Chapter 1

Introduction

1.1 High-Affinity Protein Capture Agents in Medical Diagnostics

A fundamental goal of medical diagnostics is to detect and monitor changes in biomarkers, which are substances used as an indicator of a biological state. Exemplary biomarkers are proteins, genes, mRNA, or small molecules. With the information provided by measurement of biomarkers, the current state of a patient's health can potentially be determined and predictive features can be claimed. Medical diagnostics, and in particular cancer diagnostics, is increasingly requiring measurements of large panels of biomarkers based on the complex and heterogeneous molecular composition of diseased tissues and organs. Such a multi-parameter approach, namely simultaneously measuring as many different biomarkers as possible in a single experiment, should improve the accuracy and efficiency of diagnostic assays. Through the measurement of a collection of biomarkers, multi-parameter diagnostics have the potential to offer unique molecular signatures, or fingerprints, of a patient's health status and a high level of sensitivity and specificity for diagnosing, staging, monitoring treatments over time, and predicting future disease.[1]

Genetic (DNA) and transcriptomic (mRNA) biomarker panels are already employed in the clinic on a routine basis, but technologies enabling the routine implementation of protein biomarker panels have lagged behind. This is quite unfortunate, as protein biomarker measurements are perhaps the most informative clinically. However, they are also by far the most expensive, in terms of cost per biomarker. In addition, the majority of the approximately 20,000 proteins in the human proteome are post-translationally modified at some stage in their existence, and such modifications can often change the basic function of the protein.[2] These modifications (e.g., glycosylation, phosphorylation) can only be detected by directly detecting the

modified protein. Furthermore, temporal changes in post-translational modifications, such as evolving glycosylation patterns on a given protein, have been implicated as indicators of disease stage.[3]

The dominant clinical technologies for detecting protein biomarkers are antibody based and, in fact, the gold standard protein assays, and the only ones that are highly reproducible from clinic to clinic and across geographical locations, require two antibodies per protein detected. These are sandwich assays, or enzyme-linked immunosorbent assays (ELISAs).[4] The cost and instability of antibodies generally prohibit the measurement of more than a handful of proteins in a single assay, and the cost per protein is about $50. Nevertheless, ultimately one would like to routinely assess the levels of hundreds or more proteins for disease diagnosis, or monitor a few proteins at high frequency. This will require inexpensive protein capture agents that possess the affinities and specificities of antibodies, but also exhibit chemical, biochemical, and physical stability. A technology for the rapid and scalable production of such capture agents would revolutionize disease diagnostics. It would also significantly impact benchtop research, providing the realization of quantitative and highly multiplexed assays that can replace the pauci-parameter protein measurement approaches (e.g., Western blots) that are standard today.

Non-antibody protein capture agents have been pursued for several years. The chemical nature of such capture agents is typically limited to nucleic acids, peptides, and small molecules, but a capture agent can also incorporate lipids, carbohydrates, and even other proteins. Nucleic acid aptamers[5] hold promise, but possess the intrinsic limitation of chemical diversity, as there are only 4 standard nucleobases, as compared to the 20 natural amino acids from which proteins are constructed. Other issues, such as nuclease

resistance and synthetic scale-up, comprise additional hurdles in the widespread applicability of nucleic acid aptamers. On the other hand, peptides selected from phage display libraries[6] can offer reasonable to excellent performance. However, the L-amino acids comprising such peptides are sensitive to proteolytic cleavage. Chemical stability and water solubility can be an additional limitation as they are highly sequence dependent.

A third alternative is peptide affinity agents that are identified using one-bead-one-compound (OBOC) libraries.[7] This chemical library-based approach allows for the inclusion of broad classes of amino acids, including artificial and non-natural amino acids, along with peptide mimetics.[8] This diverse chemical flexibility can be harnessed to infer attributes including biochemical, chemical, and physical stability, and water solubility. However, compromises have to be reconciled between peptide length and library diversity, since OBOC libraries of oligopeptides are practically only 10^5–10^6 elements in size.[7] In addition, even a small OBOC library of polypeptides (or polypeptide mimetics) can be challenging to build, since the synthetic purity of an on-bead peptide correlates with peptide length, and very-high purity libraries are required for affinity screening. As a result, OBOC libraries have rarely been employed for the identification of high-affinity, high-specificity protein capture agents.

