xpr,ypr,1);
%p1 = slope and p2 = intercept
% Must correct for time scale:

```

```

ts = 32*10^-3; % seconds per frame
D = p(1)/ts/6; % nm^2/s <-- For 3D diffusion
Dr = p2(1)/ts/4; % rad^2/s
%
% Now determine the length of the nanotube from Translation:
d = 0.76; %nm - diameter for (6,5) Nanotubes
Lo = 1000; %nm - initial guess for length of Nanotube
options=optimset('MaxFunEvals',1*10^10,'MaxIter',1*10^10,'TolFun',0.000000000
1);
L = fsolve(@SWNTLength,Lo,options,D,d);
%
% Now determine the length of the nanotube from rotation:
options=optimset('MaxFunEvals',1*10^10,'MaxIter',1*10^10,'TolFun',0.000000000
1);
L2 = fsolve(@SWNTLength_rot,Lo,options,Dr,d);
%
% Evaluate the Intensity Traces
Imean1 = mean(M(:,4));
Imean2 = mean(M(:,5));
Isd1 = std(M(:,4));
Isd2 = std(M(:,5));
% Write the data for this trace into the CSV file
Data(iii,1) = L;
Data(iii,2) = L2;
Data(iii,3) = Imean1;
Data(iii,4) = Imean2;
Data(iii,5) = Isd1;
Data(iii,6) = Isd2;
%
% End of outer loop for reading files:
end
%
% Post processing save and graphing tools:
% Save Data:
csvwrite('SWNTData.csv',Data);
%plot(Data(:,1),Data(:,2))

end

```

```

function F = SWNTLength(L,D,d)
% Solve function to determine length of nanotube (all inputs in units of nm):
kB = 1.3806503*10^(-23); %m2 kg s-2 K-1
T = 298; % K
eta = 51; %centipoise
% Drag = 8*pi()*(eta/1000)*(L*10^-9)/(3*log(L/d)+2*(-0.114)+0.886);
F = D - kB*T/(8*pi()*(eta/1000)*(L*10^-9)/(3*log(L/d)+2*(-
0.114)+0.886))*(10^9)^2; % Both quantities should have units of m^2/s
return
end

```

```

function F = SWNTLength_rot(L,D,d)
% Solve function to determine length of nanotube (all inputs in units of nm):
kB = 1.3806503*10^(-23); %m2 kg s-2 K-1
T = 298; % K
eta = 51; %centipoise
% Drag = 8*pi()*(eta/1000)*(L*10^-9)/(3*log(L/d)+2*(-0.114)+0.886);

```

```

F = D - kB*T/(pi()*(eta/1000)*(L^3/(10^9)^3)/(3*(log(L/d) - 0.447))); % Both
quantities should have units of m^2/s
return
end

```

## Appendix 7.4

Code for determining cell volume from a z-stack of widefield images

```

function SWNT_inCell_imageMaker
% Coded by Nigel Reuel on 9.30.2011
% Used to display the SWNT tracks within a z-stack of cell images
%
% Step 0: ....Define scale.....
%
% Read in calibration image to define the following parameters:
% (only has to be done once)----->
%{
Cal = imread('Calibration.tif');
imshow(Cal);
% Select UL, UR, and LR points
[Xcal,Ycal] = ginput(3);
%Save these points so they do not need to be clicked again:
csvwrite('Xcal.csv',Xcal);
csvwrite('Ycal.csv',Ycal);
% -----<
%}
Xcal = csvread('Xcal.csv');
Ycal = csvread('Ycal.csv');
%
% Translate the widefield pixel distance to
FCSdistance = ((Xcal(1,1)-Xcal(2,1))^2 + (Ycal(1,1)-Ycal(2,1))^2)^(1/2);
xscale = 92*FCSdistance/256; % nm/pixel
yscale = 92*FCSdistance/256; % nm/pixel
% Z scale from lab notebook
zscale = 1000; % nm/stack step;
%
% Step 1: ....Define the Cell Space....
%{
% ----- Read in the z-stack images -----
% Number of images (lowest to highest z level)
Ni = 15;
% Read one image to see how big your matrix needs to be:
temp = imread('Cell1_1.jpg');
[Nr Nc] = size(temp);
zstack = zeros(Nr, Nc, Ni);
for i = 1:Ni
    itemp = imread(['Cell1_',int2str(i),'.jpg']);
    idouble = double(itemp);
    zstack(:,:,i) = idouble;
end
% Optional image check:
% contour(zstack(:,:,1))
%
%----- Logic to define wall regions (darkest in the image)-----
%

```

