12-2003

Development of hydroxyl free radical chemistry for the surface mapping of proteins

Joshua Shane Sharp

To the Graduate Council:

I am submitting herewith a dissertation written by Joshua Shane Sharp entitled "Development of hydroxyl free radical chemistry for the surface mapping of proteins." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Jeffrey Becker, Major Professor

We have read this dissertation and recommend its acceptance:
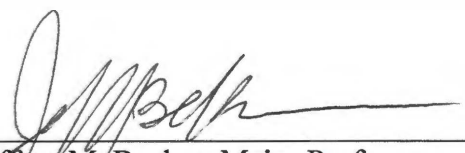
Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Joshua Shane Sharp entitled "Development of Hydroxyl Free Radical Chemistry for the Surface Mapping of Proteins." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Jeffrey M. Becker, Major Professor

We have read this dissertation
and recommend its acceptance:

John W. Koontz

Robert L. Hettich

Accepted for the Council

Vice Provost and Dean of Graduate Studies

DEVELOPMENT OF HYDROXYL FREE RADICAL CHEMISTRY FOR THE SURFACE MAPPING OF PROTEINS

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Joshua Shane Sharp
December 2003

ii

DEDICATION

This dissertation is dedicated to my wife, Shannon Sharp, and our daughter, Chloe Anne Sharp. As much satisfaction as the work presented here has brought to my life, it pales in comparison to the happiness that my family has given me. Without the loving care of my wife, all of my scientific work to this point may not have been completed. Without the boundless promise of my daughter, all of the work I could perform in the future would be without meaning.

ACKNOWLEDGEMENTS

ABSTRACT

In order to rapidly probe the structure of proteins in solution, a protocol for chemical modification of solvent accessible amino acid side chains was developed, and the sites of modification were determined by mass spectrometry to describe the surface of the protein. The knowledge gained about side chain solvent accessibility allowed the critical evaluation of structural models for the proteins examined, allowing incorrect models to be rejected and more likely models to be proposed. Methods were developed using either Fenton chemistry or photolysis of hydrogen peroxide to generate hydroxyl radicals *in situ*. The oxidation chemistry of these radicals with the side chains of various amino acids were exploited to label solvent accessible sites on several model proteins of known tertiary structure. The relative apparent rate of oxidation of the side chains was shown to be a function of the known solvent accessibility and the chemical reactivity of the amino acid. The known properties of hydroxyl radical oxidation of amino acid side chains allows hydroxyl radical surface mapping data to be used as biophysical constraints for evaluating structural models of proteins and protein-protein interactions.

Computational models of the yeast ribonucleotide reductase inhibitor protein Sml1p were evaluated using surface mapping data of the functional C14S Sml1p protein. Various full atom computational models were discredited based on the surface mapping data, and a manually adjusted computational model was generated that possessed low free energy, and agreed with surface mapping data, partial NMR data, and tryptophan anisotropy and quenching data. In addition, the interaction between peptides forming $A\beta$

fibrils, implicated in Alzheimer's disease, were examined using hydroxyl radicals. This radical mapping suggests that the model proffered by Perutz *et al*, which states that the Aβ fibril is a solvent-filled nanotube, is incorrect. Overall, chemically-generated hydroxyl radicals have been developed as a general, multi-target labeling reagent for protein surfaces. Hydroxyl radical surface maps can be used to characterize protein tertiary and quaternary structure and to apply valuable biophysical constraints for structural modeling.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

LC-MS	Liquid chromatography-mass spectrometry
NMR	Nuclear magnetic resonance
XRC	X-ray crystallography
EDTA	Ethylenediaminetetraacetic acid
ES-FTMS or	Electrospray-Fourier transform ion
ES-FTICR-MS	cyclotron resonance mass spectrometry
QIT	Quadropole ion trap
SORI-CAD	Sustained off-resonance irradiation-collisionally activated dissociation
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
HPLC	High performance liquid chromatography
MS/MS	Tandem mass spectrometry
CAD	Collisionally activated dissociation
PBS	Phosphate buffered saline
DTT	Dithiothreitol
PDB	Protein Data Bank
BSA	Bovine serum albumin
TFA	Trifluoroacetic acid
HFIP	3,3,3,3',3',3'-hexafluoro-2-propanol

CHAPTER 1—INTRODUCTION

The era of high-throughput biology began with the publication of the microbial genomes of *Haemophilus influenzae* (1), *Mycoplasma genitalium* (2), the eukaryote *Saccharomyces cerevisiae* (3), and the archaeon *Methanococcus jannaeschii* (4). Since these reports, there has been an explosion in discovery-based, high-throughput biology. Comparative genomics has led to insights in evolution, gene prediction, and gene regulation sequences (5-7). These insights have been applied to many different problems, including the elucidation of antimicrobial targets (8, 9), elucidation of host-symbiont interactions (10), and the nature of drug-efflux systems (11). Comparative genomics has even allowed researchers to begin to define the minimal set of genes required to sustain independent life (12, 13). Genome sequencing today has been greatly automated to run scores of experiments in parallel, with massive computing resources dedicated to the assembly of contigs from all of the automated sequencing results, leading to incredibly high-throughput DNA sequencing.

The amount of information afforded by high-throughput biological analyses increased greatly with the development of high-density microarray hybridization analysis (14). High-density DNA microarrays allow for the analysis of DNA homology on a genomic scale (15-18). Perhaps more importantly for high-throughput biology is the ability to rapidly compare gene expression profiles between different growth conditions. For example, expression differences between stationary and logarithmic growth phases in *E. coli* were analyzed by microarray analysis (19), as was the differences between aerobic

1

and anaerobic growth of *B. subtilis* (20), even analysis of transcription in the brewer's yeast *S. cerevisiae* during different stages of the brewing process (21). This technology allows the parallel comparative expression analysis of thousands of genes in a single experiment, resulting in very high-throughput analysis of transcriptional regulation.

More recently, the advent of high-throughput proteomics has been realized, with the great advances in mass spectrometric instrumentation and the coupling of mass spectrometry with various separations techniques allowing the rapid identification of proteins from complex mixtures (22, 23). Not only has proteomics allowed the identification of a large part of the protein complement of many organisms and tissue types, but also the relative quantitation of protein abundances between different growth states (24, 25), environmental stimuli (26), or tissue differentiation stage (27). A major key to the tremendous increase in throughput and coverage in mass spectrometry is the development of low flow rate LC-MS (22). The result of these technologies is the ability to determine hundreds of protein identities (and potentially relative quantities) in a single experiment.

Thus far, the advances in high-throughput biology have followed the central dogma of molecular biology as described above (Figure 1). As the complexity of the information sought increased from genes to transcripts to protein sequence to protein structure, the analytical requirements, and thus the complexity of the analytical techniques have increased as well. The next logical step is the high-throughput characterization of protein structure. Protein structure is widely recognized to be intimately linked to protein function. Knowledge of protein structure often allows great
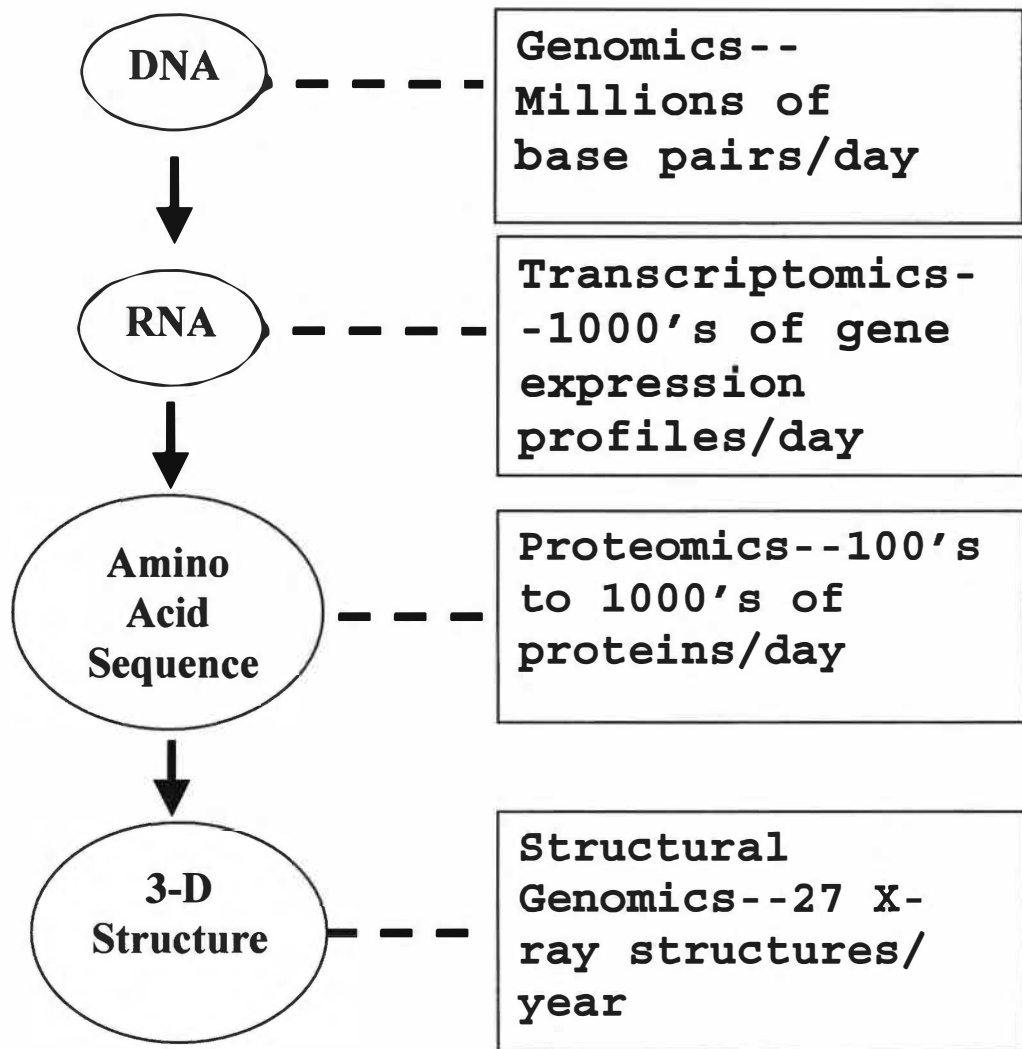
Figure 1: Progress of High-Throughput Biology

A comparison of the standard rates of progress in a single research group for high-throughput biology in systems of increasing complexity. The rate for structural genomics taken from a recent review (28).

insight into the mechanisms of various aspects of protein enzymatic function and ligand binding. For example, Figure 2 shows the structure of bovine β-lactoglobulin A at pH 6.2 (Figure 2A) and pH 8.2 (Figure 2B), two sides of the characterized Tanford transition (29). High-resolution structures allowed the characterization of the conformational change of the pH-sensitive loop covering the hydrophobic binding pocket (30). Numerous other examples of the value of protein structure in elucidation of protein function have been reported. Various initiatives have been proposed to enable the rapid elucidation of protein structures. The major comprehensive initiatives focus on the analysis of targeted proteins to represent a "fold family", by which other proteins can be modeled (31). Current initiatives have focused on NMR and X-ray crystallography to provide high-resolution structures of these targeted proteins (32). Both NMR-based and XRC-based approaches have recently reported a success rate of about 20% from gene to structure (33, 34), resulting in 80% of targeted genes yielding no structure by high-throughput methods. Estimates suggest that the structures of about 16,000 new, carefully selected proteins must be solved in order to accurately model the vast majority of protein structures. If the current rate of ten non-redundant structure solutions per week continues, sufficient data to produce accurate models for the vast majority of proteins will be possible in just over 30 years (31).

There are several factors leading to the low throughput and low success rate for structural genomics by XRC and NMR. The main factor is the need to purify large quantities of the protein of interest. Protein purification is often a very time-consuming task, and not all proteins are amenable to overexpression and purification. In addition, solution-phase NMR has limitations as to the size of proteins that can be analyzed, with a

4
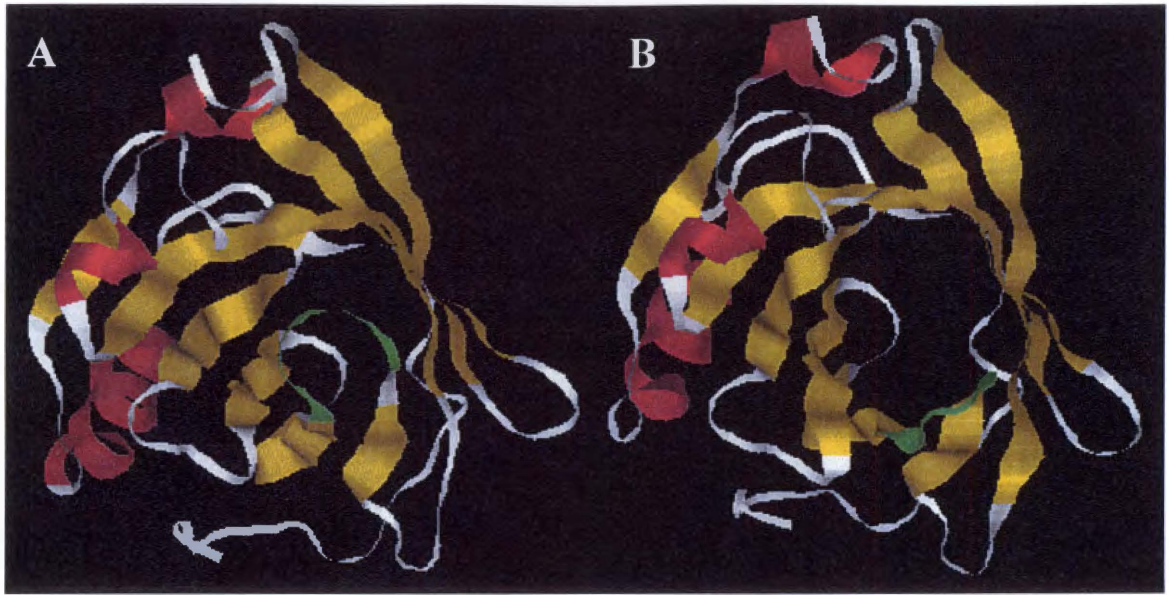
Figure 2: The Tanford Transition of β-lactoglobulin A

The value of protein structure to elucidate the mechanism of biological function.
A.-β-lactoglobulin A at pH 6.2. Loop 85-90 is colored green, and is covering the
hydrophobic pocket between the two β sheets. **B.**- β-lactoglobulin A at pH 8.2, after the
Tanford transition. Loop 85-90 (containing Glu89) moves outward, exposing the
hydrophobic pocket.

hard size limitation at about 40 kDa, and a practical limitation for rapid *de novo* structural analysis of 20 kDa (32). While XRC does not have this size limitation, the technique does require a high-quality crystal of the protein of interest, and many proteins do not readily crystallize. With the amount of protein identifications possible by proteomics approaching thousands per experiment, and the low success rate and low throughput of structural genomics initiatives, an alternative technique that can combine true high-throughput analysis (preferably without the need for purification) with reliable structural information would be an invaluable addition to the toolbox of high-throughput biology.

Every structural genomics initiative to date has included as an essential part the technique of computational structural prediction. The high speed of computational prediction as compared with any experimental technique almost ensures the continued reliance of structural genomics initiatives on computational modeling. Computational modeling algorithms are divided into several groups based on the methods used to generate a structural model. For the purposes of this discussion, two groups of modeling algorithms will be considered. The first group heavily utilizes previously solved protein structures to predict the structure of similar proteins. This group includes protein threading (35), fold recognition and comparative modeling (36) algorithms. With current modeling technologies, reasonably accurate structural models can be generated for proteins that share >30% sequence identity with a protein of known structure (37, 38). Of organisms with completed genomes, reasonably accurate structural models could be generated for only 5-10% of the genes sequenced, according to a report made in 2001 (31). These models generally focus on backbone alignments only, with side chain configurations often disregarded.

6

The other group of modeling algorithms is often referred to as *ab initio* modeling. These modeling techniques attempt to generate protein structures without reliance on homology with known structures, using force field equations and energy minimization techniques to generate protein structures from primary sequences. In recent competitions in protein modeling known as Critical Assessment in Structure Prediction (CASP), the Rosetta algorithm by Dr. David Baker's group has consistently outperformed other algorithms in the *ab initio* category, and is approaching levels of confidence and accuracy previously reserved for comparative modeling/threading techniques (39). In addition, protein comparative modeling is unsuited to modeling truly novel folds, while *ab initio* modeling techniques enjoy much better (although still very limited) success.

Protein structure modeling techniques have been shown to have greatly improved accuracy and reliability when biophysical constraints are applied to the modeling process. Several studies have been published utilizing sparse NMR datasets to constrain protein modeling (40, 41), as well as known active site geometries and disulfide bonds (42). Since biophysical constraints greatly improve the accuracy of computational models, a high-throughput method for generating reliable biophysical constraints for proteins would be a great aid in structural genomics, allowing for reasonably accurate models to be generated for a large number of proteins in a short amount of time. As opposed to threading techniques that tend to rely upon backbone alignments, Rosetta relies heavily upon full atom representations to improve the accuracy and reliability of the modeling process (43). Therefore, unlike algorithms that only model backbone positions, previous knowledge of the biophysical properties of amino acid *side chains* could be utilized to evaluate full-atom representations generated from models provided by Rosetta.

7

One biophysical property of proteins that can be measured experimentally is solvent accessibility. Several methods for determination of solvent accessible regions of a soluble protein have been utilized previously. The study of the binding of monoclonal antibodies has been used in several analyses (44-47). One allows monoclonal antibodies to bind to the epitope of interest, and then observes changes in the binding affinities under different conditions. A reduced binding affinity is interpreted as a reduced accessibility of the epitope. This method requires the availability of monoclonal antibodies for each epitope that one wishes to study, making this approach unsuitable for high-throughput analyses. Also, the large size of the antibody and the large size of the epitope to which the antibody binds reduce the resolution of the surface map; the change in antibody accessibility can occur anywhere within the epitope, and changes in structure that preclude antibody binding may or may not eliminate solvent accessibility.

Another popular method utilized involves incomplete digestion of the protein by a panel of proteases and measuring which sites are susceptible to proteolysis (48-63). One partially digests the protein with the panel of proteases and then tries to determine which potential cleavage sites are acted upon. However, there are many limitations with this approach. First, the investigator must be very careful that only one cleavage per protein molecule occurs. If more than one cleavage occurs, it is very possible that the protein structure can change significantly, preventing further cleavages from being representative of the native structure. A thorough sampling of a protein requires a battery of experiments using different proteases for each protein to be probed, reducing the throughput. Also, the number of reliable proteases limits the sites that can be probed.

8

Finally, proteases are large enzymes, and thus do not accurately represent the solvent accessibility of the sites they act upon.