Small molecule ligands can exhibit a high affinity for their protein targets, but selectivity is limited since they only sample a small part of the protein.[9] One small-molecule method that is relevant to the work of this thesis is that of in situ click chemistry,[10] which was originally developed by K. B. Sharpless and M. G. Finn. Their goal was to identify small molecule enzymatic inhibitors that could be catalytically assembled using the scaffold of the protein target itself. Some of these studies started

with a known small molecule inhibitor that was then divided into two components, each of which was expanded into a small library of building blocks. One library contained molecules functionalized with an azide group, and the other library contained molecules functionalized with an acetylene group. During the screening of the target protein against the molecular libraries, the protein plays an active role in the selection and covalent assembly of a new inhibitor. In these systems, the protein accelerates the Huisgen 1,3-dipolar "click" cycloaddition by holding elements from each library in close proximity. The protein exhibits exquisite selectivity; it only promotes the formation of a 1,2,3-triazole between those library elements that can be brought into precise molecular proximity on the protein surface. The result is a biligand inhibitor with an affinity that approaches the product of the affinities of the individual molecular components. Furthermore, the triazole itself can contribute to the binding affinity observed for this inhibitor.

In Chapters 2 and 3 of this thesis, I will discuss how the chemical flexibility of comprehensive, OBOC libraries of oligopeptides may be combined with *iterative* in situ click chemistry to select a high-affinity, high-specificity triligand capture agent against the protein biomarker carbonic anhydrase II (CA II), for both the human and bovine varieties ($K_D \approx 45$ and 64 nM, respectively). Furthermore, this triligand capture agent can be used in a dot blot assay to detect those proteins at the ≥20 ng level from 10% porcine serum. Results from Western blots, sandwich (ELISA-like) assays, and protein activity assays, with the triligand implemented as the primary capture agent, are presented in Chapter 3.

The triligand is built from peptides comprised of non-natural and artificial amino acids, including amino acids containing azido and acetylene functionalities. For this

selection scheme, the OBOC method was utilized first to identify an anchor (1°) ligand for CA II which contained a terminal acetylene-containing amino acid. This screen resulted in a 7-mer peptide that binds CA II with $K_D \approx 500$ μM, which is a suitable affinity value for further maturation. Then, the protein target was utilized to template the covalent coupling between two peptide ligands, the pre-identified 1° ligand and a secondary (2°) ligand, which was selected by the protein target and the 1° ligand from a comprehensive OBOC library of 2° ligands displaying a terminal azide-containing amino acid. This in situ click chemistry screen resulted in a biligand that binds CA II with $K_D \approx 3$ μM. After modifying the biligand with a terminal acetylene-containing amino acid, this capture agent became the new anchor for selection of a 3° ligand. A final protein-templated in situ click chemistry screen between the biligand anchor and a comprehensive OBOC library of 3° ligands (azides) resulted in the triligand capture agent. Interestingly, the triligand does not bind to the enzymatically active binding site of CA II—a result that argues for the generality of this approach.

This *iterative* in situ click chemistry approach has several significant advantages over both traditional in situ click chemistry[10] and traditional OBOC peptide libraries[7] for affinity agent screening. These include: (1) Production of the capture agent requires no prior knowledge of affinity agents against the protein of interest, but can potentially take advantage of such ligands if they exist. (2) The approach permits the sampling of a very large chemical space. (3) The process can be repeated to produce tetraligands, pentaligands, and other higher-order multi-ligands with an accompanying increase in affinity and specificity from cooperative interactions. (4) The approach may be harnessed to produce branched capture agents, thus providing low molecular weight capture agents that mimic the 3-D folded structures of antibodies or polypeptides.

(5) The capture agents can be designed, *ab initio*, to contain desirable features such as chemical, biochemical, and thermal stability, water solubility, fluorophore conjugation, and ability for highly oriented attachment to a substrate or surface in a monoparameter or multiparameter assay. (6) The final capture agents may be prepared in gram quantities and stored as a powder under ambient conditions. Chapters 2 and 3 have been taken in part from *Angewandte Chemie International Edition* **2009**, *48*, 4944–4948 (see also Appendix A).