```

% Are nearest neighbors below set threshold?
%
% Threshold:
MeanZ = mean(mean(mean(zstack)));
%Thr = MeanZ-10*std(std(std(zstack)));
Thr = MeanZ*.75;
% Neighbor size:
NS = 1;
%
% Cell wall location matrix
CellLoc = zeros(1,3);
count = 1;
for k = 7:10
    tic
    % Calculate z location
    zmid = round(Ni/2);

    Zpos = (k-zmid)*zscale;
    for i = 1+NS:Nr-NS
        for j = 1+NS:Nc-NS
            if mean(mean(zstack(i-NS:i+NS,j-NS:j+NS,k))) < Thr
                % Mark this spot
                Xpos = i*xscale;
                Ypos = j*yscale;
                CellLoc(count,:) = [Xpos Ypos Zpos];
                count = count+1;
            end
        end
    end
    plot3(CellLoc(:,1),CellLoc(:,2),CellLoc(:,3),'.')
    toc
end
csvwrite('CellLoc2.csv',CellLoc);
%
%}
%
CellLoc = csvread('CellLoc2.csv');
%
plot3(CellLoc(:,1),CellLoc(:,2),CellLoc(:,3),'.')
hold on
%}
%
% Step 2: Insert the significant traces into the volume cell:
%
% Specify number of files and run the trace prep program to extract the
% meaningful traces:
%NumFiles = 18;
%Nt = feval(@SWNTTracePrep,NumFiles);
% Now plot the data!
%
% Loop through the files to do the analysis:
%for iii = [7 5 16]
for iii = 1:18

M = csvread(['SigT_',int2str(iii),'_ForAnalyzer.csv']);
% Initial visualization

```