Specific chemical modification of solvent accessible side chains is another method of protein surface characterization that has been utilized often in the literature (51, 64-74). These methods each allow for the modification of a single class of functional groups on amino acid side chains, such as the free amines of lysine or the sulfhydryl of cysteine. However, although these methods are quite powerful, they only probe a very limited range of amino acids, requiring batteries of different experiments, reagents, and reaction conditions to probe a significant fraction of the amino acids of a protein. A nonspecific protein label would avoid the need for multiple experiments, increasing the speed and reliability of the data acquisition process as well as increasing reproducibility by minimizing the number of different reaction conditions required.

Nonspecific hydrophilic photolabeling experiments were performed previously in order to tag surface proteins in erythrocytes (75), study the denaturation of ribonuclease A (76), discern the topography of rhodopsin in the membrane (77), and measure pH-induced conformation changes in receptor proteins (78). However, in these studies, the detection of the site of labeling was limited to simply detecting which proteins were labeled by autoradiography, protease digestion followed by autoradiography, or acid hydrolysis followed by analysis of differences in amino acid composition. These detection methods have inherently poor resolution. Analysis of proteolytic fragments can, at best, only localize the labeling site to within one polypeptide; amino acid composition analysis can only determine approximately how many of each amino acid are labeled, not their location. Also, for detection using these methods, the reaction is

9

required to go to saturation. Saturating the protein with the photolabel makes it possible that the first labeling reactions will cause structural changes that are then reflected in later labeling reactions. In addition, experimentation has shown that the overall efficiency of the photolabeling reaction is low, requiring an excess of label and long irradiation times (personal observation).

Solution phase hydrogen-deuterium exchange is a commonly used technique to study solvent accessible surfaces and hydrogen bonding of proteins (79-91). Hydrogen-deuterium exchange is a powerful technique, which will probe the backbone of any protein residue. Analysis of hydrogen-deuterium exchange data is often very complicated, though, with the effects of hydrogen bonding, solvent accessibility, and complex back-exchange kinetics all contributing to the observed exchange rates (92). As such, any hydrogen-deuterium exchange experiment requires extensive optimization to characterize the back-exchange rates for the protein studied. Therefore, hydrogen-deuterium exchange, while an excellent technique for targeted applications, is unsuitable for high-throughput experimentation.

Another promising method is that of untethered Fe-EDTA cleavage of proteins in the presence of hydrogen peroxide and ascorbate (93-97). This method has been shown to yield nonspecific cleavages and has been used successfully to map protein-protein interactions and protein-DNA interactions. However, there are several limitations with this method. Once again, this method involves cleavage of the protein backbone, and a single cleavage event can drastically alter the structure of the protein. Therefore, one must be very careful to ensure that only a single cleavage event per molecule occurs; otherwise, incorrect structural information may be obtained. Also, the cleavage reaction

results in many modifications to protein side chains before the first cleavage event; these modifications make accurate assignment of the sites of cleavage difficult (98). However, mapping the sites of protein side chain modification under very limited oxidation conditions could be very useful for describing the surface of a protein.

The use of a synchrotron X-ray beamline has been proposed to generate hydroxyl radicals very rapidly in solution. The hydroxyl radicals oxidize certain amino acid side chains, and then mass spectrometry is used to determine the sites of oxidation (99-103). Good correlation was demonstrated in most cases between oxidation and solvent accessibility, and the value of the technique for probing protein folding and unfolding has been demonstrated (99, 103). Unfortunately, there are some practical difficulties with a synchrotron X-ray approach. Standard low-flux or pulse x-ray sources are not ideal for these studies due to potential damage to the analyte protein (102), and access to a synchrotron beamline is not readily available to all researchers. Another method was developed for surface oxidation of proteins using high electrospray needle voltages with oxygen as the nebulizer gas (104). However, this technique invokes questions as to the structure of the protein in the aerosol spray under a high electric field, as well as difficulties in subsequent tryptic digestion.

Matheson *et al* listed a series of criteria for that must be met for a useful probe of protein surfaces (76).

1) The labeling process itself should not alter the surface topography of the protein

2) The presence of the surface labeling compound(s) in solution before labeling should not alter the surface features of the protein

3) It should be demonstrated that residues known to be buried in the native protein are not labeled but can be labeled in the denatured protein

11

4) In order to correctly interpret the labeling patterns, it should be determined to what extent the labeling reagent is nonspecific

The hydroxyl radical meets all of these requirements. This radical has been utilized to label a variety of model proteins, and it has been shown that the radical will not interact with side chains that are buried in the native structure, showing that the process itself does not modify the protein topography, and that the presence of the labeling compounds also does not modify the protein topography (101, 104). In addition, we have shown that residues known to be buried and unmodified in native folded lysozyme were oxidized readily upon UV-induced protein denaturation (discussed in Chapter 4). Finally, the rich history of radiation biology provides an excellent collection of data on the reactivity of side chains with hydroxyl radicals (105). Complete and thorough analyses on the rates of modification for each of the twenty standard amino acids have been performed, describing the reactivities of the amino acids in solution (106). These findings are summarized in Table 1. The hydroxyl radical has been shown to label a variety of amino acids, most notably sulfur-containing amino acids (Cys, Met) and aromatic amino acids (Trp, Tyr, Phe) (100). These amino acids are usually found in the hydrophobic interior of the protein. Therefore, information on a hydrophobic amino acid that is actually on the surface of the protein is highly useful from an *ab initio* modeling perspective; the force field equations attempt to bury the hydrophobic amino acids, so constraints that require the amino acid on the surface eliminates a large number of inaccurate structures.

Table 1 Pseudo-First Order Rate Constant of Free Amino Acid Oxidation.

Side Chain	Rate constant $(M^{-1} s^{-1})$	Side Chain	Rate constant $(M^{-1} s^{-1})$
Ala	7.7×10^7	**Leu**	**1.7×10^9**
Arg	3.5×10^9*	Lys	3.5×10^8
Asn	4.9×10^7	**Met**[1]	**8.9×10^9**
Asp	7.5×10^7	**Phe**	**6.9×10^9**
Cys	**3.4×10^{10}**	**Pro**	**4.8×10^8**
Glu	2.3×10^8	Ser	3.2×10^8
Gln	5.4×10^8	Thr	5.1×10^8
Gly	1.7×10^7	**Trp**	**1.3×10^{10}**
His	**5.0×10^9***	**Tyr**	**1.3×10^{10}**
Ile	**1.8×10^9**	Val	7.6×10^8

Compiled from Buxton *et al* (106). Side chains in bold are those observed previously to have an addition of 16 Da as a result of hydroxyl radical surface mapping (100).

*These amino acids with relatively high rate constants have a complex series of major reaction products that cause the measured rate of addition of 16 Da to be much lower than the previously reported rate of oxidation; therefore, the rate constants of these amino acids are not indicative of the observed relative reactivity in surface mapping experiments.

[1]Methionine is known to be oxidized regardless of solvent accessibility (101).

The need clearly exists for a high throughput method for accurate determination of protein structure, preferably based on actual biophysical data. Current structural techniques generally require prior purification of the protein of interest, inherently limiting their speed as well as limiting the range of proteins which can be analyzed to those which can be successfully purified in a native form. It has also been established that computational modeling methods, which are currently the backbone of high-throughput structural analyses, often yield inaccurate structures, but can be improved through the use of biophysical constraints (40, 42, 107, 108). The hydroxyl radical satisfies all criteria for a valid surface probe set out by Matheson *et al* (76). Previous work has also demonstrated that protein surfaces can be rapidly and accurately probed using hydroxyl radicals, and the sites of oxidation can be accurately and sensitively determined using tandem mass spectrometry (101, 104).

In order to meet the need for rapid generation of biophysical constraints for computational modeling, rapid chemical methods for generating hydroxyl radicals without denaturing proteins were developed. ***We hypothesize that chemically generated hydroxyl radicals would accurately probe the surface of proteins utilizing a simple, rapid reaction to label a variety of solvent accessible side chains, many of which are hydrophobic in nature. Further, we hypothesize that these labeling events can serve as constraints to eliminate faulty computational models.*** The proposed scheme to probe protein surfaces is presented in Figure 3. Proteins are labeled utilizing hydroxyl radicals generated from hydrogen peroxide, either by transition metal-catalyzed Fenton chemistry, or by photolysis of hydrogen peroxide by UV irradiation. The labeled protein is characterized by ES-FTMS to determine the extent of oxidation. The labeled protein is

14
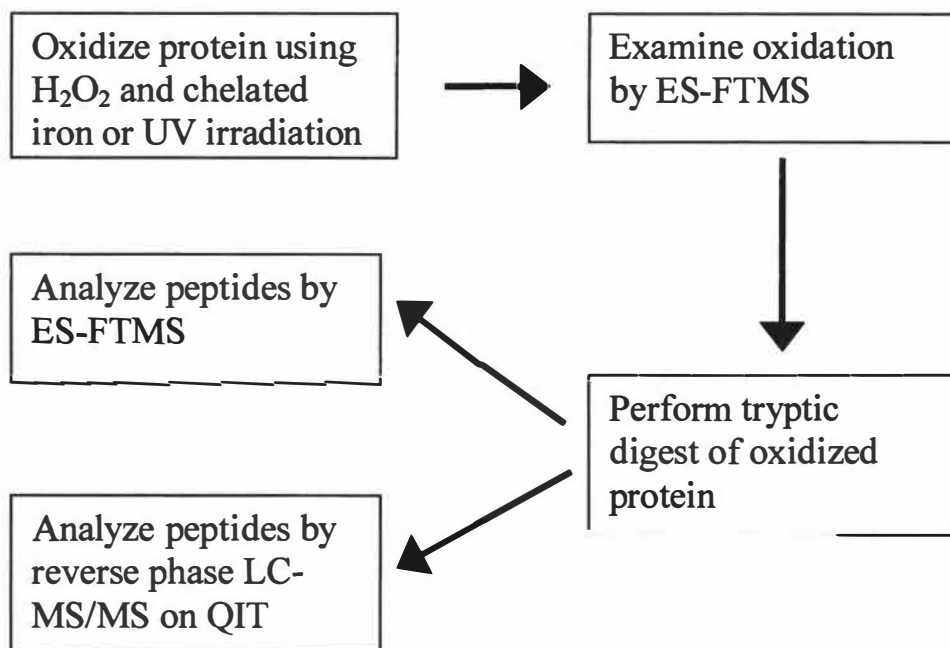
Figure 3: Scheme for Surface Mapping of Proteins

A diagram of the scheme utilized for mapping the solvent accessible surface of proteins

by hydroxyl radical oxidation.

then digested with trypsin, and the resulting peptides are measured by ES-FTMS and by LC-MS/MS using a quadropole ion trap. The labeled amino acid side chains must be accessible to the hydroxyl radical, and therefore to the solvent. Differences in the rate of oxidation will be observed for amino acid side chains with different solvent accessibilities.

The data generated from protein surface mapping will not be sufficient to allow the generation of accurate, high-resolution models. Neither NMR and X-ray crystallography will be supplanted by the successful use of this technique. However, surface mapping has several advantages over these two techniques, as summarized in Table 2. Although the data provided by hydroxyl radical surface mapping is significantly lower resolution, it does not require pure sample, does not have a size limitation, is not as labor intensive, and requires less sample than NMR and XRC. Hydroxyl radical surface mapping was developed as yet another tool in the field of structural biology, especially for rapid characterization of unknown protein samples. Presented herein is work performed to develop a rapid, simple method for utilizing the hydroxyl radical to describe the surface of a protein, and ultimately to use these data as constraints to improve protein structure prediction accuracy.

Table 2 Comparison of Three Structural Biology Techniques

Technique	Sample Amount	Purity	Size Limitation	Solution Phase?	Dynamics?	Resolution
X-ray crystallography	~20-50 mg	≥ 95%	none	no	no	very high (typically ≥1-3 Å)
NMR spectroscopy	~3 mg	≥ 95%	<30-40 kDa	yes*	yes	high (typically ≥ 2-4 Å)
Hydroxyl radical mapping	<.2 mg	Low; usable with simple mixtures	none	yes	no— aggregate average solvent accessibility	Undefined —depends on protein, fairly low

*Solid-state NMR is not yet considered a method of choice for complete structural resolution (109), and thus is not considered here.

CHAPTER 2—MATERIALS AND METHODS

The approach described here is designed to rapidly probe protein surfaces with hydroxyl radicals. The basic strategy is to generate hydroxyl radicals *in situ* with either native or denatured protein of interest. The amount of oxidation is gauged by mass spectrometry to ensure that the oxidation is controlled and limited to a moderate extent. The protein is then digested and the sites of oxidation are determined by mass spectrometry. The sites of oxidation of the denatured protein serve as a control to determine potential oxidation targets, and the native state oxidation determines which side chains are solvent accessible.

Chemical oxidation of protein solutions

Two stock solutions were prepared for the chemical oxidation reactions. Stock solution #1 consisted of 20 mM sodium ascorbate, 50 mM sodium phosphate (Fisher Scientific, Pittsburgh PA), adjusted to pH 6.5 using sodium hydroxide. Stock solution #2 consisted of 6 μM $NH_4Fe(SO_4)_2$ (Fisher Scientific), 13 μM EDTA (Fisher Scientific), 50 mM sodium phosphate (Fisher Scientific), adjusted to pH 6.5 using sodium hydroxide. These stock solutions and the reagent (3% H_2O_2) were used according to a protocol from Heyduk (98) adapted to result in limited side chain oxidation of apomyoglobin. Lyophilized protein (apomyoglobin, Sigma-Aldrich, St. Louis MO) was reconstituted at 200 μM concentration in 50 mM sodium phosphate. For reactions, ascorbate and the iron-EDTA solutions were added in a 1:10 dilution (stock volume : total volume) to the

buffered apomyoglobin. Finally, the H_2O_2 was added to a 0.3% final concentration to initiate the oxidation. In order to quench the reaction, an equal volume of 2M Tris (pH 5) was added.

Photochemical oxidation of protein solutions

All solvents were purchased from Sigma-Aldrich at the highest purity available and used as supplied without further purification, unless otherwise designated. The protein of interest was reconstituted in 1 mL phosphate buffered saline (PBS, 0.01M sodium phosphate, 0.15M sodium chloride, pH 7.5) at a concentration of 0.1 to 2 mg/mL in a 2mL microcentrifuge tube (VWR International, West Chester PA). The protein concentration did not seem to affect the amount of protein oxidation within this concentration range with these proteins. Each sample had 1mL of 30% hydrogen peroxide (Sigma-Aldrich) added to it, for a final hydrogen peroxide concentration of 15% v/v. For denatured C14S Sml1p analyses, guanidine hydrochloride (Sigma-Aldrich) was added to the mixture to a final concentration of 6M. Each sample was then irradiated in the microfuge tube with the cap open in a Stratalinker 2400 (Stratagene, LaJolla CA) at a distance of ~10 cm from the UV source, using the manufacturer's UV light bulbs with maximum illumination at a wavelength of 254 nm for a set amount of time, with the time corrected internally by a UV detector to ensure constant a UV output of ~4000 $\mu watts/cm^2$. The oxidized sample was immediately extracted using a tC2 SepPak (Waters, Milford MA), according to manufacturer's recommendations, to stop the reaction by removing the hydrogen peroxide. The sole exception to this protocol was the analysis of the Aβ 1-40 monomer and fibril. The monomer oxidation was quenched by

19

solid phase extraction using a C18 Lite Sep-Pak (Waters); the oxidation of the fibrils was halted by pelleting, using a 15 minute centrifugation at 4°C in a benchtop microcentrifuge followed by a wash with water. The wash was repeated twice to quench the reaction. An aliquot of 100 µL of intact oxidized protein was retained for ES-FTMS analysis of the extent of oxidation.

Chemistry of hydroxyl radical-mediated oxidation of amino acid side chains

Oxidative damage of proteins has long been an important topic of research, especially damage resulting from ionizing radiation (105) or from accumulation of reactive oxygen species as associated with aging (110, 111). A large body of research is available on the chemistry of radical-mediated oxidation of proteins, both at the side chains and the backbone. Understanding of this chemistry is essential to characterization of the resulting oxidation products.

A potential oxidation site in all peptides and proteins is at the Cα carbon of the protein backbone. Figure 4A shows the reaction mechanism of oxidative cleavage at a protein backbone (112, 113). Oxidation at the protein backbone occurs at a fairly low rate primarily due to the low steric accessibility of the hydrogen of the Cα carbon and the difficulty of the backbone to assume the necessary planar trigonal conformation (114). In the conditions described above for protein oxidation, no main chain degradation was detected.

Oxidation at aliphatic side chains, including Ile, Leu, Val, and Ala, is initiated by abstraction of a hydrogen from a C-H bond, resulting in a side chain carbon radical (Figure 4B). The carbon radical will then interact with molecular oxygen in solution to

20
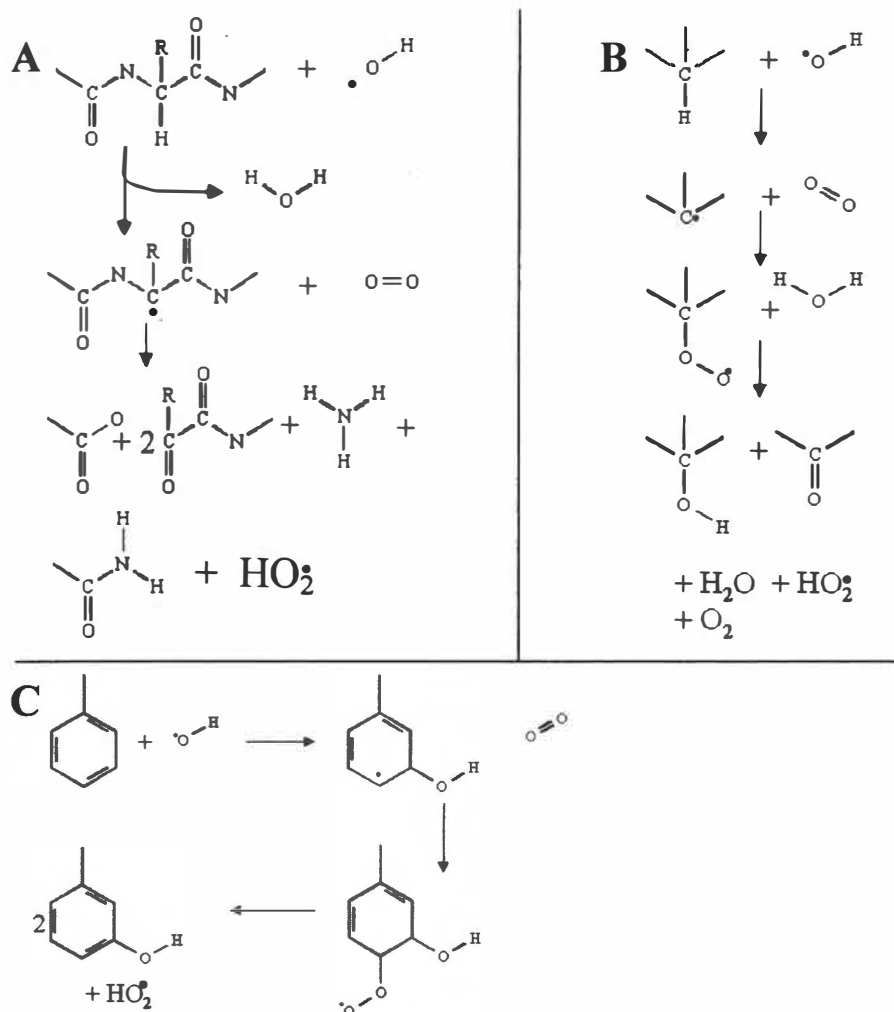
Figure 4: Hydroxyl Radical-Protein Chemistry

Reactions of hydroxyl radicals with various areas of the proteins. **A.** The oxidative cleavage of the protein backbone; the two major pathways both lead to a keto acid, with one resulting in a free amide and the other releasing ammonia and a carboxylic acid. **B.** The oxidation of aliphatic groups. Both hydroxyl and carbonyl substitution are possible pathways, although hydroxyl substitution is much more prevalent. **C.** The oxidation of aromatic rings. The addition to the phenyl group occurs in Phe, Tyr, and Trp, resulting in the production of a hydroperoxyl radical.

result in either a hydroxyl or carbonyl substitution (112, 115, 116). In other amino acids with aliphatic side chains ending in functional groups (such as acid and basic amino acids as well as Asn, Gln, Ser and Thr), hydrogen abstraction in the aliphatic portion of the side chain often results in the loss of the functional group (115, 117). In the absence of oxygen, the carbon radicals will interact with one another and dimerize, resulting in protein crosslinking either through the Cα carbon (118-121) or through the side chain-centered radical (116). All reactions were performed in non-evacuated solutions, and no crosslinking of protein was observed.