Protein-templated in situ click chemistry is a low-yielding reaction requiring precise alignment of the azide and alkyne with respect to each other and the protein. Therefore, only a small fraction (<<1%) of the peptides on a particular bead will be converted to multi-ligands. In Chapter 4, both colorimetric and quantitative polymerase chain reaction (QPCR)-based methods for detection and quantitation of the formation of on-bead, protein-catalyzed multi-ligand capture agent will be discussed. The low but detectable yield per protein-catalyzed in situ click reaction—approximately 0.000005% for bCAII—confirms the exquisite demands of the process. This result encouraged us to develop more sophisticated screening strategies that incorporated anti-selections (following the selections) and also direct detection of the bead-bound products of the protein-catalyzed click reaction. In other words, we developed screens that identified the protein target, secondary screens that identified the in situ click product, and even tertiary screens that identified potential side-reactions. These new screening strategies were applied toward the selection of a biligand capture agent ($K_D \approx 140$ nM) against the blood-based cancer biomarker prostate-specific antigen (PSA). The rapid assembly of the biligand capture agent by the protein-catalyzed process was expedited to two weeks by utilization of a previously reported anchor ligand[11] and the new selection/anti-

selection strategies, and demonstrates the potential feasibility of a high throughput route toward production of high-affinity, high-specificity protein capture agents.

1.2 Assay Platforms for Multi-Parameter Protein Measurements

While the high-affinity protein capture agents of Chapters 2 to 4 represent a significant technology advance, they are just one component of what is necessary for highly multiplexed measurements of protein biomarkers. In addition, it is also important to develop or optimize the actual assay platforms that can enable sensitive multi-parameter protein measurements using these capture agents. There are a number of drawbacks associated with the existing gold-standard approaches. As mentioned above, the gold standard for protein diagnostic assays are ELISA assays, and the standard clinical procedures are to extract a few milliliters of blood from a patient, centrifuge that blood to separate plasma (or serum) from whole blood, and then carry out ELISA assays for one or two proteins in 96-well plate format under diffusion-limited conditions.

One drawback of this approach involves the stability of the antibodies utilized within the ELISA assays. ELISA assays require at least two antibody capture agents for detection of the protein biomarker—a monoclonal surface-immobilized antibody for protein capture, and a secondary enzyme-linked polyclonal antibody which binds to a second epitope on the protein. Binding of the secondary antibody is visualized by applying a colorimetric substrate which, for example, changes color or yields a fluorescence signal in the presence of enzyme. In Chapter 3, I will describe how peptide multi-ligands, identified by in situ click chemistry, show feasibility as capture agents in ELISA and other standard biological assays such as Western blots. Using multi-ligand capture agents instead of antibodies in these platforms avoids problems often associated

with antibody use—namely high cost, poor stability, and subtle variations in performance (e.g., sensitivity).

A second drawback of the current gold-standard clinical approach is that it is slow. During the time between blood draw and assay completion (typically a few hours to a few days), the biospecimen may degrade, so that the measured protein levels no longer reflect the patient status at the time of the blood draw. In addition, the few milliliters of blood that are drawn make it easier to handle the blood, but, in principle, the same protein assays could be accomplished with only a few microliters of plasma or serum (and thus, with a significantly reduced amount of patient discomfort).

While multi-ligand capture agents avoid the inherent problems of antibody instabilities (and potentially antibody costs), they do not change the inherently large sample volume, lengthy assay time, or number of measurement parameters per assay. The use of microfluidics to miniaturize and expedite protein assays can solve many of these problems.[1a] Other technologies, such as label-free nanoelectronic sensors, can provide further advantages. Silicon nanowire (SiNW) nanoelectronic sensors[12] can provide quantitative multi-parameter measurements from nanoliter to microliter volumes of protein biomarkers in real time. The "label-free" characteristic of these sensors means that no secondary antibodies are required to detect the binding between the protein of interest and capture agent. SiNWs fabricated by the SNAP technique[13] represent ultra-dense arrays of electronically addressable nanowires, where each wire may be functionalized with a different protein capture agent. When the protein of interest specifically binds to the capture agent, both the electrical conductance of the nanowire and the electrical capacitance between the nanowire and the surrounding solution is altered. These electrical changes may be directly correlated to the absolute