```

X = M(:,1);
Y = M(:,2);
Z = M(:,3);
I1 = M(:,4);
I2 = M(:,5);
% Normalize the I1 and I2 vectors to be between 0 and pi()/2
I1max = max(I1);
I1min = min(I1);
I2max = max(I2);
I2min = min(I2);
I1n = I1./(I1max-I1min)*pi()/2;
I2n = I2./(I2max-I2min)*pi()/2;
% Scale of the pixels:
scaleXFCS = 92; % nm/pixel
scaleYFCS = 92; % nm/pixel
scaleZFCS = 20; % nm/pixel
% Center point of FCS in widefield
Xcenter = (Xcal(3,1)-Xcal(1,1))*xscale;
Ycenter = (Ycal(3,1)-Ycal(1,1))*yscale;
%
% First attempt, don't worry about the small rotation
%
L = length(I1);
Xcorrect = zeros(L,1);
Ycorrect = zeros(L,1);
Zcorrect = zeros(L,1);
for i = 1:L
    Xcorrect(i,1) = X(i,1)*scaleXFCS+Xcenter;
    Ycorrect(i,1) = Y(i,1)*scaleYFCS+Ycenter;
    Zcorrect(i,1) = Z(i,1)*scaleZFCS;
end
% Plotting code:
% Fancy Plot: Show the trace with the intensity mapped along it:
% W = 3;
C = I1;
Cmin = min(I1);
Cmax = max(I1);
p = feval(@cline,Xcorrect,Ycorrect,Zcorrect,C,Cmin,Cmax);
figure_name_fig = [int2str(111),'.png'];
saveas(gcf,figure_name_fig,'png');
%set(gcf,'Position',[0 0 365 343])
% End of loop file to go through the trace segments:
end
end

```





## 8. Conclusion and Future Outlook

This thesis focuses on the development of label-free sensors for glycans and proteins based on single walled carbon nanotube (SWNT) fluorescence. Impetus for this work came from our growing realization of the integral role glycans play in human health and disease and the lack of convenient tools for glycobiochemists. Existing tools for glycan analysis include many pretreatment steps for liberating and labeling the sugars (Fig 1.4). In Chapter 1 we review many nanoengineered candidates for label and liberation free glycan analytics. Carbon nanotube-based sensors emerge as a worthy candidate for these analytics due to their sensitivity and capability for massive multiplexing. In Chapter 2 we model how such quantitative nanosensors, SWNT or otherwise, could be used as a weak affinity dynamic microarray to augment the sensitivity and profiling capabilities of existing lectin microarrays. This is simulated with a kinetic Monte Carlo model with parameters estimated from the Consortium of Functional Glycomics (CFG) public database of fluorescent microarray data. The model is applied to a few relevant examples (screening biotherapeutics and monitoring clinical biomarkers), but what it really highlights is 1) the usefulness of quantitative nanosensors and 2) the need for much larger libraries of robust kinetic data to make real predictions. Obtaining these massive data sets of kinetic parameters between glycans and lectins (and other carbohydrate recognition domains, such as custom boronic acid polymers) will be an area of ongoing work for these sensors.

Next we detail the ‘proof of concept’ experimental work, showing the ability to transduce glycan-lectin binding events using carbon nanotubes (Chapter 3). This first generation platform consisted of chitosan-wrapped SWNT functionalized with NTA chelation groups that held divalent cations ( $\text{Ni}^{2+}$ ). The nickel provided the signal transduction (quenching when close to the SWNT surface) as well as a convenient docking site for a sensor protein. When the analyte protein binds to the sensor protein, the increased steric hindrance (pulling the sensor from the SWNT surface) increases the SWNT signal. Two glycan-lectin pairs are explored (GlcNAc to GafD and Fuc to PA-III) both as free glycans and as model glycoprotein probes (by biotinylating the sugar and attaching to streptavidin). The sensor signal was collected and integrated from ensembles of nanotubes in chitosan hydrogels. The dynamic response closely resembles the surface plasmon resonance ‘sensograms’ presented in Chapter 1 and are likewise analyzed with a similar kinetic model. The fit provides the forward and reverse kinetic rates and dissociation constant ( $K_D$ ). The sensors were also interrogated individually and we found that a small subset of ‘good responders’ were responsible for the majority of the integrated signal modulation. With this proof of concept established we then embarked on applying the tool to a more practical problem.