Oxidation at unsaturated cyclic side chains does not occur via a hydrogen abstraction mechanism; rather, the hydroxyl radical directly inserts itself into an unsaturated C-C bond system, resulting again in a side chain carbon radical (Figure 4C). In the presence of molecular oxygen, the addition of an oxygen atom occurs via the introduction of a hydroxyl group to the unsaturated system(122-124). In tryptophan, the addition preferentially occurs at the double bond in the indole ring (125, 126). Histidine undergoes addition to the imidazole ring, followed by a complex sequence of side reactions (Figure 5) (127-130). The most abundant of these reactions is the conversion of histidine to asparagine and aspartic acid.

Oxidation at sulfur-containing side chains occurs preferentially at the sulfur atom. In methionine, oxidation occurs though an addition reaction at the sulfur locus to yield methionine sulfone and methionine sulfoxide (131), although a very low yield of oxidative cleavage of the C-S-C linkage has been reported (132). The oxidation of methionine can also occur without direct interaction with a hydroxyl radical (101, 133, 134), perhaps through a radical transfer reaction in which a methionine buried in a

22
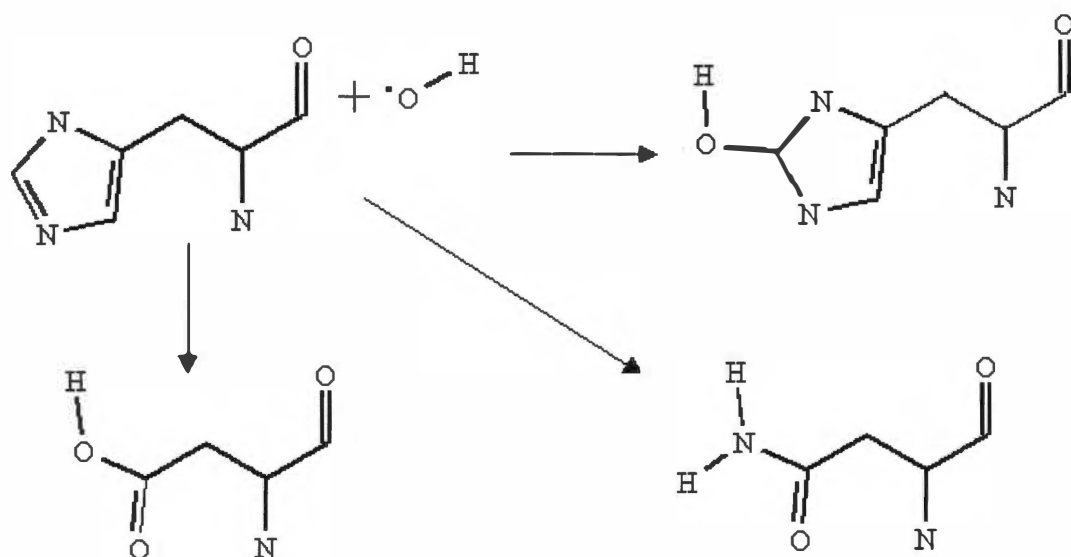
Figure 5: Hydroxyl Radical-Histidine Chemistry

Major reaction pathways of the hydroxyl radical with the histidine side chain. Various

other minor products have also been described.

protein transfers an electron to a carbon radical on the surface of the protein. After this electron transfer, the sulfur-centered radical is oxidized by molecular oxygen (101). Radical-mediated oxidation at free cysteines at neutral pH occurs through abstraction of the hydrogen, followed by interaction with molecular oxygen, yielding final products of sulfuric and sulfonic acids (135, 136). Oxidation at disulfide bonds results in the final formation of both sulfuric and sulfonic acids, with cleavage of the disulfide bond occurring (137-139). A summary of the mass shifts resulting from the most common oxidation products detected are shown in Table 3.

Analysis of oxidized proteins and peptides by ES-FTMS

To analyze the extent of protein oxidation, the oxidized protein was analyzed by ES-FTMS. All ES-FTMS mass spectra were acquired with an IonSpec (Lake Forest, CA) 9.4-Tesla HiRes electrospray Fourier transform ion cyclotron resonance mass spectrometer (Figure 6). Ions were generated by direct infusion with an Analytica electrospray source (~2 μL/min flow rate), desolvated in a heated glass capillary, accumulated in an external hexapole (140), transferred into the high vacuum region with a quadropole lens system (141), and then detected in the cylindrical analyzer cell of the mass spectrometer. Ion detection was achieved in an ultra-low vacuum regime (~2 x 10^{-10} Torr) through the use of multiple stage pumping. Initial pumping was achieved using a mechanical pump, lowering the pressure to the millitorr range. The next stage of pumping was achieved using a turbopump, lowering the pressure to ~10^{-5} Torr. Finally, two cryopumps lowered the base pressure to 1-2 x 10^{-10} Torr. A broadband mass possible. Mass scale calibration was accomplished with ubiquitin. The high-resolution

Table 3 Common Product Mass Shifts From Protein Oxidation

Residue	Mass Shift
Cys	+16, +32, +48
Met	+16, +32
His	+16, -23, -24
Pro	+16, +14
Tyr, Trp, Phe, Ile, Leu	+16
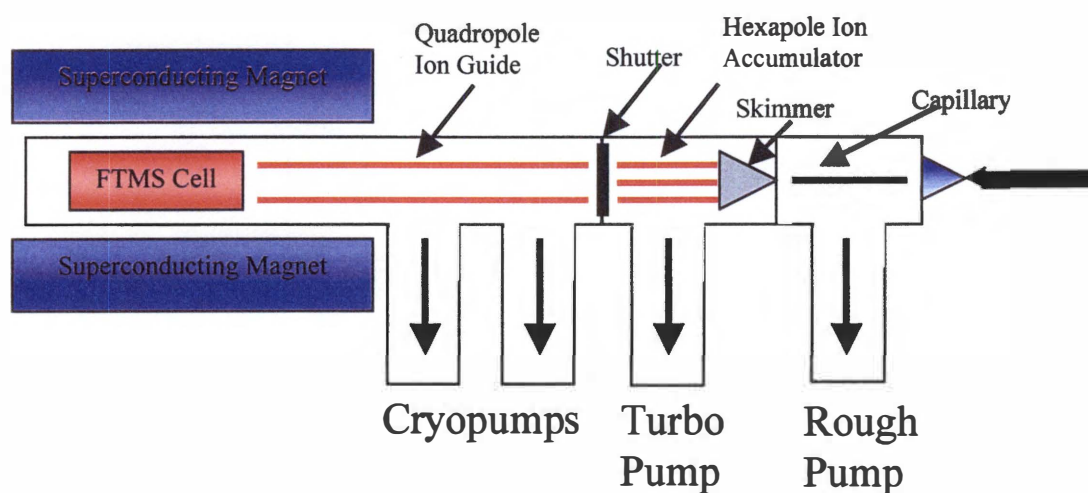
Figure 6: Diagram of IonSpec FT-ICR

Schematic of the IonSpec ES-FTMS instrument. Ions are introduced through the Analytica electrospray ion source and transferred through a heated glass capillary into a mechanically-pumped region, through a skimmer, and into a turbopumped rf-only hexapole for storage at 2×10^{-5} Torr. The ions are then gated through a shutter, down a quadropole ion guide into the Penning cell in a high magnetic field. The Penning cell is pumped to $\sim10^{-10}$ Torr by two cryopumps for low pressure, high resolution analysis.

resolution of about 150,000 (full width of peak at half maximum) at m/z 1000 was mass measurement enables isotopic resolution of multiply charged ions. Thus, the charge state of a multiply-charged ion can be determined solely by its isotopic spacing (142).

Deconvoluted molecular mass spectra were generated with the IonSpec software from the electrospray mass spectra by multiplying the masses of the electrospray ions by their respective charges, and then subtracting the masses of the protons added. This "unfolds" the multiply charged ion mass spectrum into a more easily interpreted molecular mass spectrum. Errors in mass measurement for the multiply charged ions will be scaled proportional to the charge in the calculation of the molecular masses in the deconvoluted mass spectra. By calibrating on the calculated values of the most abundant isotopic peaks of six different charge states (7^+ to 12^+) of bovine ubiquitin, the deconvoluted molecular mass spectrum yielded a measured molecular mass for ubiquitin that was within 0.030 Da of the calculated value. This external calibration procedure generally provides a molecular mass measurement accuracy of ≤ 5 ppm for most proteins up to at least 20 kDa. Measurements of lower mass peptides, such as those obtained from a proteolytic digestion, usually yield molecular mass accuracies in the few millimass range (≤ 1 ppm). The ability to accurately determine the mass/charge ratio of ions as well as the charge state from the isotopic spacing allows the oxidation state of proteins to be followed accurately, as well as the identity of oxidized peptides to be determined by accurate mass measurement.

Ion collisional dissociation was conducted by isolating an ion of interest (either a peptide or protein) within the analyzer cell of the mass spectrometer, and then accelerating the ion into a molecular nitrogen or argon target gas by sustained off-

27

resonance irradiation collisionally activated dissociation, or SORI-CAD (143). For the SORI-CAD experiments, the ion excitation was accomplished with an rf pulse (~1 KHz lower in frequency than the ion's cyclotron frequency) applied for two seconds in duration and with an amplitude in the range of 1-4 volts (peak-to-peak). A pulsed valve was used to admit the argon or nitrogen collision gas into the high vacuum region to a maximum pressure of ~5 x 10^{-6} Torr during the ion excitation step. A base pressure of ~1 x 10^{-9} Torr was re-established prior to ion detection.

Trypsin digestion of oxidized protein

Measurement of the extent of oxidation is insufficient to generate much solvent accessibility data; if the actual sites of modification are determined, they can be deduced to be solvent accessible in the protein. In order to identify the sites of oxidation, the proteins were dried to near completeness in a SpeedVac (Savant Instruments, Holbrook NY) and resuspended in 1mL 8 M guanidine HCl (Sigma-Aldrich), 10 mM DTT (EM Science, Gibbstown NJ), 50 mM Tris buffer (Sigma-Aldrich), and 10 mM calcium chloride (Sigma-Aldrich) pH 7.5. The proteins were heated at 60° C for one hour. They were then diluted with 9 mL of 50 mM Tris buffer, 10 mM calcium chloride, and 20 μg of sequencing grade modified trypsin (Promega, Madison WI) was added to each sample. The tryptic digests were incubated overnight in a 37° C shaker-incubator, and then the peptides were extracted using a C18 SepPak (Waters, Milford MA), and their accurate masses determined by ES-FTMS as described above. The accurate mass measurement allows unambiguous and sensitive detection of oxidized peptides, as well as determining

28

the approximate apparent rate of oxidation for the peptide by comparison of the abundance of the oxidized peptide abundance with the unoxidized version.

Analysis of tryptic peptides by LC-MS/MS

For LC-MS/MS analyses, the C18 SepPak eluant was dried to near completeness in a SpeedVac and resuspended in ~200 μL 0.1% acetic acid. For each tryptic digest, ~50 μL of the resuspended eluant was placed in a sampling vial (Sun International Trading, Ltd., Wilmington NC) for analysis. All samples were injected using a FAMOS autosampler (LC Packings, San Fransisco CA) over a 20 μL low dispersion loop onto a C_{18} (218MS5.315, 300 μm i.d. x 15cm, 300 Å with 3 μm particles) reverse phase column (Vydac, Hesperia CA). HPLC was performed using an Ultimate HPLC system (LC Packings) interfaced to an LCQ-Deca ion trap mass spectrometer (Thermo Finnigan, San Jose CA) equipped with an electrospray source. The sample was washed for five minutes in 100% Buffer A (5% acetonitrile, 95% water, 0.5% formic acid), then the HPLC was run using a linear gradient of 100% Buffer A to 100% Buffer B (70% acetonitrile, 30% water, 0.5% formic acid) over a time of fifty minutes at a flow rate of 4 μL/min, followed by a five minute wash in 100% Buffer B using a plug-in to the Xcalibur software provided by LC Packings. For all C14S Sml1p samples, as well as for all Aβ 1-40 samples, a second gradient was also run to increase sequence coverage. The second gradient was identical, except the linear A to B gradient was run for thirty minutes instead of fifty. The column was connected to the electrospray source with 100 μm I.D.

fused silica. The LCQ-Deca mass spectrometer was run in data-dependent mode with dynamic exclusion activated, and both MS and MS/MS spectra were acquired throughout the LC run. Apomyoglobin tryptic digestion LC-MS runs were examined periodically to ensure column integrity and mass calibration.

The resulting data was analyzed in two steps. The first step employed TurboSEQUEST (144), using a database consisting of the sequence of the protein analyzed. The threshold was set at 50000, with an upper mass limit of 5000, a lower mass limit of 300, a precursor mass tolerance of 1.4 Da, a fragment mass tolerance of 0.5 Da, a peptide mass tolerance of 1.5 Da, a group scan tolerance of 2, a minimum group count of 1, and differential modifications of 48 Da on cysteines (triply oxidized version most commonly observed) and 16 Da on Trp, Met, Tyr, Phe, His, Leu (105), Ile (personal observation), and Ala (solely to aid in identification of other oxidation sites). All masses that were identified as oxidation products were interpreted manually. All MS/MS spectra that were not identified as an unoxidized tryptic peptide were then analyzed manually for unassigned oxidized peptides. As the peptide fragments in the mass spectrometer, cleavage at the amide bond is the prevalent mode of fragmentation. Fragments that are charged on the N-terminal portion of the peptide are b-type ions, and are numbered from the N-terminus of the parent peptide. Fragments that are charged on the C-terminal portion of the peptide are y-type ions, and are numbered from the C-terminus of the parent peptide (145). When an amino acid is oxidized, the difference between the fragment masses equals the amino acid plus the mass of oxygen (or oxygens, in the case of multiply-oxidized residues). Therefore, the site of oxidation can be determined by the neutral mass loss in the tandem mass spectrometry experiments (Figure 7).

30
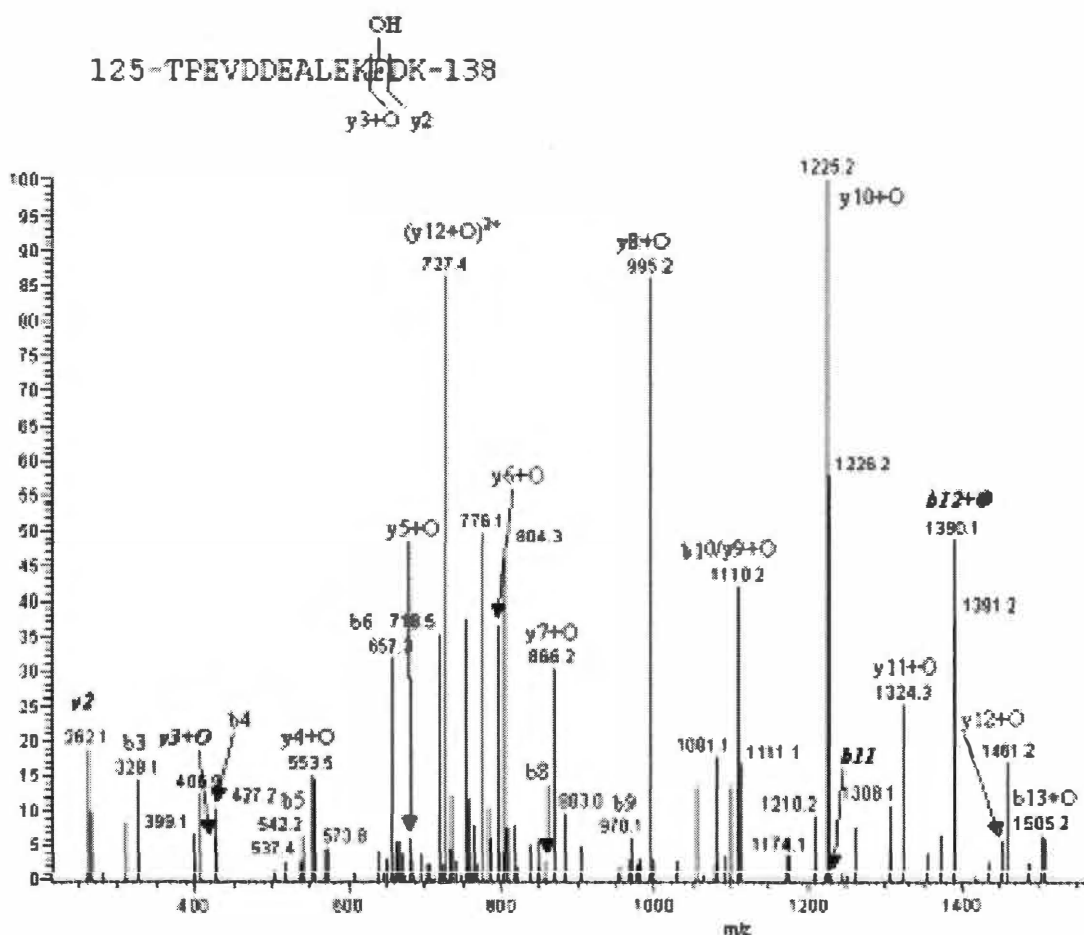
Figure 7: Sample MS/MS Spectrum to Identify Oxidation Site

MS/MS spectrum of peptide 125-138 of bovine β-lactoglobulin A from quadropole ion trap. Peptide fragments were labeled according to the site of cleavage and which of the resulting fragments retained the charge. A shift of 16 daltons more than would be expected between the b11 and b12 ions (and also the y2 and y3 ions) proved the presence of oxygen on F136.