amount of protein in the solution. However, SiNW sensors can be challenging to deploy. For example, working with Si surfaces can be challenging because unprotected Si forms a native oxide (SiO_2), and this insulating layer can significantly reduce the detection sensitivity of the nanowire sensors via dielectric shielding. In addition, the native oxide on silicon also has a low isoelectric point, meaning that under physiological conditions (= pH 7.4), SiO_2 surfaces are negatively charged.[14] These surface charges can potentially limit the sensitivity of silicon nanowire field effect biosensors through Debye screening at the sensor surface.[15] Finally, the native oxide layer contains electrical defect sites at the Si-SiO_2 interface.[16] For high surface area devices, such as SiNWs, this phenomenon can reduce charge carrier mobilities significantly.[16,17]

In Chapter 5, a general method for the non-oxidative functionalization of single-crystal silicon (111) is described. To prevent the formation of this oxide, the silicon (111) surface was modified with an acetylene (-C≡C-H) monolayer of ~100% surface coverage. An electroactive monolayer of a benzoquinone-masked primary amine was subsequently formed on the acetylene-passivated surface via Cu(I)-catalyzed Huisgen 1,3-dipolar cycloaddition ("click" chemistry). Molecules presenting a carboxylic acid group were finally immobilized onto regions where the benzoquinone had been reduced and cleaved to reveal the underlying amine on the surface. This strategy provides a general platform to incorporate most organic and biological molecules, such as proteins, antibodies, or multi-ligand capture agents, on highly passivated silicon (111) surfaces with minimal surface oxidation. This work can be further extended toward the non-oxidative biopassivation of silicon-on-insulator (SOI) wafers, whose topmost 30–50 nm single-crystal silicon layer is the substrate in the fabrication of SiNW sensors. Chapter 5

has been taken in part from the *Journal of the American Chemical Society* **2006**, *128*, 9518–9525 (see also Appendix D).

1.3 REFERENCES

1. (a) Fan, R.; Vermesh, O.; Srivastava, A.; Yen, B. K. H.; Qin, L.; Ahmad, H.; Kwong, G. A.; Liu, C.-C.; Gould, J.; Hood, L.; Heath, J. R. *Nat. Biotechnol.* **2008**, *26*, 1373–1378. (b) Hood, L.; Heath, J. R.; Phelps, M. E.; Lin, B. *Science* **2004**, *306*, 640–643. (c) Phelan, M. L.; Nock, S. *Proteomics* **2003**, *3*, 2123–2134.
2. Larsen, M. R. *Methods Mol. Biol.* **2003**, *251*, 245–262.
3. Dudkin, V. Y.; Miller, J. S.; Dudkina, A. S.; Antczak, C.; Scheinberg, D. A.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2008**, *130*, 13598–13607.
4. Engvall, E.; Perlmann, P. *J. Immunol.* **1972**, *109*, 129–135.
5. (a) Cox, J. C.; Hayhurst, A.; Hesselberth, J.; Bayer, T. S.; Georgiou, G.; Ellington, A. D. *Nucleic Acids Res.* **2002**, *30*, e108–e108. (b) Lee, J. F.; Hesselberth, J. R.; Meyers, L. A.; Ellington, A. D. *Nucleic Acids Res.* **2004**, *32*, D95–D100. (c) Famulok, M.; Mayer, G.; Blind, M. *Acc. Chem. Res.* **2000**, *33*, 591–599. (d) Gold, L.; Polisky, B.; Uhlenbeck, O.; Yarus, M. *Annu. Rev. Biochem.* **1995**, *64*, 1094–1110. (e) Burmeister, P. E.; Lewis, S. D.; Silva, R. F.; Preiss, J. R.; Horwitz, L. R.; Pendergrast, P. S.; McCauley, T. G.; Kurz, J. C.; Epstein, D. M.; Wilson, C.; Keefe, A. D. *Chem. Biol.* **2005**, *12*, 25–33. (f) Proske, D.; Blank, M.; Buhmann, R.; Resch, A. *Appl. Microbiol. Biotechnol.* **2005**, *69*, 367–374. (g) Thiel, K. *Nat. Biotechnol.* **2004**, *22*, 649–651.
6. Smith, G. P.; Petrenko, V. A. *Chem. Rev.* **1997**, *97*, 391–410.
7. Lam, K. S.; Lebl, M.; Krchňák, V. *Chem. Rev.* **1997**, *97*, 411–448.
8. (a) Lam, K. S.; Lebl, M.; Krchňák, V.; Wade, S.; Abdul-Latif, F.; Ferguson, R.; Cuzzocrea, C.; Wertman, K. *Gene* **1993**, *137*, 13–16. (b) Liu, R.; Marik, J.; Lam, K. S. *J. Am. Chem. Soc.* **2002**, *124*, 7678–7680. (c) Alluri, P. G.; Reddy, M. M.;