There are many useful applications for a glycan sensor (Chapter 1), but the problem we decided to tackle first was that of antibody production (Chapter 4). It is a tractable problem, in that the glycoprotein (IgG) is well understood. The Fc domain of the IgG has two conserved sugars that can change based on a variety of processing parameters. There are a set of glycans that are desirable (terminal sialic acid, terminal galactose, core fucose) as well as glycoforms that are unwanted due to poor pharmacokinetics (hypermannose forms, immunogenic glycans). The first problem to overcome in adapting our first generation sensor to antibody detection was the inherent promiscuity of antibodies; they had a greater tendency for non-specific adsorption. Also, they were much larger than the streptavidin probes used previously and thus a new hydrogel platform was developed. In the end polyacrylamide (PAAm) demonstrated reduction in nonspecific adsorption as well as the ability to tune pore size. This platform was then used to measure IgG titer, KD distributions, and extent of hypermannosylation. Also, local productivity of cell islands could be measured with the spatial array. Future work for this project included making a field portable instrument and making nanotube containing hydrogels more conducive to cell growth. The portable detector was the subject of the next chapter and the latter is an ongoing area of research in our group.

With success of demonstrated lab-based measurements, the next hurdle was performing such measurements in the field with online bioreactors (Chapter 5). The lab-based platforms we had used prior (Chapter 3-4) were too capital intense and were not conducive for portability (large inverted microscope, floated optics stage, laser, liquid nitrogen cooled detector). So we designed a smaller, portable unit that used an LED for excitation and a much smaller, electronically amplified detector (Fig 8.1). We also changed the sensor hydrogel to agarose to allow for a more rapid production protocol (1 day vs. 1 week) and also one that can be automated (all steps done in ambient conditions). The portable detector was demonstrated in our Cambridge, MA lab as well as on the Novartis campus in Basel, Switzerland with many different types of IgG products. Ongoing work for this area includes further improvements in the portable detector optics to increase signal to noise and to improve multiplexing capabilities. Along this vein, we have also designed another detector that allows for multiplexing of a small printed grid of hydrogel spots by raster scanning the excitation source (laser) across a larger surface area (Fig 8.2). This way a larger array of spots can be printed and interrogated at the same time (Fig 8.3). This experimental tool is now being used to print smaller and smaller arrays so that 100-200 independent sensor sites can be monitored simultaneously with less than a second frame rate. This should allow for the long-held goal of multiplexing many different His-tagged sensor proteins (i.e. lectins) and obtaining their kinetic parameters with an analyte protein (i.e. glycoprotein) in a single experiment, thus realizing the desire for larger and larger  $K_D$  matrices to use for our predictive modeling work (Chapter 2).

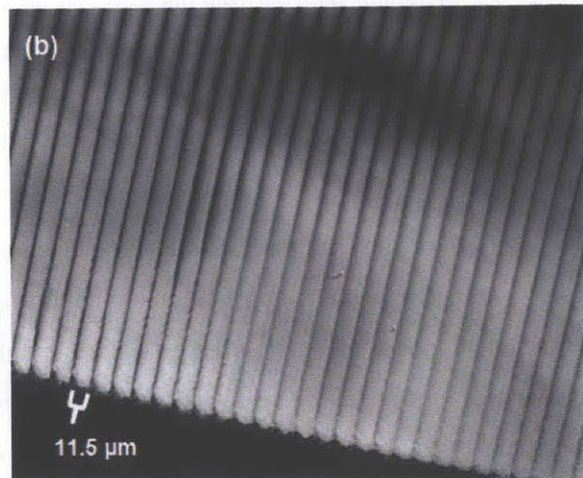