LC-MS/MS utilizing the QIT has several advantages over ES-FTMS that necessitate its use for oxidized peptide characterization. The addition of a chromatography step prior to mass spectrometry decreases the complexity of each spectrum by separating the proteins in time. This separation step increases the dynamic range of the technique; peptides of widely differing abundances can be detected from the same mixture by separating the abundant peptide from the less abundant one. The use of reverse phase chromatography has actually allowed the separation of peptides differing only by a single oxygen atom, allowing the detection of low level oxidation events even in the presence of much greater amounts of the unoxidized peptide. In addition, the LCQ Deca instrument used for the LC-MS/MS runs couples nicely with TurboSEQUEST, the computational tool used to perform the initial screening for oxidized peptides. Finally, CAD on the QIT is more thorough and more robust than SORI-CAD using ES-FTMS. The increase in CAD efficiency allows for high quality fragmentation spectra, which are essential to accurately determine the actual sites of oxidation. The combination of the high accuracy mass measurements from ES-FTMS with the robust fragmentation and increased dynamic range from LC-MS/MS on the QIT allows increased identification and characterization of a complex mixture of peptides.

Analysis of side chain solvent accessibility

The side chain solvent accessibility of a given protein model or structure was calculated using the fifty structures determined by NMR spectroscopy for lysozyme, PDB accession number 1E8L (146); the ten structures determined by NMR spectroscopy for β-lactoglobulin A, PDB accession number 1CJ5 (147); the X-ray crystal structure of horse

heart myoglobin, PDB accession number 1YMB (148) modified as described in Chapter 3; and the various Rosetta and manually generated models for C14S Sml1p as discussed in Chapter 5. Each PDB file was split into individual structure files (when necessary for multiple NMR structures), and the solvent accessible area of each reactive carbon and sulfur for each individual NMR structure, XRC structure, or computational model was calculated using GETAREA 1.1 (149), calculating the solvent accessible area in Å^2 for each atom of the structural model input. The averages and standard deviations for the solvent accessible area of each reactive carbon and sulfur (105) were calculated across all NMR structures for lysozyme and β-lactoglobulin A. XRC structures are generally considered more accurate and higher resolution, but NMR structures reveal protein dynamics, which are very important when utilizing a method that examined the average solvent accessibility of a side chain. For improved protein characterization, both the XRC and the NMR structure were examined and compared to the oxidation sites when possible.

Bacterial expression and purification of C14S Sml1p

The cloning, expression, and purification of the C14S mutants is described in Gupta *et al* (150). The expression construct causes the addition of a Gly-Ser to the N-terminus of the protein, which is known not to alter the biophysical properties of the protein (150). As previous work has been performed on the C14S mutant without the N-terminal addition, the C14S nomenclature will be maintained here, although the actual mutation in this protein is at Cys16. The elution fractions containing C14S Sml1p were examined by SDS-PAGE using a 15% acrylamide gel followed by analysis by

electrospray Fourier transform ion cyclotron resonance mass spectrometry (ES-FTMS) as described previously (151). The concentration of C14S Sml1p was determined by Coomassie Protein Assay Kit (PIERCE, Rockford, IL) using BSA as a standard.

Computational modeling of Sml1p

Partial sequence of Sml1p (residues 1-80 and 51-104) and the whole amino acid sequences of Sml1p were subjected to structure prediction using Rosetta/I-sites (152). The top five most commonly recurring models generated by Rosetta for each submitted sequence were returned. Full atom representations for all models were constructed using Modeller (153). Models that are consistent with experimental data for each segment were selected for manual construction of the constrained structural models of Sml1p. The quality of each model was evaluated as described earlier (150). Different, lower energy rotamers of side chains Phe87 and Arg26 were substituted for the manually constructed model, followed by energy minimization. The models were refined using Deep View (154) (http://www.expasy.ch/spdbv/).

Preparation of Aβ 1-40 peptides and fibrils

Aβ 1-40 fibrils were grown from chemically synthesized Aβ 1-40 peptide (Keck Biotechnology Center, Yale University). Briefly, Aβ 1-40 peptide was treated with TFA (Pierce) and HFIP (Acros) to remove any preexisting aggregates (155). Following disaggregation, the peptide was then dissolved stepwise in equal volumes of 2 mM NaOH and 2× PBS, pH 7.4, containing 0.1% sodium azide to give a peptide concentration of

34

approximately 50 μM. The peptide was then ultracentrifuged for 18 h at 50000g to remove any remaining aggregates. The soluble fraction after this step was used for monomer oxidation at a final concentration of ~0.1 mg/mL.

The fibril formation reaction was initiated by addition of a small quantity of sonicated Aβ 1-40 fibrils at a weight ratio of 1:750 and incubation at 37 ° C in PBS, pH 7.4. Fibril growth was monitored using thioflavin assay (156) until complete (~7 days). Fibrils were suspended by vigorous shaking prior to oxidation, and oxidation was performed at a final concentration of ~0.1 mg/mL. After fibril oxidation, the fibrils were disaggregated by incubation overnight in TFA, followed by dilution in PBS and solid phase extraction by C18 Lite SepPak (Waters). Following extraction, the remaining TFA was removed by evaporation in a SpeedVac (Savant Instruments).

CHAPTER 3—SURFACE MAPPING OF APOMYOGLOBIN

BY FENTON CHEMISTRY OXIDATION

The data presented below has been previously published as "Sharp, J. S., Becker, J. M., and Hettich, R. L. Protein surface mapping by chemical oxidation: Structural analysis by mass spectrometry. *Anal Biochem* (2003), **313**, pg 216-225". *All experimental work and data analysis were performed by Joshua S. Sharp.*

Introduction

One approach to generate hydroxyl radicals for protein solvent accessible surface studies is the use of chelated iron to catalyze the formation of hydroxyl radicals from hydrogen peroxide, known as Fenton chemistry. Previous use of this method was limited to utilizing the hydroxyl radicals to cleave the protein backbone, followed by determination of the sites of cleavage by gel electrophoresis (93). Gel electrophoresis is a low resolution, low accuracy mass measurement technique, limiting the quality of information gained from solvent accessible surface analyses by this method. Adding to the inherent inaccuracy of gel electrophoresis is the fact that amino acid side chain oxidation reactions occur on a shorter timescale than backbone reactions, resulting in the addition of an unknown number of oxygens to the protein (98). Consequently, the fragments resulting from oxidative cleavage of the protein backbone have a range of oxygens attached to the amino acid side chains, altering the fragments' masses and electrophoretic mobilities by an unknown amount and reducing the accuracy of oxidative cleavage site determinations.

We report an experimental protocol to use chemically generated hydroxyl radicals to probe the solvent accessible surface areas of apomyoglobin by determination of the sites of side chain oxidation by mass spectrometry. This method has the benefits of requiring only commonly available chemicals, fast reaction times, and well-characterized chemistry (157). Apomyoglobin was used because its secondary and tertiary structure has been extensively characterized by NMR (158-161), it is commercially available at high purity, easily digested, and has no prosthetic groups to complicate chemistry or mass spectrometric analysis. We present the results from this proof-of-principle experiment as evidence that selective amino acid side chain oxidation by chemically generated hydroxyl radicals can be used for the rapid analysis of protein solvent accessible surface areas.

Results

Chelated iron-catalyzed oxidation of apomyoglobin

Solutions of apomyoglobin were oxidized for varying amounts of time using chelated iron and hydrogen peroxide to generate solution phase hydroxyl radicals, which then react with the amino acid side chains to oxidize the protein. Chelated iron does not bind specifically to proteins, helping to ensure the non-specificity of the oxidation reaction (93, 98, 162). Analysis of the oxidation products by ES-FTMS showed that oxidation of apomyoglobin occurs in a time-dependent manner, and several amino acid side chain oxidation events can be detected before significant backbone cleavage occurs. Figure 8 shows the deconvoluted mass spectra of four time points in the presence of a 10-fold dilution of catalyst to slow the rate of oxidation. The zero time point was collected
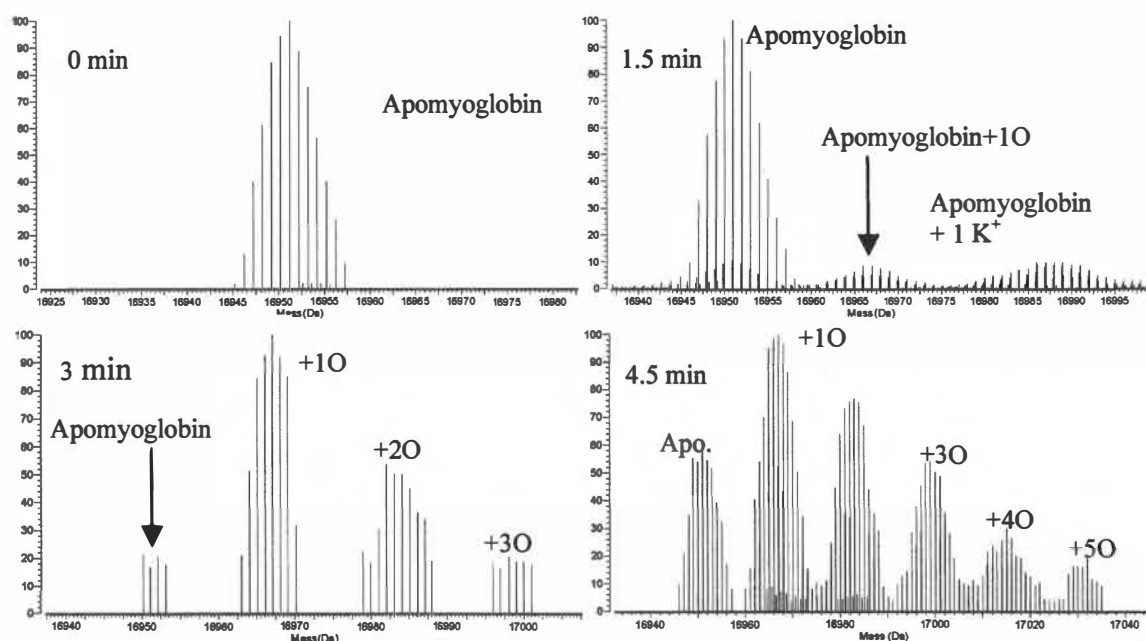
Figure 8: Apomyoglobin Oxidation Timecourse

Deconvoluted mass spectra of a time-course of apomyoglobin oxidation. The reactions were quenched at 90 second intervals and spectra obtained by ES-FTMS, showing that the number of oxygen atoms added to each molecule of apomyoglobin is time-dependent. At 4.5 minutes, in addition to the oxidation observed, significant fragmentation has begun to occur (data not shown).

after performing the reaction in the presence of 1M Tris buffer and allowed to sit on the bench top for one hour, showing that the presence of 1M Tris is sufficient to quench the oxidation reaction. Note that there are no oxidation events at the zero time point; the observed packet of ions is due to the naturally occurring ^{13}C isotopic distribution. A mass shift of 15.995 Da per oxidation event is characteristic of oxidation events. The increase in oxidation events per molecule is clearly time-dependent, with the addition of up to five oxygens as well as cleavage of the protein backbone by the hydroxyl radicals occurring at 4.5 minutes (data not shown).

One concern with the development of this technique is ensuring that the oxidation reactions probe the native structure of the protein. Multiple oxidations could conceivably alter the structure of the protein. Therefore, we used direct infusion ES-FTMS to monitor the oxidation status of the intact protein, and chose a time point that represented a small number of oxidation events per protein molecule distributed across the reactive, solvent accessible side chains. Thus, we could help ensure that the oxidation events themselves do not invalidate the data gathered by altering the structure of the protein with initial oxidation events and then oxidizing a site that is buried in the native structure. No evidence of fragmentation or cross-linking is observed in this mass spectrum of oxidized apomyoglobin.

Analysis of oxidation sites by tandem mass spectrometry

In order to determine which amino acid side chains are solvent accessible, we determined the sites of oxidation by both direct infusion ES-FTMS with SORI-CAD fragmentation and by reverse phase capillary LC-MS/MS in data-dependent mode.

The use of high resolution ES-FTMS allows for high accuracy mass measurements to ensure the identity of oxidized peptides, while LC-MS/MS is a robust, partially automated method for interrogating the complex mixture to assign oxidation sites. Full sequence coverage of non-oxidized peptides was obtained by ES-FTMS for all peptides that did not contain one of the two methionines in the sequence. Oxidized peptides containing the two methionines were detected, as well as oxidized peptides containing either Phe151 or Trp7 and Leu11. The high resolution and mass accuracy of the ES-FTMS measurements ensured that the peptides were oxidized and not a different adduct. Oxidized peptides containing all sites identified by LC-MS/MS were also found by ES-FTMS. Listed in Table 4 are the residues that we were able to identify as sites of oxidation by tandem mass spectrometry. Oxidized peptides were identified by their masses compared to an *in silico* digestion of apomyoglobin using PROWL (163).

Figure 9 shows the MS/MS spectrum that identifies Phe151 as an oxidized residue. The presence of an abundant non-oxidized b5 ion with no (b5+O) ion present shows that the peptide is oxidized at or C-terminal to Phe151. The presence of an abundant (b6+O) ion with no non-oxidized b6 ion detected shows that oxidation of this peptide occurred exclusively at Phe151. The additional presence of b4 and (b7+O) further confirm this assignment. As calculated from the crystal structure of horse heart myoglobin, Phe151 has 20.75$Å^2$ side chain solvent accessible area.

Table 5 summarizes the MS/MS spectra that reveal the sites of oxidation. In the case of peptide 119-133, the CAD spectrum allowed assignment of Met131 as the sole oxidation site. The presence of the (y3+O) ion with the non-oxidized y2 ion shows that the oxygen is present on Met131. The lack of non-oxidized y3 ion, as well as the lack of

Table 4 Sites of Apomyoglobin Oxidation

Measured Mass (Da)	Peptide	Calculated Oxidized Mass (Da)	Site of Oxidation by LC-MS/MS	Surface Area by X-ray Model (Å^2)
1830.889	1-16	1830.889	Trp7	13.42
1830.889	1-16	1830.889	Leu11	40.44
1494.716	48-56	1494.723	Met55	5.74
1517.660	119-133	1517.656	Met131	0.00
956.454	146-153	956.460	Phe151	20.75

A table of all oxidized peptides detected by ES-FTMS and the sites of oxidation on each determined by LC-MS/MS. The calculated masses were determined using PROWL (163). The measured and calculated masses all represent the deprotonated peptide. The side chain solvent accessible areas were calculated using GETAREA 1.1 (149) using the X-ray structure of horse heart myoglobin (1YMB) (148).
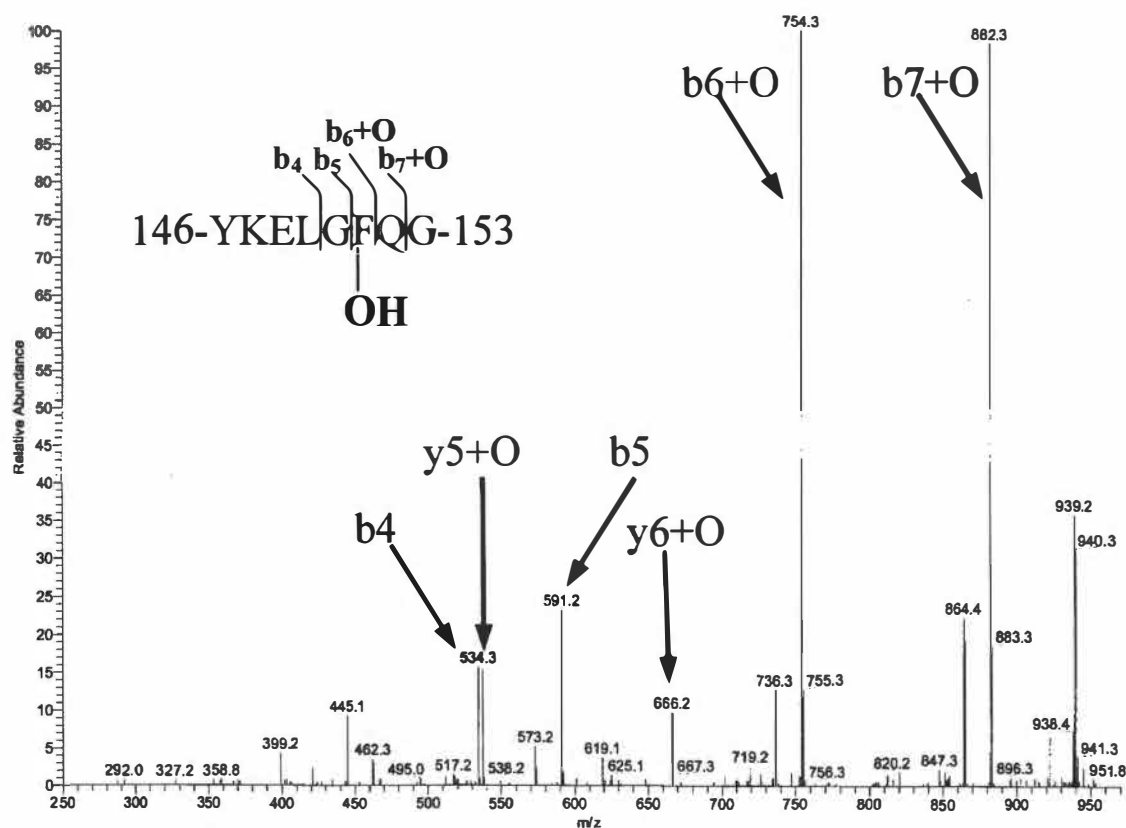
Figure 9: MS/MS Spectrum to Identify Oxidation in Lysozyme

A representative ion trap MS/MS spectrum used to determine sites of oxidation. The presence of an abundant non-oxidized b5 ion with no (b5+O) ion present shows that the peptide is oxidized at or C-terminal to Phe151 of lysozyme. The presence of an abundant (b6+O) ion with no non-oxidized b6 ion detected shows that oxidation of this peptide occurred exclusively at Phe151. All oxidation site assignments were performed in a similar manner.

Table 5 MS/MS Fragmentation of Oxidized Peptides

Peptide	Detected b ions	Detected y ions	Oxidized Residues
1-GLSDGEWQ QVLNVWGK-16	b9, b9+O, b10, b10+O, b11+O, b14+O, b15+O	y5, y6, y6+O, y7, y7+O, y9, y9+O, y10+O, y12+O, y13+O, y14+O	W7, L11
51-TEAEMK ASEDLKK-63	b2	y8, y9+O, y11+O	M55
119-HPGDFGAD AQGAMTK-133	b7, b8, b9	y2, y3+O, y5+O, y6+O	M131
146-YKELGFQG-153	b4, b5, b6+O, b7+O	y5+O, y6+O	F151

A table of the MS/MS fragments from oxidized peptides. Four oxidized peptides were found by LC-MS/MS. The MS/MS spectra were analyzed manually and assignments made for various fragment ions. The fragments assigned are listed in the table above, along with the assignments of the oxidation sites made from the fragmentation patterns.

a $(y2+O)$ ion, shows that the oxidation is exclusively on Met131. In the X-ray crystal structure of horse myoglobin, Met131 is completely buried in the interior of the protein and has 0.00 $Å^2$ of solvent accessible side chain surface area.