Bachhawat-Sikder, K.; Olivos, H. J.; Kodadek, T. *J. Am. Chem. Soc.* **2003**, *125*, 13995–14004. (d) Tornøe, C. W.; Sanderson, S. J.; Mottram, J. C.; Coombs, G. H.; Meldal, M. *J. Comb. Chem.* **2004**, *6*, 312–324.

9. Fabian, M. A.; Biggs III, W. H.; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lélias, J.-M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. *Nat. Biotechnol.* **2005**, *23*, 329–336.

10. (a) Lewis, W. G.; Green, L. G.; Grynszpan, F.; Radić, Z.; Carlier, P. R.; Taylor, P.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 1053–1057. (b) Manetsch, R.; Krasiński, A.; Radić, Z.; Raushel, J.; Taylor, P.; Sharpless, K. B.; Kolb, H. C. *J. Am. Chem. Soc.* **2004**, *126*, 12809–12818. (c) Bourne, Y.; Kolb, H. C.; Radić, Z.; Sharpless, K. B.; Taylor, P.; Marchot, P. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1449–1454. (d) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021. (e) Whiting, M.; Muldoon, J.; Lin, Y.-C.; Silverman, S. M.; Lindstrom, W.; Olson, A. J.; Kolb, H. C.; Finn, M. G.; Sharpless, K. B.; Elder, J. H.; Fokin, V. V. *Angew. Chem. Int. Ed.* **2006**, *45*, 1435–1439.

11. (a) Wu, P.; Leinonen, J.; Koivunen, E.; Lankinen, H.; Stenman, U.-H. *Eur. J. Biochem.* **2000**, *267*, 6212–6220. (b) Pakkala, M.; Jylhäranta, A.; Wu, P.; Leinonen, J.; Stenman, U. H.; Santa, H.; Vepsäläinen, J.; Peräkylä, M.; Närvänen, A. *J. Pept. Sci.* **2004**, *10*, 439–447. (c) Koistinen, H.; Närvänen, A.; Pakkala, P.;

Hekim, C.; Aaltonen, J.; Zhu, L.; Laakkonen, P.; Stenman, U.-H. *Biol. Chem.* **2008**, *389*, 633–642.

12. (a) Bunimovich, Y. L.; Shin, Y.-S.; Yeo, W.-S.; Amori, M.; Kwong, G. A.; Heath, J. R. *J. Am. Chem. Soc.* **2006**, *128*, 16323–16331. (b) Zheng, G.; Patolsky, F.; Cui, Y.; Wang, W. U.; Lieber, C. M. *Nat. Biotechnol.* **2005**, *23*, 1294–1301.
13. Melosh, N.; Boukai, A.; Diana, F.; Gerardot, B.; Badolato, A.; Petroff, P.; Heath, J. R. *Science* **2003**, *300*, 112–115.
14. Hu, K.; Fan, F.-R. F.; Bard, A. J.; Hillier, A. C. *J. Phys. Chem. B* **1997**, *101*, 8298–8303.
15. (a) Lud, S. Q.; Nikolaides, M. G.; Haase, I.; Fischer, M.; Bausch, A. R. *ChemPhysChem* **2006**, *7*, 379–384. (b) Neff, P. A.; Wunderlich, B. K.; Lud, S. Q.; Bausch, A. R. *Phys. Status Solidi A* **2006**, *203*, 3417–3423.
16. Buriak, J. M. *Chem. Rev.* **2002**, *102*, 1271–1308.
17. Israelachvili, J. *Intermolecular and Surface Forces* (London: Academic Press, 1985).