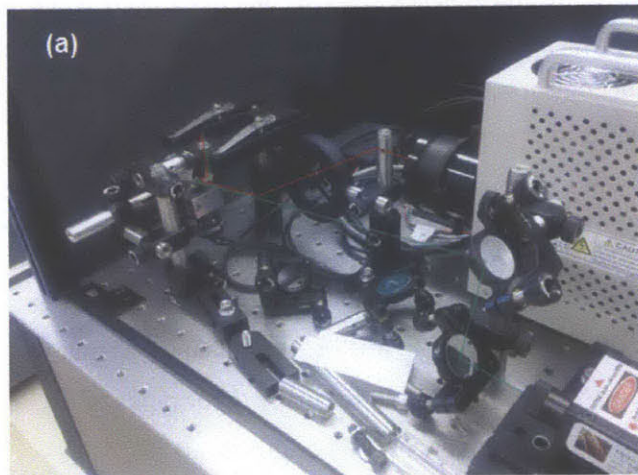


\$150-250k, throughput = 1 sample/hr

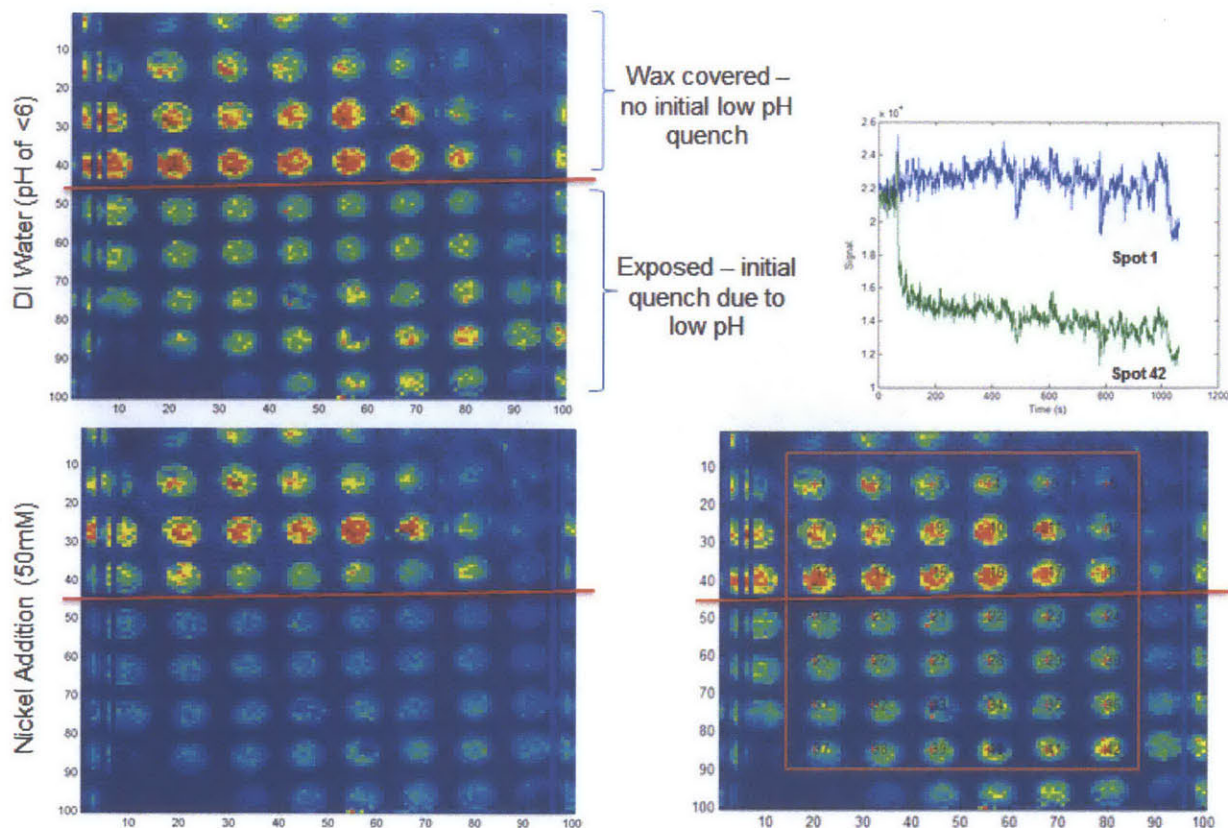


\$5-10k, throughput = 100 samples/hr

**Figure 8.1** – Evolution of experimental setups going from capital intense lab-based systems to portable detectors



**Figure 8.2** – Microarray scanner setup a) Optics – excited by a 561nm laser, galvo mirrors sweep the focused spot on the sample slide up to a 1cm<sup>2</sup> area. b) Close up of sweep pathway as recorded in a photo-bleached dye showing a 11.5μm spot size.



**Figure 8.3** - Example data set from microarray scanner of a 7x6 grid of sensor spots. a) Half of the array is covered by melted parafilm (initially brighter because the low pH water partially quenches the exposed SWNT – lower half) b) Upon nickel addition the exposed sensor sites reduce in brightness. c) Algorithms have been written in Matlab to extract the average signal from each spot, presented are intensity traces from two – spot 1 that was protected from the nickel quench and spot 42 that was exposed.

In the course of this research there were multiple times that large data sets of single nanotubes were analyzed (Chapter 3). These single nanotube signal traces contained embedded event states obscured by noise. We developed an efficient state finding algorithm (NoRSE – see Chapter 6) that we used in this work to evaluate  $K_D$  of a large body of SWNT. It can be used for many other types of experiments and has since been adopted by a few other experimenters outside of this field.

Nanotube sensors immobilized in hydrogels (chitosan, PAAm, and agarose) have been the dominant focus of this work, but there are many future applications where it is desirable for the SWNT sensor to be a free-floating probe. These could be *in vivo* probes for a specific glycosylated biomarker or to adhere to cells with a unique glycocalyx (such as cancer cells). Thus this thesis also explored the ability to track nanotubes *in vivo* using an advanced particle tracking microscope customized for SWNT sensing (Chapter 7). The study revealed the fundamentals of SWNT rotational movement as tracked with

polarized excitation and correlated these motions to solution viscosity. Nanotubes were tracked in living cells, demonstrating normal, confined, and convective diffusion. The SWNT sensors could also monitor the internal viscosity of the cells. In future work the microscopes orbital speed should be increased to improve the ability to differentiate signal fluctuations caused by SWNT rotation *versus* sensor transduction. This would then allow for interesting studies of SWNT sensors, tracking binding events, while traveling around a cell *in vivo*.

This work has taken an abstract idea of improving glycan analytics *via* the use of quantitative nanosensors. It has focused on fluorescent carbon nanotubes as a good candidate for glycan sensing due to signal stability and capability to multiplex. The sensor was shown to work experimentally with ideal, homogenous glycoprotein standards. The sensor platform was improved for antibody measurement from cell supernatants and real fermentation bioreactors. The platform was then further refined for portable sensing and demonstrated on-site at a major pharmaceutical company. Along the way a useful algorithm for finding binding events in noisy single nanotube sensor responses was created as well as demonstrating the practical limitation of tracking SWNT sensors three-dimensionally *in vivo*. We look forward to the utilization of these sensors in larger and larger arrays (Fig 8.3) to realize the long-standing goal of simple and robust glycoprofiling. Once realized this will have a very large impact in biopharmaceutical manufacturing, clinical diagnostics, and beyond.