Surprisingly, the fragments from CAD of peptide 1-16 allow for the assignment of Trp7 and Leu11 as *two distinct oxidation sites in a single peptide with only one oxidation event per peptide.* The presence of both oxidized and non-oxidized fragment ions for many b and y ions shows that there is more than one oxidation site on the peptide. In the y ion series, y14 through y10 are only found as oxidized fragments, while y9 and $(y9+O)$ are present, showing that the N-terminal oxidation site is at Trp7. The y6 and $(y6+O)$ ion are also both present, while only the non-oxidized y5 ion is present, showing that the C-terminal oxidation site is Leu11. According to the crystal structure of horse myoglobin, the side chain of Trp7 has 13.42 $Å^2$ of solvent exposed surface area, while the side chain of Leu11 has 40.44 $Å^2$. The ability to resolve oxidation isomers greatly increases the utility of the technique.

The fragments from the CAD of peptide 51-63 were used to determine the oxidation of Met55 of apomyoglobin. The presence of an abundant $(y11+O)^{+2}$ ion as well as a $(y9+O)^{+2}$ ion with no unoxidized $(y9)^{+2}$ present show that all oxidation events detected at this peptide must occur at or C-terminal to Met55. However, the presence of a $(y8)^{+2}$ ion with no $(y8+O)^{+2}$ ion detected shows that the sole site of oxidation on this peptide occurs at Met55. As shown in Table 4, according to the X-ray structure of myoglobin, Met55 has 5.74 $Å^2$ side chain solvent accessible surface area, which is relatively small.

A calculation based on the X-ray crystal structure of holomyoglobin of the solvent accessible side chain areas of all residues in apomyoglobin that are chemically susceptible to modification by hydroxyl radicals is presented in Figure 10. The residues that we found to be oxidized are not the most solvent accessible according to the structure of holomyoglobin. A partial NMR structure of sperm whale apomyoglobin, however, shows that three regions of apomyoglobin are disordered (161), causing the residues in these disordered regions to be extremely solvent accessible. As shown in Figure 10, Phe151 and Met55 are directly in one of the disordered regions, implying that their side chain solvent accessibilities to be much higher than the calculations based on the X-ray crystal structure of holomyoglobin suggested.

Trp7 and Leu11 are not in a disordered region, and yet were found to be oxidized by our method. As shown in Figure 11, Trp7 and Leu11 are protected from solvent in holomyoglobin by a region that is disordered in apomyoglobin. Thus, in apomyoglobin, Trp7 and Leu11 have a much greater solvent accessible side chain area than that calculated from the X-ray structure of holomyoglobin. Therefore, the oxidation sites identified by our method are shown by established structural methods to be highly solvent accessible in the structure of apomyoglobin.

Discussion

With the exception of methionines, all of the residues oxidized and identified in this study are highly solvent accessible according to the NMR structure of apomyoglobin. The data presented above (most notably the complete lack of false positives) provide evidence that the use of chemically generated hydroxyl radicals as probes to determine

45
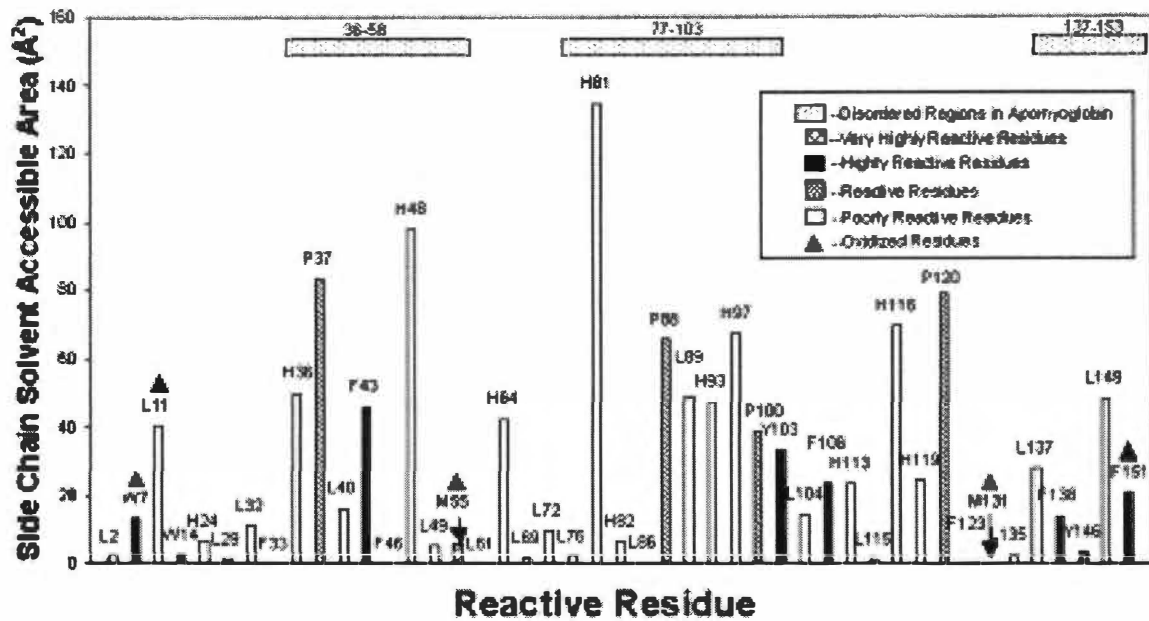
Figure 10: Analysis of Oxidation Sites in Apomyoglobin

Side chain solvent accessibilities of all amino acid residues in apomyoglobin that are

either very highly reactive (cysteine or methionine, crossed); highly reactive

(phenylalanine, tyrosine, or tryptophan, black); reactive (proline, diagonal); or poorly

reactive (histidine or leucine, white) (100). Residues determined to be oxidized by

tandem mass spectra are labeled with black triangles. The vertical axis represents the

side chain solvent accessible area in Å^2 as calculated from the crystal structure of

holomyoglobin, PDB identifier 1YMB (148). Plotted across the top of the graph are

dotted bars representing disordered regions as determined from the partial NMR structure

of sperm whale apomyoglobin (161).

46
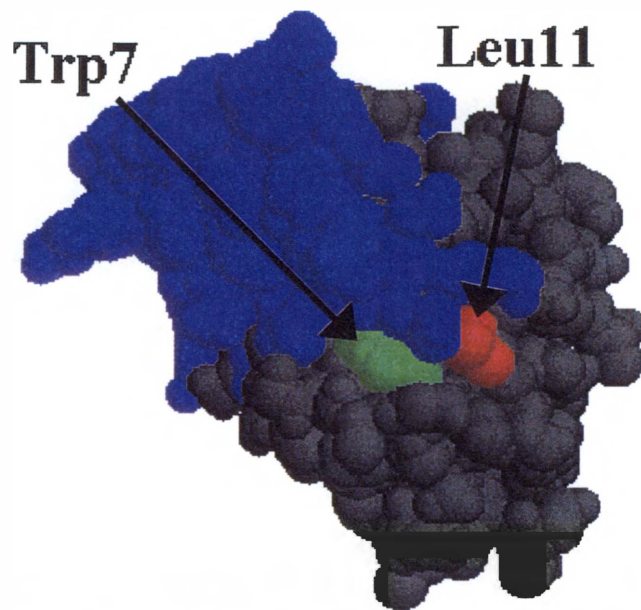
Figure 11: Effect of Disordered Regions on Oxidation Rates

The X-ray crystal structure of holomyoglobin, PDB identifier 1YMB (148). Shown in

green is the side chain of Trp7; shown in red is the side chain of Leu11. The blue areas

are all regions determined by the partial NMR structure of sperm whale apomyoglobin to

be disordered (161).

protein surface residues is feasible, with the exception of methionine oxidation. The short reaction time (30 seconds) along with relatively short digestion and analysis times suggest that this method would be useful as a rapid analysis of certain aspects of protein structure.

The fact that Met131 has essentially no solvent accessible area, even considering the disordered regions in apomyoglobin, and yet is still oxidized suggests that there is a different mechanism for the oxidation of methionine that is not dictated by solvent accessibility. These data are supported by observations in studies with lysozyme using synchrotron radiation that also found buried methionines to be oxidized (102). Kiselar *et al* postulated in previous work that methionines are oxidized by an intramolecular radical transfer mechanism, in which a hydroxyl radical would first abstract a hydrogen from a solvent accessible reactive residue. Then, another side chain close to the radical site would donate an electron to the solvent accessible radical. That solvent accessible residue would then pick up a proton, transferring the radical to the electron-donating group. Molecular oxygen would then oxidize the radical, which would not necessarily be solvent accessible (101). Met131, the only buried methionine in the sequence, is in close proximity to both Leu11 and Trp7, which were both shown to be oxidized. Therefore, we propose that a similar radical transfer mechanism is responsible, at least in part, for the oxidation of Met131. In addition, peptides with oxidized methionines can be detected in ES-FTMS spectra of tryptic digests of apomyoglobin that have undergone no oxidation reactions (data not shown), suggesting that methionine oxidation also occurs, at least in part, during or after digestion.

48

According to the X-ray structure of horse myoglobin, Leu11 has 40.44 Å^2 of solvent accessible side chain surface area. The previous observation that the main region in holomyoglobin that partially protects Trp7 and Leu11 from solvent is disordered in apomyoglobin shows that the actual solvent accessible regions of the Trp7 and Leu11 side chains are greater than those calculated from holomyoglobin. Even so, Phe43 is within a disordered region and has a higher chemical reactivity than leucine. Phe41 should also have a higher solvent accessibility than Leu11, as the group is actually within a disordered region; however, no oxidation of Phe43 is observed. The fact that an amino acid residue with a higher chemical reactivity within a disordered region is not oxidized while Leu11 is oxidized suggests that either Phe43 is not fully exposed to solvent even within a disordered region, or that a factor other than solvent accessibility and amino acid identity, such as the local chemical environment of the amino acid side chain, may play a role in the rate of oxidation. Unfortunately, due to the uncontrolled Fenton chemistry catalyzed by the iron bound to the heme group, holomyoglobin was not able to be analyzed to test the effect of disordered regions on the rate of oxidation.

While only limited oxidation was performed in this initial experiment, the failure to observe the oxidation of certain residues which one would expect to be oxidized by no means invalidates the approach presented in this report. All residues (except methionine) that were found to be oxidized are highly solvent accessible in the native solution phase structure of apomyoglobin as determined by NMR (161). The data presented in this study suggest that the technique presented here may have applications in a variety of structural analyses, including protein folding and protein-protein interaction studies. X-ray synchrotron studies, which use the same hydroxyl radical probe, have been utilized to

examine protein unfolding (103) and protein structure (102). A chemical means for generating the hydroxyl radical probe should allow for more widespread use of this rapid structural technique. Though this method has many advantages, there are some drawbacks. A potential drawback with this method is that the high concentrations of iron salt, EDTA, and ascorbate could be causing conformational changes in the analyte protein. Previous studies using Fe-EDTA catalyzed reactions to cleave the protein backbone and determining sites of cleavage by gel electrophoresis have thus far agreed with experimental data gathered by other methods (93, 162). Therefore, existing empirical evidence suggests that the chemical oxidation method presented here accurately probes the surface of proteins. Another drawback is that, due to the nature of the hydroxyl radical-generating reaction, this method is not applicable to metal-binding proteins. The data generated by mass spectrometry can determine down to the amino acid side chain which area is oxidized; in order to obtain a more complete description of the surface of a protein, however, it may be necessary to increase the numbers of oxygen atoms attached to each protein molecule. Further testing of other model systems is also required to ensure that the oxidation method described in this report does indeed probe solvent accessible areas. Efforts to determine other factors that dictate the kinetics of side chain oxidation also should be pursued.

CHAPTER 4—RAPID ANALYSIS OF PROTEIN SOLVENT ACCESSIBLE SURFACES BY PHOTOCHEMICAL OXIDATION AND MASS SPECTROMETRY

The data presented below has been accepted for publication pending revision as "Sharp, J. S., Becker, J. M., and Hettich, R. L. Rapid Analysis of Protein Solvent Accessible Surfaces by Photochemical Oxidation and Mass Spectrometry. *Anal Chem* (2003), in review". *All experimental work and data analysis were performed by Joshua S. Sharp.*

Introduction

A promising nonspecific reagent that has been utilized for determining solvent accessibility is the hydroxyl radical. Several methods utilizing hydroxyl radicals as a solvent accessibility probe have been published. The first experiments utilized Fenton chemistry to generate the radicals, which were then used as a protein cleavage reagent (93, 98, 162). This method had some success, but was limited by the inherently poor resolution of gel electrophoresis, as well as the various side chain reactions that occurred prior to cleavage and altered the fragments' electrophoretic mobilities.

More recently, a method has been developed that utilized a high-energy X-ray synchrotron beamline to generate the hydroxyl radical probe from water for protein surface labeling (100, 101, 103, 164-167). This method has the distinct advantages of being a pulse labeling technique (reactions require milliseconds) that does not require the presence of chemicals that could alter protein structure. The main drawback to this experimental method is the requirement for a high-energy X-ray synchrotron beamline, a

resource that is not readily available to most researchers. Hydroxyl radicals generated by Fenton chemistry were also used to label certain amino acid side chains, and the site of modification determined by reverse phase liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) (134). Although this technique provided solvent accessibility information for a model protein (apomyoglobin), the high concentrations of iron salts and EDTA required for the reactions generated some concern about the maintenance of native protein structure, reproducibility of results, and the ease of sample preparation for mass spectrometric analysis.

The hydroxyl radical will rapidly label a variety of amino acid side chains; most notably, it will target sulfur-containing residues and aromatic residues, although oxidation of certain aliphatic residues has been reported (99-101, 134, 164-166). Previous measurements of the pseudo-first order rate constants for the oxidation of amino acid side chains have shown that, of the amino acids previously observed to be oxidized by this technique, the rate of oxidation is Cys>Trp, Tyr>Met>Phe>His>Ile>Leu>Pro. Each of these oxidation events occur significantly faster than oxidative backbone cleavages(106). One advantage of the hydroxyl radical as a probe of solvent accessibility is the fact that the hydroxyl radical preferentially interacts with sulfur-containing and aromatic residues that are often buried in protein structure. As such, the analytical value of each identified oxidation site is greater than if the technique labeled residues expected to be found on the surface of the protein, such as charged residues.

The purpose of this study was to develop a simple, rapid, and reliable method for generating hydroxyl radicals and to utilize these radicals to probe the solution-phase structure of two model proteins with well-characterized, stable NMR structural models,

52

hen egg white lysozyme and bovine β-lactoglobulin A, while avoiding the sample clean-up difficulties and high salt concentrations required for the previous approach utilizing Fenton chemistry (134). These two proteins have very different secondary and tertiary structures, and serve nicely as test models for a surface mapping protocol. The rate of oxidation of any one side chain should be influenced by the solvent accessibility of the reactive atoms in the side chain, as well as the inherent chemical reactivity of the side chain. Hydroxyl radicals were generated by UV irradiation of hydrogen peroxide, and the model protein was allowed to react with the hydroxyl radicals. The resulting oxidized protein was analyzed by high resolution, high accuracy electrospray Fourier transform mass spectrometry (ES-FTMS) to determine the extent of oxidation. The oxidized protein was then digested with trypsin, the resulting peptides were analyzed by LC-MS/MS in order to determine the residues that were oxidized, and then analyzed by ES-FTMS in order to both confirm the identity of oxidized peptides and to gather more comprehensive information on oxidized peptides that were not identified by LC-MS/MS. Finally, the sites of oxidation were analyzed against the known solution-phase NMR structures of the model proteins in order to determine the extent to which solvent accessibility influenced the rate of oxidation.

Results

Photochemical oxidation of lysozyme

Solutions of lysozyme and hydrogen peroxide were irradiated for varying amounts of time to generate hydroxyl radicals, and the amount of oxidation was determined by ES-FTMS. A control was also carried out in which the protein was

incubated at room temperature in hydrogen peroxide with no irradiation for 30 minutes. The amount of oxidation was found to increase with the UV irradiation time in a near linear fashion up to a point, after which cooperative oxidation was observed (Figure 12). After five minutes of oxidation, no backbone fragmentation was detected; therefore, no significant backbone oxidation is occurring. However, some fragmentation was detected after oxidation for one hour. When the sites of oxidation were determined by LC-MS/MS for the heavily irradiated sample, several buried cysteines were found to be oxidized (data not shown). Similarly, when irradiation was performed in the absence of hydrogen peroxide for forty minutes, no oxidation occurred, but the solution turned cloudy and the lysozyme began to precipitate. These data support the conclusion that, upon extensive irradiation, the protein began to unfold independently of oxidation, exposing sites that are highly reactive but were previously buried. These newly exposed, highly reactive sites were rapidly oxidized, leading to the apparent cooperative oxidation observed.

Previous published work reports that UV irradiation can damage disulfide bonds that are important in maintaining the structure of lysozyme (168). Since the unfolding seemed to occur at least partially independent of protein oxidation, it is probable that the protein was being damaged by the extensive UV irradiation. However, after a five minute irradiation period, no cooperative oxidation was observed, no fragmentation of the protein was detected, no buried residues were found to be oxidized (discussed below), and in the absence of hydrogen peroxide, no clouding or precipitation of lysozyme was detected. In addition, after a five minute oxidation, no alteration in the circular dichroism spectrum of lysozyme is detected (Figure 13). However, after one hour of oxidation,

54
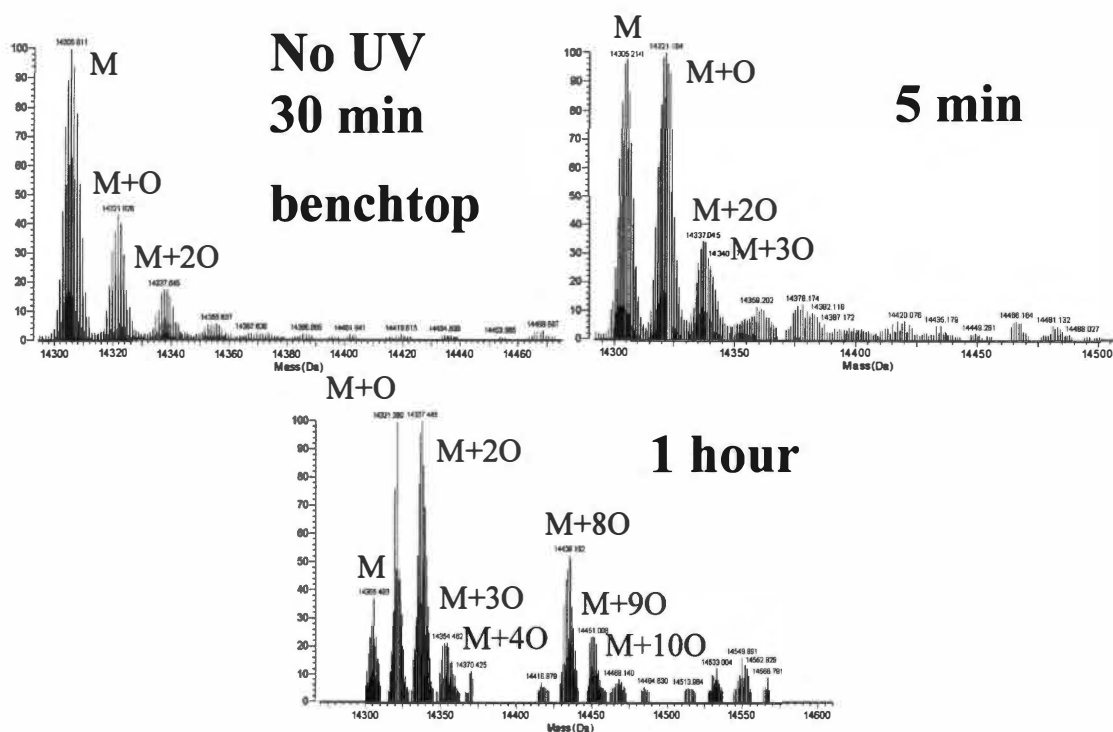
Figure 12: Timecourse Oxidation of Lysozyme

ES-FTMS spectrum of oxidation of lysozyme as a function of UV irradiation time. 0.5 mg/mL lysozyme in PBS and 15% H_2O_2 was irradiated for varying amounts of time using UV light. As the time increased, the amount of oxidized protein increased until one hour, where about a third of the protein unfolds and highly reactive, previously buried residues are exposed to solvent and oxidized very rapidly (cooperative oxidation). Some oxidation occurred when the protein was incubated for 30 minutes in 15% H_2O_2 with no UV irradiation at ambient temperatures; to prevent uncontrolled oxidation, solid phase extraction of oxidized protein was always performed directly after irradiation. The mass of unmodified lysozyme is labeled M, with the additions of multiple oxygens noted.
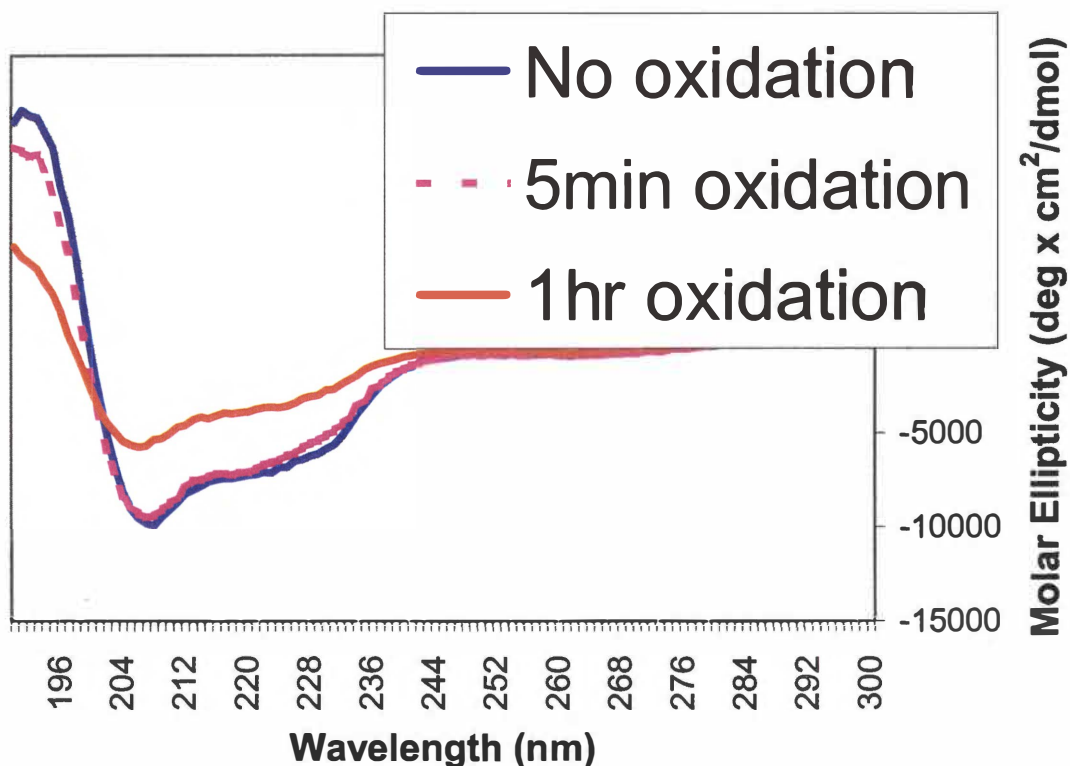
Figure 13: Effect of Oxidation on CD Spectrum of Lysozyme

Circular dichroism spectra of hen egg white lysozyme (blue line), lysozyme after five minutes of UV irradiation in the presence of 15% H_2O_2 (pink dashes), and after one hour of UV irradiation in the presence of 15% H_2O_2 (red line). The deconvolution of the spectra show little change between the control and the five minute oxidation spectra, while after one hour, approximately one third of the helical structure of lysozyme has been converted to random coil. This supports the hypothesis that roughly a third of the protein is in an almost fully denatured conformation, while the rest is in a near native state.

significant changes in the CD spectra are detected, showing that significant alterations in secondary structure occurs after extensive irradiation and oxidation. Due to the unaltered CD spectra after a five minute oxidation, we propose that limited UV irradiation and oxidation does not cause substantial structural changes in lysozyme.

LC-MS/MS analysis of oxidized lysozyme tryptic fragments

Lysozyme oxidized for five minutes was denatured and reduced, and then digested with trypsin. The tryptic digest mixture from oxidized lysozyme was loaded onto a C18 column for LC-MS/MS analysis to determine the sites of oxidation.

Previous work examining peptide oxidation by X-ray synchrotron radiolysis-generated hydroxyl radicals determined that sulfur-containing residues (Cys, Met) are the most highly reactive, followed by aromatic residues (Phe, Tyr, Trp), followed by proline, and finally histidine and leucine (100). Our results presented here, as well as previous results examining ionizing radiation-induced oxidation (105), suggests that histidine is oxidized more readily than proline and aliphatic residues, but probably not as fast as aromatic residues. Also, we have evidence from examining R67 dihydrofolate reductase from *E. coli* that isoleucine can also be oxidized by this method (J. Sharp, unpublished results).

After the LC-MS/MS experiment, the measured peptides were screened computationally for differential modifications using SEQUEST (144), as previously described. Double oxidation of methionine was observed during manual interpretation of MS/MS spectra. However, incorporation of this higher oxidation states into the SEQUEST searching parameters caused several oxidized peptides not to be identified;

57

therefore, parameters for lower level oxidation events were used for computational screening. All spectra which could contain an oxidized peptide were interpreted manually. Previous work using Fenton chemistry (134) and X-ray synchrotron radiolysis (99-102, 164-166) have shown that methionine can be oxidized regardless of its solvent accessibility, possibly via an electron transfer mechanism through a reactive residue present on the surface followed by subsequent oxidation by molecular oxygen (101, 102). Therefore, while all methionine oxidation events are measured, they are ignored for solvent accessibility analyses. The overall sequence coverage, including all unoxidized peptides and oxidized peptides identified manually in the LC-MS/MS run included the entire sequence of lysozyme except for residues 113-114. The non-methionine oxidized residues of lysozyme as detected by LC-MS/MS are shown as colored in Figure 14. When plotted onto the X-ray crystal structure of lysozyme, PDB accession number 193L, with water removed for clarity (169), the four oxidized residues can clearly be seen to be on the surface of the protein.

In order to gain a quantitative understanding of the effect of solvent accessibility on side chain oxidation, the average solvent accessibility of all reactive carbons and sulfurs of each side chain as calculated from fifty NMR structures was plotted, PDB accession number 1E8L (146). In Figure 15, the solvent accessibility of all sulfurs in cysteine and all reactive carbons in aromatic side chains is plotted. Cys6 has the greatest solvent accessibility of all cysteines for the reactive sulfur group, while the other cysteine sulfurs are essentially buried. The only cysteine identified as oxidized by the photochemical method is Cys6; none of the other highly reactive cysteines were seen to be oxidized, supporting the supposition that our technique only oxidizes exposed
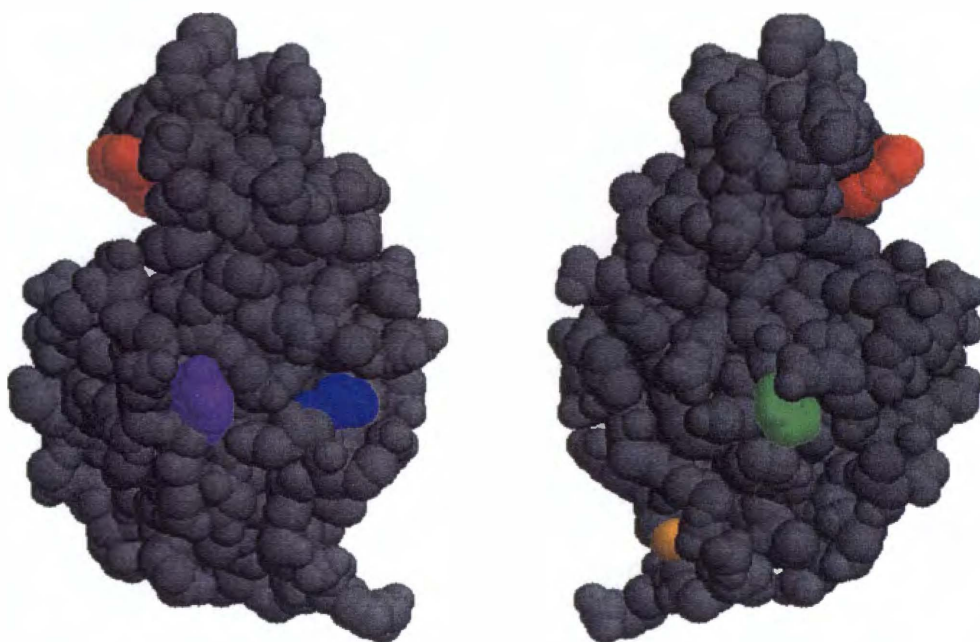
Figure 14: Sites of Oxidation Plotted on Lysozyme Structure

Two views of the X-ray crystal structure hen egg white lysozyme with waters removed for clarity (PDB accession number 193L). Oxidized residues confirmed by MS/MS are colored, with the sulfur of Cys6 is shown in orange, with the side chain of His15 (blue), Tyr20 (purple), Phe34 (green), and Trp62 (red) also indicated. Additional oxidized residues were detected by ES-FTMS (as shown in Table 6), but are not included here since the exact oxidation site was not confirmed.
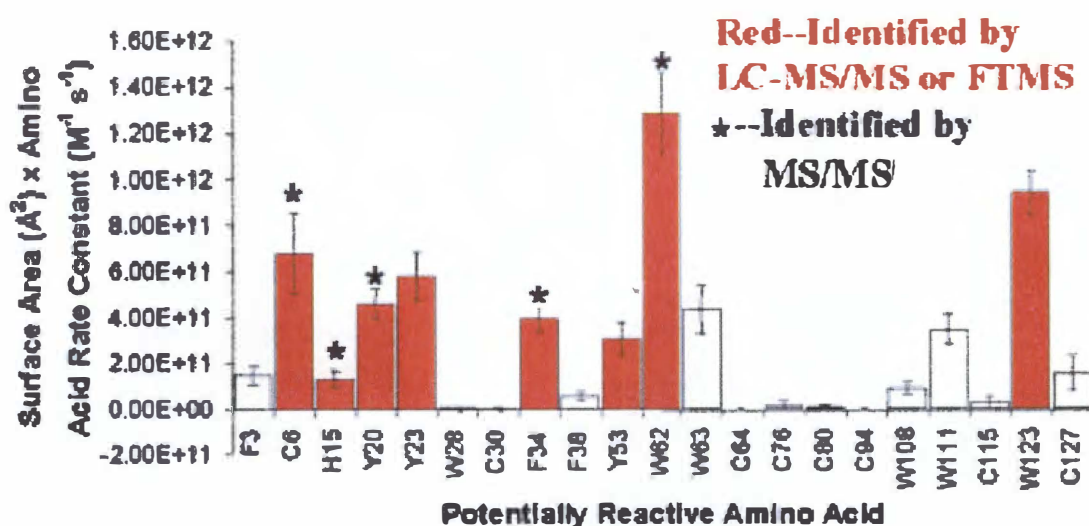
Figure 15: Quantitative Analysis of Oxidation of Lysozyme

The average solvent accessibility (in Å^2) of all reactive carbons and sulfurs in all

cysteine, aromatic, and histidine residues in hen egg white lysozyme as calculated from

50 NMR structures (PDB accession number 1E8L) times their known inherent chemical

reactivity ($M^{-1} s^{-1}$). Error bars represent one standard deviation. Residues that were

determined to be oxidized by LC-MS/MS or SORI-CAD are marked with an asterisk,

while all residues that either were the primary oxidation targets for peptides known to be

oxidized by ES-FTMS or were positively identified by LC-MS/MS or SORI-CAD are

shaded. All solvent accessibilities were calculated using GETAREA 1.1 with default

settings.

residues. Among the aromatic residues, the most accessible side chain is Trp62, and it was indeed found to be oxidized. His15 and Tyr20 were also found to be oxidized, and as shown by their average side chain solvent accessibility, they are exposed to solvent. Quantitation of oxidation ratios by tandem mass spectrometry on peptides with multiple oxidation events can be achieved by comparison of the oxidized and unoxidized fragment peak heights. While such a comparison is not entirely straightforward, the ratios of oxidized to unoxidized fragment ions between His15 and Tyr20 in the LC-MS/MS experiment suggest that the majority of oxidation on peptide 15-22 occurred on Tyr20, with considerably less oxidation occurring on His15. The LC-MS/MS experiment did not detect oxidized Tyr23, Phe34, Trp63, Trp111, or Trp123 (all of which are more reactive and have side chains at least as solvent accessible as His15); however, oxidized forms of peptides containing these residues were detected by ES-FTMS (Table 6, below). In addition, SORI-CAD experiments showed that the exclusive site of oxidation on peptide 34-45 is Phe34.

In order to determine if the extent of oxidation is indicative of the amount of solvent accessibility, the peptides resulting from the tryptic digest of oxidized lysozyme were analyzed by ES-FTMS. Table 6 shows the oxidized peptides detected by direct infusion ES-FTMS. The sequence of lysozyme is shown above the table; the regions of lysozyme for which a peptide was detected (either oxidized or unoxidized) by ES-FTMS are underlined, residues that were confirmed as oxidized by MS/MS are in bold, and peptides which were shown to be oxidized by ES-FTMS but with no oxidation sites identified by LC-MS/MS are italicized. In the table itself, all masses listed are monoisotopic masses, and residues which were positively identified as oxidized by

61

Table 6 ES-FTMS Measurement of Oxidized Peptides from Lysozyme

```
      1     *    10     *    20     *    30     *    40     *    50
  1 KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGS
 51 TDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVS
101 DGNGMNAWVAWRNRCKGTDVQAWIRGCRL
```

Peptide (residues)	Measured Deconvoluted Masses (Da)	Oxidization Forms	Ratio of Oxidized/ Unoxidized *	Reactive Residues
15-21	873.408	M	.0114	**H15**, L17, **Y20**
	889.409	M+O		
62-68	935.377	M	.0620	**W62**, W63, C64
	951.368	M+O		
117-125	1044.538	M	.1324	W123, I124
	1060.536	M+O		
22-33	1267.601	M	.0445	Y23, L25, W28, C30
	1283.583	M+O		
34-45	1427.669	M	.0101	**F34**, F38
	1443.674	M+O		
62-73	1433.711	M	.1177	**W62**, W63, C64, P70
	1449.721	M+O		
98-112	1674.788	M	17.0959 (.0995)*	I98, **M105**, W108, W111
	1690.788	M+O		
	1706.740	M+2O		
46-61	1752.828	M	.0063	Y53, I55, L56, I58
	1768.801	M+O		
97-112	1802.890	M	17.9175 (.1189)*	I97, **M105**, W108, W111
	1818.902	M+O		
	1834.896	M+2O		
15-33	2123.016	M	.0464	**H15**, L17, **Y20**, Y23, L25, W28, C30
	2139.000	M+O		

*-The ratio in parentheses is of mono-oxidized to di-oxidized peptide.

MS/MS are marked in bold, while residues which are probable sites of oxidation, but for which there are not reliable LC-MS/MS spectra, are underlined. Phe34, which is both underlined and bold in Table 6, was the only amino acid which was identified as oxidized based solely on SORI-CAD data from ES-FTMS. A sequence coverage of 89.9% of the amino acids was obtained by ES-FTMS analysis, with the notable lack of residues 6-13, containing the only oxidized cysteine detected by LC-MS/MS. No unoxidized peptide 6-13 was detected by the LC-MS/MS experiment, so it is probable that Cys6 is oxidized nearly 100% in this experiment.

Tyr23, Trp63, Trp111, and Trp123 are shown to be more solvent accessible than His15, and yet they were not found to be oxidized by MS/MS. The overall rate of oxidation for aromatic residues was shown to be faster than that of histidine (100); therefore, one would expect to observe oxidation on these residues. Failure to detect all peptides in a tryptic digest is common for mass spectrometry experiments, due to dynamic range issues, poor peptide ionization, or other factors. Automated LC-MS/MS experiments have the added problem of time resolution. As peptides elute from the reverse phase column, they are isolated and fragmented by the mass spectrometer in the order of signal intensity. The mass spectrometer is programmed to only fragment each particular peptide twice, ignoring it for a period of time after analyzing it. If an oxidized peptide coelutes with other, more abundant peptides, the oxidized peptide may not be fragmented in the time that it elutes from the LC column. Therefore, the failure to detect all peptides expected to be oxidized is not unusual; these peptides probably were oxidized, just not detected under these conditions. Examination of the ES-FTMS data in Table 6 shows that oxidation probably did occur on these residues and was not detected

63

by our LC-MS/MS method. In the peptide containing residues 22-33, the most highly solvent accessible, reactive residue by far is Tyr23. The ratio of oxidized to unoxidized peptide for 22-33 was higher than that of peptide 15-21, which contains both His15 and Tyr20. When the solvent accessibility of the Tyr23 is compared to that of Tyr20 and His15, it is found that Tyr23 has the greatest solvent accessibility (Figure 15).

Several solvent accessible tryptophan residues (Trp63, Trp11, and Trp123) were not detected by our LC-MS/MS method. The proximity of Trp63 to Trp62 made identification of the lesser oxidation site by tandem mass spectrometry very difficult. In order to determine the sites of oxidation from a mixture of oxidation isomers, MS/MS peaks must be measured for both the oxidized and the unoxidized fragmentation events that occur between the two oxidation sites. Since in the case of Trp62 and Trp63 the potential oxidation events would occur directly next to each other, there are only two ions that could differentiate the two events (the y ion and the b ion with the fragmentation site between residue 62 and 63). As the oxidation event at Trp62 was very abundant, the unoxidized b ion (and/or the oxidized y ion) from fragmentation between Trp62 and Trp63 representing an oxidation at Trp63 would be at a much lower signal-to-noise. Since there are only two potential ions that would allow the identification of the oxidation at Trp63, and the oxidation at Trp62 is expected to be much more prevalent, it remains possible that Trp63 was oxidized to a lesser extent, and simply could not be accurately identified.

It is also probable that the oxidized peptides containing Trp111, and Trp123 were present but simply not identified by the LC-MS/MS experiment. The ES-FTMS data show that, in both cases, there was a detectable level of oxidized peptide (Table 6). Due

64

to the fact that Met105 is present on the same peptide as Trp111, and that methionines can be doubly oxidized, accurate differentiation between the peptides with a single oxidation at Met105 and a single oxidation at Trp111, and peptides with a double oxidation at Met105 cannot be made quantitatively. Trp123, though, is the only readily oxidizable residue on peptide 117-125. The ratio of oxidized to unoxidized peptide for peptide 117-125 was very similar to that of peptide 62-68, the smallest peptide containing Trp62, which is known to be oxidized. Since the solvent accessible areas of the two residues are also similar, these data support the idea that solvent accessibility directly influences the rate of oxidation of an amino acid side chain by our method.

In order to estimate the effect of solvent accessibility on the rate of protein oxidation, the ratio of oxidized to unoxidized peptide was measured for all peptides in lysozyme where oxidation probably occurred on an aromatic side chain. Since only one cysteine was oxidized to a detectable level, and methionines were oxidized regardless of solvent accessibility, we could not plot the effect of solvent accessibility for sulfur-containing residues. In order to try to best estimate the effect of solvent accessibility on the rate of oxidation, we assumed (when necessary) that all oxidation on a peptide occurred on the most accessible, most reactive residue. Residue Phe34 was not included due to a low signal to noise ratio, which made accurate determination of the ratio of oxidized to unoxidized peptide difficult. While this assumption is overly simplified, it allows for a reasonable estimation of the effects of solvent accessibility on the rate of oxidation. We then plotted the side chain solvent accessible area of each candidate residue against the ratio of oxidized to unoxidized peptide on which the candidate residue resides (Figure 16). Since the only difference between the oxidized and unoxidized

65
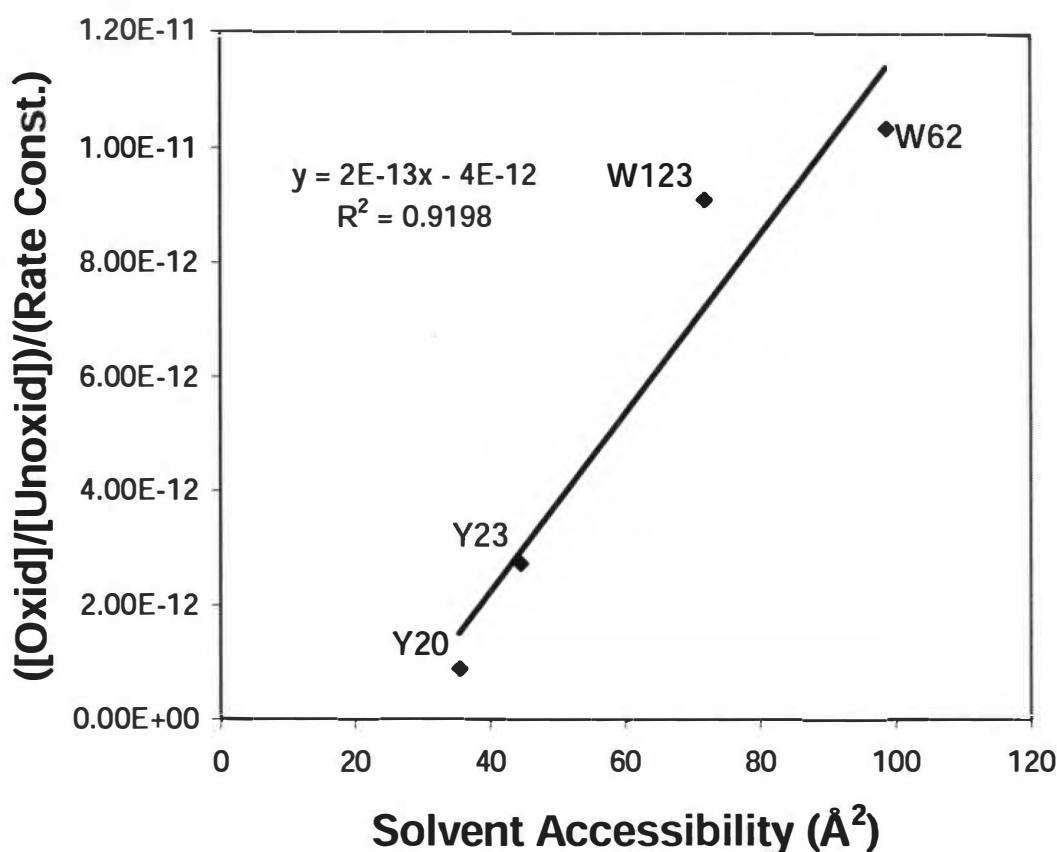
Figure 16: Rate of Oxidation as a Function of Solvent Accessibility

A plot of the solvent accessible area of the aromatic residues most likely to be oxidized on oxidized peptides detected by ES-FTMS versus the ratio of oxidized to unoxidized peptide signal in the ES-FTMS spectra times the inherent chemical reactivity of the amino acid (M^{-1} s^{-1}). Due to the nearly identical chemical structure of the oxidized peptide to its unoxidized form, the ratio of peptide signal should be almost identical to the ratio of peptide in solution.

versions of the peptide is one oxygen, the two peptides should ionize similarly, making the relative abundances of oxidized and unoxidized peptide peak heights indicative of the actual relative abundances of each form of the peptide in solution. The plot shows a definite influence of solvent accessibility on the rate of oxidation. The lack of absolute linearity could be due to several factors, such as partial oxidation at other residues on the peptide, fluctuations in the ES-FTMS signal causing incorrect ratios, differences in the chemical reactivity of a side chain due to local chemical influences from neighboring residues, or slight differences in ionization potential between the oxidized and unoxidized version. However, the plot shows a definite influence of solvent accessibility, and these data support the use of this protocol to measure solvent accessible residues.

Photochemical oxidation of β-lactoglobulin A

In order to further support the use of photochemical oxidation to probe the solvent accessible surface of a protein, we performed the oxidation of bovine β-lactoglobulin A, which is slightly larger than lysozyme, contains fewer disulfide bonds, and is mainly comprised of β-sheets as opposed to lysozyme's largely α-helical secondary structure. It was observed that, after 5 minutes of UV irradiation, a similar extent of oxidation was observed for β-lactoglobulin A as that observed for lysozyme. No cooperative oxidation was observed for β-lactoglobulin A (data not shown), supporting the premise that the protein remained folded throughout the oxidation process.

LC-MS/MS analysis of oxidized β-lactoglobulin A tryptic fragments

Oxidized β-lactoglobulin A was denatured, reduced, and digested with trypsin. The tryptic digest of oxidized β-lactoglobulin A was analyzed by LC-MS/MS and the data searched computationally by SEQUEST for oxidation events. A total of four non-methionine oxidation sites were found for β-lactoglobulin A in this experiment. These oxidation sites are shown on the X-ray crystal structure of β-lactoglobulin A, PDB accession number 1BEB, monomer with no waters show (170), in Figure 17. All oxidized side chains appear to be solvent accessible from the X-ray crystal structure.

Quantitative analysis from ten NMR structures of β-lactoglobulin A, PDB accession number 1CJ5 (147), showed a trend previously noted with lysozyme. Figure 18 shows the average solvent accessibility of the reactive sulfur of each cysteine and the reactive carbon in each aromatic residue in β-lactoglobulin A. Cys160 clearly has the greatest solvent accessibility of all cysteines in β-lactoglobulin A, and Cys160 is the only cysteine determined to be oxidized by LC-MS/MS. The sulfur of Cys66 is also significantly solvent accessible; however, the peptide is not detected in the LC-MS/MS experiment in either an oxidized or unoxidized state. Therefore, the extent of oxidation of Cys66 cannot be determined from this experiment; it may be completely oxidized or completely unoxidized. This observation is consistent with our hypothesis that the solvent accessibility of the reactive sulfur of cysteine directly influences the rate of oxidation. Detailed analysis of ES-FTMS data (below) will shed light on the oxidation state of Cys66; however, poor sequence coverage by LC-MS/MS makes this comparison difficult based solely on data obtained from the QIT.
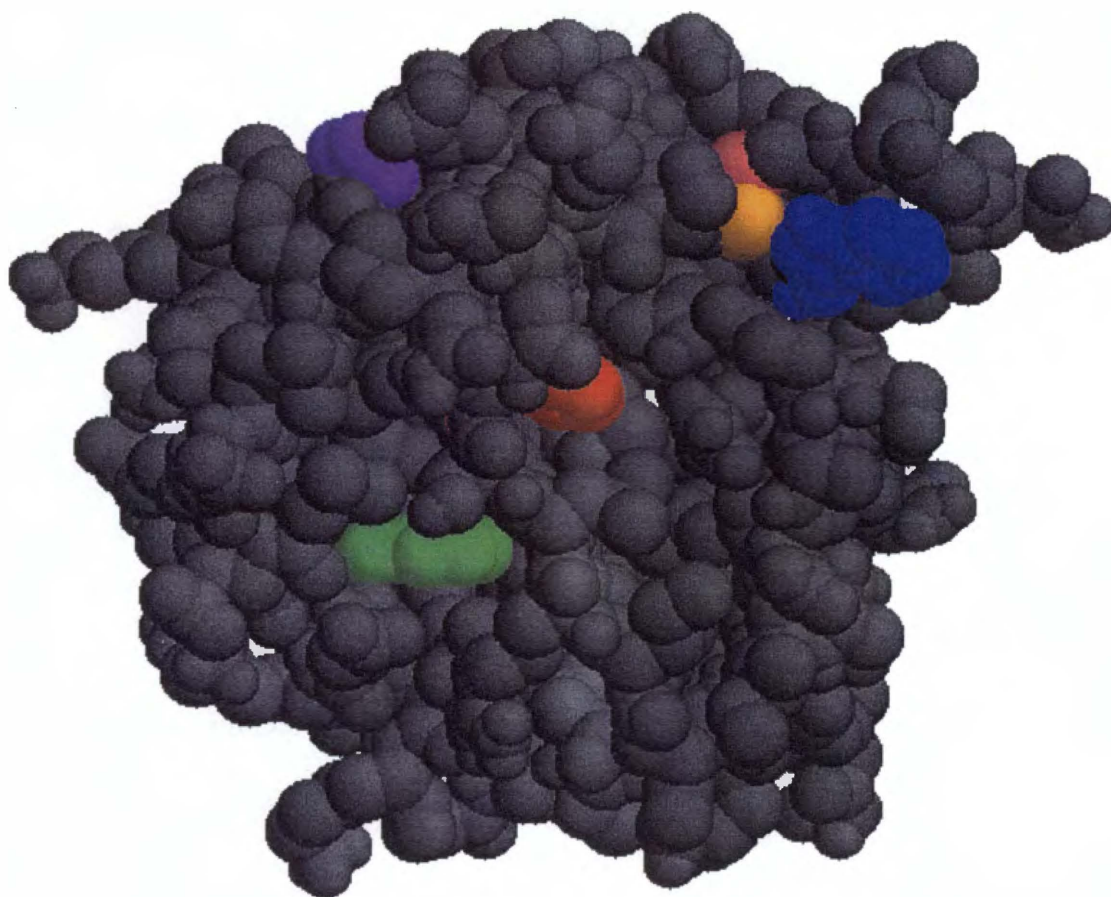
Figure 17: Sites of Oxidation Plotted on β-lactoglobulin A Structure

The X-ray crystal structure of bovine β-lactoglobulin A (PDB accession number 1BEB, monomer with waters removed for clarity). Oxidized residues confirmed by LC-MS/MS are colored, with the sulfur of Cys160 is colored orange, with the side chain of Pro126 (purple), Phe136 (green), and Phe151 (red) also indicated. Trp61 (blue) and Cys66 (pink) were detected by ES-FTMS (as shown in Table 6), but the exact oxidation site was not confirmed.
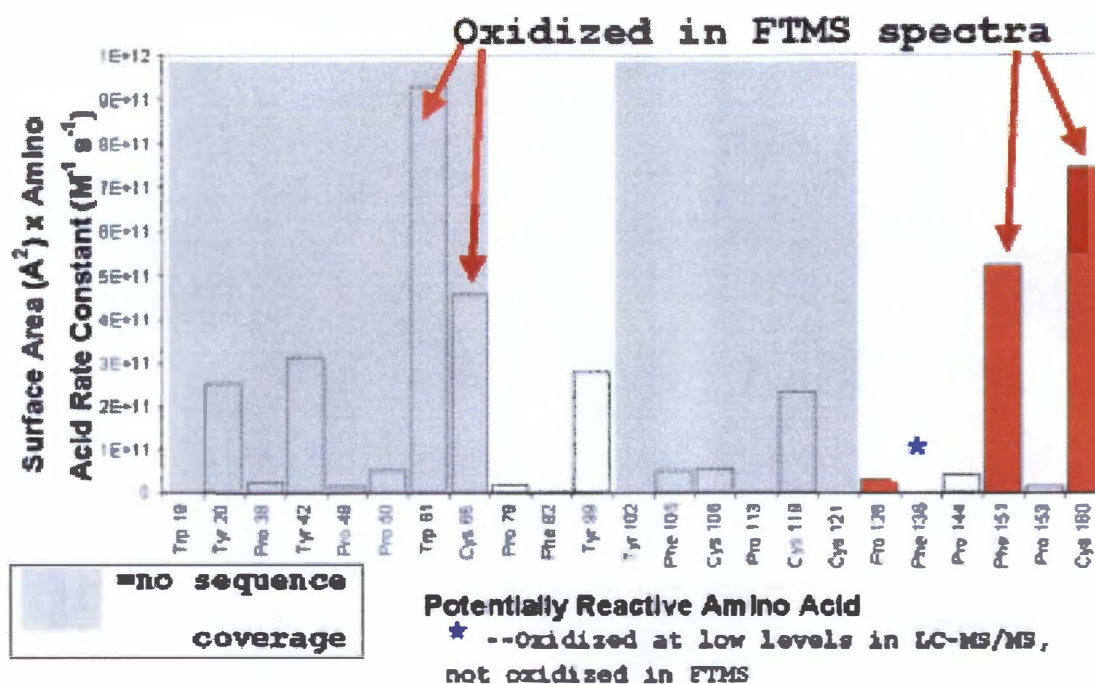
Figure 18: Quantitative Analysis of Oxidation of β-lactoglobulin A

The average solvent accessibility (in Å^2) of all reactive carbons and sulfurs in all

cysteine, aromatic, and histidine residues in bovine β-lactoglobulin A as calculated from

10 NMR structures (PDB accession number 1CJ5) times their known inherent chemical

reactivity ($M^{-1}\ s^{-1}$). Error bars represent one standard deviation. Residues that were

determined to be oxidized by LC-MS/MS are colored red (or marked with an asterisk

when the colored bar is not visible), while all residues that were the primary oxidation

targets for peptides known to be oxidized by ES-FTMS are marked with red arrows.

Regions of the protein which were not detected in the LC-MS/MS run as either oxidized

or unoxidized peptides are shaded gray. All solvent accessibilities were calculated using

GETAREA 1.1 with default settings.

Figure 18 also shows the solvent accessibility of the reactive carbons of aromatic and histidine residues in β-lactoglobulin A. Trp61 is clearly the most solvent accessible of all of the reactive side chains. However, the peptide that contains Trp61 (which is the same peptide that contains Cys66) was not detected in either an oxidized or unoxidized state in the LC-MS/MS experiment, so the extent of oxidation of Trp61 versus that of Cys66 cannot be determined by this experiment. The failure to detect a peptide containing Trp61 by LC-MS/MS could be due to poor peptide ionization or coelution with a peptide with a very high electrospray response (a dynamic range problem). Phe151 was found to be oxidized by this experiment, and has very solvent accessible reactive side chain carbons. An oxidation event also occurred at Phe136, which was shown by NMR to be relatively (but not completely) buried. When examining the LC-MS/MS selected ion chromatograms, peptide 125-138 was very prevalent in its unoxidized form, and it eluted from the column in detectable amounts for much longer than most peptides (over eight spectra). However, the oxidized form of the peptide was only detectable in one MS spectrum, and at very low levels (oxidized to unoxidized ratio of approximately .04 in the one LC-MS spectrum).

In order to quantitatively measure the rates of oxidation, as well as to try to identify other regions of the protein where oxidation occurred, we examined the peptides by ES-FTMS (Table 7). The sequence of bovine β-lactoglobulin A is shown above the table; the regions of β-lactoglobulin A for which a peptide was detected (either oxidized or unoxidized) by ES-FTMS are underlined, residues that were confirmed as oxidized by LC-MS/MS are in bold, and peptides which were shown to be oxidized by ES-FTMS but with no oxidation sites identified by LC-MS/MS are italicized. Significantly better

71

Table 7 ES-FTMS Measurement of Oxidized Peptides of β-lactoglobulin A

```
    1    *   10    *   20    *   30    *   40    *   50

  1 LIVTQTMKGLDIQKVAGTWYSLAMAASDISLLDAQSAPLRVYVEELKPTP

 51 EGDLEILLQKWENDECAQKKIIAEKTKIPAVFKIDALNENKVLVLDTDYK

101 KYLLFCMENSAEPEQSLVCQCLVRTPEVDDEALEKFDKALKALPMHIRLS

151 FNPTQLEEQCHI
```

Peptide (residues)	Measured Deconvoluted Masses (Da)	Oxidization Forms	Ratio of Oxidized/ Unoxidized	Reactive Residues
142-148	836.474	M	123.0101 (.0198)*	L143, P144, **M145**, H146, I147
	852.464	M+O		
	868.461	M+2O		
149-162	1657.774	M	.0109 (>20)**	L149, **F151**, P153, L156, **C160**, H161, I162
	1673.811	M+O		
102-124	2722.278	M+3O	-	Y102, L103, L104, F105, C106, **M107**, P113, L117
	2738.228			
61-83	2688.194	M	1.9425	W61, C66, I71, I72, I78, P79, F82
	2704.386	M+O		
15-40	2722.380	M+O	-	W19, Y20, L22, **M24**, I29, L31, L32, P38, L39
	2738.380			

*-The ratio in parentheses is of mono-oxidized to di-oxidized peptide.

**-The ratio in parentheses is from the LC-MS/MS experiment, where the tri-oxidized and unoxidized peptide co-eluted in the LC run.

sequence coverage was obtained by direct infusion ES-FTMS than by LC-MS/MS. In the table itself, all masses listed are monoisotopic masses, and residues which were positively identified as oxidized by LC-MS/MS are marked in bold, while residues which are probable sites of oxidation, but for which there is not reliable MS/MS spectra, are underlined. Unfortuntely, as only two non-methionine ([oxidized]/[unoxidized]) ratios could be determined, it was not possible to attempt to plot the correlation between solvent accessibility and apparent rate of reactivity.

Peptide 149-162 contains two oxidation sites (Phe151and Cys160), making the actual quantitation of the sites difficult. However, all detected Cys160 oxidation events had Cys160 triply oxidized, while the detected Phe151 oxidation event had only Phe151 oxidized. Therefore, we suggest that the ratio of mono-oxidized to unoxidized peptide 149-162 represents the rate of Phe151 oxidation, while the ratio of tri-oxidized to unoxidized peptide 149-162 represents the rate of Cys160 oxidation. As was the case when examining lysozyme, a cysteine that was known to be very extensively oxidized (either to the point of complete oxidation, as with lysozyme, or to almost complete oxidation with Cys160 of β-lactoglobulin A) and easily detected by LC-MS/MS was undetected in subsequent ES-FTMS analysis. Samples were analyzed immediately after tryptic digestion by LC-MS/MS, while the peptides were stored for about two weeks at -75° C in 0.1% acetic acid prior to ES-FTMS analysis. It is possible that peptides that have been triply oxidized on a cysteine (forming a sulfonic acid) are not stable for long periods of time in an acidic solution.

LC-MS/MS experiments showed that Phe136, which is buried in β-lactoglobulin A, is oxidized. When examining peptide 125-138 by ES-FTMS, without the increase in dynamic range offered by separations techniques, no oxidation event can be detected (even when attempting to isolate the oxidized peptide in the Penning trap). In addition, examination of the selected ion chromatogram of oxidized peptide 125-138 showed *very* little oxidized peptide present. We suggest, therefore, that an oxidation event at Phe136 was detected only due to the very high ionization efficiency of the peptide, its excellent fragmentation pattern, and its excellent separation from other peptides in our LC-MS/MS experiment; the actual amount of Phe136 oxidized was well below background. One oxidized peptide was found that had no oxidized residue detected by LC-MS/MS. This peptide (residues 61-83) contains Trp61, which is by far the most solvent exposed reactive residue, as well as Cys66, which is the second-most solvent accessible cysteine residue. The oxidized to unoxidized ratio was also very high, which corresponds with the high solvent accessibility of the Trp61 residue and the high reactivity and medium solvent accessibility of Cys66. As mentioned previously, oxidation of cysteine results in the addition of two or three oxygens; only the addition of one oxygen to peptide 61-83 was detected. Therefore, all oxidation on peptide 61-83 probably occurred on Trp61. The number of peptides oxidized on β-lactoglobulin A was too few for meaningful quantitative analysis to determine the relationship between solvent accessible area and rate of oxidation.

Discussion

Previous work utilizing X-ray synchrotron radiolysis to probe the solvent accessible surface of lysozyme reported a smaller set of oxidized residues than those reported here, detecting Cys6, Phe34, Trp62, and Trp123 as the non-methionine oxidized side chains by LC-MS/MS. Their work clearly demonstrated that the solvent accessibility relates almost linearly with the rate of oxidation in identical residues in lysozyme (101, 102). While we did not detect Phe34 or Trp123 by our LC-MS/MS experiments, ES-FTMS analysis showed that Phe34 and Trp123 were probably oxidized by our method. In addition, we were able to detect oxidized peptides from almost every region of lysozyme by ES-FTMS. While our ability to detect a greater amount of oxidized peptides may be a result of the greater performance of ES-FTMS as opposed to the matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) instrumentation utilized in by Kiselar *et al*(101), another possible explanation is an increase in the extent of oxidation. As such, from an oxidation coverage standpoint, our method compares favorably with current reported results.

The overall purpose of this study was to provide a useful method for rapidly probing the surface area of proteins in solution. As shown from the results presented here, this technique is suitable for the oxidation of solvent accessible residues. Through the use of both LC-MS/MS and direct infusion ES-FTMS, all oxidation events that would be expected to occur are detected, with no appreciable oxidation of buried residues. While not every solvent accessible residue was oxidized to a level detectable by our LC-MS/MS experiment, those oxidized side chains that were detected were all solvent accessible to some extent, and the extent of solvent accessibility (along with the inherent

75

chemical reactivity of the residue) had a direct influence on the rate of oxidation. The extent of oxidation was comparable to (if not more extensive than) previous X-ray synchrotron studies performed on lysozyme (101), however, the pattern of oxidation detected by LC-MS/MS analysis was somewhat different.

Photochemical oxidation has the advantages of being a simple protocol for labeling of a variety of residues (mainly hydrophobic), most of which are normally expected to be buried in protein structures. Since it is residues that are typically buried that are labeled, while there may only be four to five non-methionine residues oxidized in the protein, the value of each oxidation event is much greater than if, for instance, the method labeled charged residues that would be expected to be on the surface. The ability of this technique to provide structural data for targeted proteins in a rapid, relatively simple fashion without the necessity for an X-ray synchrotron beamline could be of great use to structural biologists to test structural models or investigate conformational changes. Unfortunately, the requirement for hydrogen peroxide prevents the use of this technique on transition metal-binding proteins due to uncontrolled Fenton chemistry that would occur. Also, X-ray synchrotron radiolysis surface mapping still maintains the distinct advantage of being a pulse labeling method, and thus more useful for measuring protein-protein interactions and protein folding kinetics. We are currently working with pulsed UV sources in an attempt to develop this technique into a pulse labeling method.

The recent reports in the area of constrained computation modeling show that the use of experimental constraints in computation protein modeling greatly increase the accuracy of the model (41, 42). The potential for utilizing a technique such as photochemical oxidation to generate experimental constraints on protein structure in a

high throughput manner may enable structural genomics to provide structural models of reasonable accuracy at a turnover rate comparable to that of genomics and proteomics, allowing for great leaps in systems biology. We are currently exploring possibilities for utilizing photochemical oxidation data to improve computational models of proteins with unknown three-dimensional structures.

CHAPTER 5—APPLICATION OF PHOTOCHEMICAL SURFACE MAPPING OF C14S SML1P TO CONSTRAINED COMPUTATIONAL MODELING

Introduction

Sml1p is a small protein from the yeast *Saccharomyces cerevisiae* that binds to the large subunit (Rnr1p) of the ribonucleotide reductase (RNR) complex and inhibits the activity of this enzyme for catalyzing the rate-limiting step of the *de novo* synthesis of deoxynucleotides (dNTPs) (171, 172). During S-phase and after DNA damage, the intracellular level of Sml1 is substantially reduced through its degradation (173). The degradation of Sml1p activates RNR, increasing the levels of the dNTP pools to enhance the capacity of DNA repair (171). The RNR activity must be tightly regulated, since failure to control the concentration or relative amounts of dNTP pools may lead to cell death or genetic abnormalities (172, 174). Structural characterization of Sml1p is critical for understanding the nature of the interaction between this protein and RNR.

Although there is some biochemical information about the function of Sml1p, very little structural data is available for this important checkpoint protein. Sml1p is known to bind to Rnr1p with 1:1 stoichiometry, in which the binding domain of Sml1p involves the C-terminal 33 residues (175, 176). NMR studies have indicated that Sml1p has two large alpha-helices oriented in an anti-parallel manner (177); overall, this protein generally is described as a loosely folded structure in solution. However, there is no

three-dimensional structural information available for either Rnr1p or Sml1p, much less for the Sml1p/Rnr1p complex. In order to better understand the structure-function relationship of this important complex, a reliable three-dimensional structural model for the components must be generated.

Protein structure prediction methods are playing, and will continue to play, an important role in high throughput structural genomics due to the limitations of structural determination methods such as x-ray crystallography and nuclear magnetic resonance (NMR). The progress in structure prediction techniques has been assessed objectively by the biennial community-wide experiment, Critical Assessment of Protein Structure Prediction (CASP) (178-180). These assessments have shown that rapid progress is being made in the prediction techniques of protein structures. One key group of protein structure prediction methods is comparative modeling and threading, which rely on the detectable similarity between the target sequence and at least one solved protein structure in the PDB database (181). If there is no reliable template protein structure identified by sequence- or structure-based approaches, a target protein could not have its structure predicted using such methods. This is the case for Sml1p. Both sequence homology and threading methods failed to detect any suitable template structure (150).

The second group of prediction methods, called *de novo* or *ab initio* methods, has no such limitation since it does not rely on the existing structural folds. The *ab initio* method is based on the first principal that the native state of the protein is at the global energy minimum. Significant advances have been achieved in the field of *ab initio* protein structure prediction as seen in the recent CASP3 (178), CASP4 (39) and CASP5 (http://predictioncenter.llnl.gov/casp5/). Rosetta is probably the best method in the *ab*

initio prediction category (39). It is a pseudo-*ab initio* prediction method since it uses small segment information from a structural database. A significant feature of Rosetta is the separation of local and nonlocal interactions in which each short segment of the chain samples possible local structures adopted by the short sequence in known protein structures, followed by folding to the native state when local segments are oriented such that low energy nonlocal interactions are achieved throughout the protein, favoring paired β-strands and *buried hydrophobic residues*. In contrast with some modeling techniques that only attempt to successfully align the protein backbone, the addition of full-atom representation and refinement to the backbone-Cβ structure returned by Rosetta is a very important factor for improving the accuracy and reliability of the modeling process (43).

Although the overall topology of a protein or a domain can be predicted very well, structure prediction by *ab initio* approaches generally has lower prediction accuracy than that of comparative modeling techniques when they are applicable (182). One strategy to improve the prediction accuracy is to apply partial experimental data as constraints in computational modeling. There are a number of studies that combine sparse NMR data with computational techniques, such as *ab initio* and threading, for protein structure prediction (40, 183). Effectively combining the data and modeling techniques can yield significantly better prediction results (183).

One method for generating experimental structural data for biophysical constraints is through the use of hydroxyl radicals to modify solvent exposed amino acid side chains. The sites of modification are determined by enzymatic digest, followed by mass spectrometry. Several previous reports of the successful use of this method to examine model proteins are available, utilizing high-flux X-rays (99, 101, 102), Fenton

chemistry (134), and photochemical oxidation (133) to generate hydroxyl radicals *in situ*.

These hydroxyl radicals then oxidize certain solvent accessible amino acid side chains,

preferentially oxidizing sulfur-containing residues, followed by aromatics, then followed

by histidine, proline, leucine, and isoleucine (100, 105, 106). Hydroxyl radical surface

mapping has also been utilized to examine protein conformation changes (165, 167),

largely-known protein structures (184), protein-protein interactions (166), and protein

folding (164). The rate of oxidation has been shown to be directly related to the solvent

accessible side chain area (102, 133) and the inherent rate constant of the oxidized amino

acid (106, 133), and the measurement is reproducible in different laboratories (102, 133).

However, no attempts have been reported to utilize hydroxyl radical solvent

accessibilities as constraints for computational modeling of a protein with no previously

solved three-dimensional structure.

Here, we present data from hydroxyl radical surface mapping experiments

examining the native solution structure of the C14S mutant of Sml1p. Hydroxyl radical

surface mapping measures the solvent accessibility of amino acid side chains. An

important feature of Rosetta nonlocal modeling is the burial of hydrophobic side chains,

and the use and refinement of full-atom representations are necessary for accurate and

reliable modeling by Rosetta (14). Therefore, critical analysis of the solvent accessibility

of amino acid side chains (especially hydrophobic amino acid side chains, which are

preferentially measured by our method) of the refined full-atom representations generated

from Rosetta backbone-Cβ models gives a measure of the overall accuracy of the full-

atom models. The solvent accessibility data generated are utilized, along with other

previously reported NMR data, to evaluate and improve pseudo-*ab initio* computational

81

models generated by the Rosetta algorithm. A computational model is presented which agrees with various experimental parameters.

Results

Oxidation of denatured C14S Sml1p

In order to determine which amino acids will be oxidized based solely on their inherent chemical reactivity and any effects stemming from the primary sequence, oxidation of C14S Sml1p was performed in the presence of 6M guanidine HCl. The denatured intact protein was examined by ES-FTMS to determine the overall amount of oxidation. The base peak represented C14S Sml1p plus the addition of eight atoms of oxygen; a smaller peak representing the addition of nine atoms of oxygen was also detected (data not shown). As Sml1p contains eight methionine residues, and methionine is known to be highly oxidized regardless of solvent accessibility (101, 133), this degree of oxidation is expected. No significant degradation products were detected.

In order to determine the exact sites of oxidation, we utilized tryptic digestion followed by C18 reverse phase LC-MS/MS to determine the exact side chains that were oxidized as described previously (133, 134). Amino acids should be oxidized largely based on the inherent rate constant of the corresponding free amino acid. In C14S Sml1p, we expect to detect oxidation events at Met3, Tyr9, Phe10, Tyr11, Met30, Phe33, Met41, Met46, Met53, Met66, Trp67, Met82, Phe 87, Met95, Phe96, Met104, and Phe106. This set of residues represents all amino acids with a rate constant for hydroxyl radical interaction above $6 \times 10^9\,M^{-1}\,s^{-1}$ in their free amino acid forms, and is indicative of the amino acids we normally observe in unfolded proteins. Generally, aliphatic amino

acids are not readily oxidized due to the limited amount of oxidation performed to preserve the structure of the protein.

After the oxidation, tryptic digestion, and LC-MS/MS analysis of denatured C14S Sml1p, we were able to unambiguously identify eighteen oxidation sites. Including both oxidized and unoxidized peptides, 100% sequence coverage for C14S Sml1p was obtained. The results of the oxidation of denatured C14S Sml1p are shown in Figure 19. Each of the amino acids expected to be oxidized was detected, with the exception of Tyr9. Tyr9 resides in a peptide directly adjacent to F10 and Tyr11, both of which were heavily oxidized (as estimated from fragment ion abundances in LC-MS/MS experiments); therefore, the failure to detect the oxidation of Tyr9 is probably due to the complex fragmentation pattern resulting from the MS/MS of three oxidation isomers differing in the site of oxidation by one residue. In addition to the amino acids expected to be oxidized, Pro22 and Leu25 were also found to be oxidized by LC-MS/MS; however, no other oxidation sites were detected.

Oxidation of native C14S Sml1p

Since we only failed to detect one highly oxidizable amino acid in the entire sequence, we judged the denatured data a sufficient baseline from which to perform analyses of C14S Sml1p in its native structure. The oxidized protein was measured by ES-FTMS to determine the extent of oxidation, and to help ensure that the protein was not denatured or fragmented by the oxidation process as described elsewhere (133). The native C14S Sml1p was oxidized to approximately the same extent overall that the denatured protein was; that is, the base peak was the addition of eight oxygen atoms, with

83

```
 1      *     10      *     20      *
GSMQNSQDYFYAQNRSQQQQAPSTL
RTVTMAEFRRVPLPPMAEVPMLSTQ
NSMGSSASASASSLEMWEKDLEERL
NSIDHDMNNNKFGSGELKSMFNQGK
VEEMDF
```

Red-native
Blue-denatured
Purple-both

Figure 19: Sites of Oxidation of Native and Denatures C14S Sml1p

The sequence of C14S Sml1p. The addition of a N-terminal Gly-Ser from the expression construct places the actual mutation site at Cys16→Ser16. Residues oxidized only in the native state are colored red, residues oxidized only in the denatured state are colored blue, and residues oxidized in both the native and denatured states are colored purple. Methionines are colored orange, and are not informative as they are oxidized regardless of solvent accessibility.

a smaller peak representing the addition of nine oxygens and no unmodified or degraded protein detectable (data not shown). The fact that denatured Sml1p did not show more oxidation events per protein molecule than the native Sml1p strongly suggests that saturating levels of hydroxyl radicals were not produced in the experiment; therefore, the denatured protein interacted with the same, limited number of radicals as the native, just with the oxidation events spread across more residues.

Tryptic digestion was then performed on the oxidized native C14S Sml1p sample, and the resulting peptides were analyzed by LC-MS/MS. The sites of oxidation of native C14S Sml1p are shown in Figure 19. All methionines were found to be oxidized in the native state. This finding is not surprising, as several previous reports have shown that methionine is oxidized regardless of the solvent accessibility of the residue (99, 101, 104, 134, 164-167, 184). However, several non-methionine residues that were found to be oxidized in the denatured C14S Sml1p were not detected in this assay (Pro22, Phe33, Phe107, and Phe106). These data, as well as the fact that we could detect and fragment the unoxidized version of the peptides, strongly suggests that Pro22, Phe33, Phe107, and Phe106 are buried in the native structure. The fact that these four residues are buried undermines a previous conclusion based on partial NMR data that Sml1p had no tertiary structure in solution (177).

The non-methionine amino acids that were oxidized in the native structure must be solvent accessible, as shown previously in several model systems by several oxidation techniques (101, 102, 134). In the native state, we were able to assign Tyr9 (although the signal-to-noise of the particular fragment ion required for this assignment was fairly poor); other than this single outlier, no other residues were detected in the native structure

that were not detected in the denatured structure. Tyr9, Phe10, Tyr11, Leu25, Trp67, and Phe96 were all shown to be oxidized in the native structure by LC-MS/MS. The data gathered on these ten residues were utilized to evaluate pseudo-*ab initio* computational models.

Computational modeling by Rosetta

As reported earlier, Sml1p does not exhibit significant sequence similarity with any protein structure in PDB, and fold-recognition methods failed to identify a structural fold with high confidence (150). Rosetta/I-sites, which is more applicable to novel structural fold and a top performer at CASPs in the new fold category, was applied to Sml1p structure prediction. The top five models generated by Rosetta for Sml1p are shown in Figure 20 A-E. All of these models have almost identical secondary structures (an α-helix near the N-terminus, and α-helix near the middle of the sequence, and an α-helix near the C-terminus); they differ mainly in the relative orientation and contacts between the N-terminal domain and the C-terminal domain. It has been shown that for about 40% of the examined proteins that are shorter than 150 amino acids, one of the five top models generated by Rosetta has good global similarity to the true structure (182).

Evaluation of computational models by surface mapping analysis

After the computational models were generated, each was evaluated by application of the surface mapping constraints. For each model, the product of the solvent accessible reactive area and the respective free amino acid rate constant was plotted for each of the known reactive amino acid residues in C14S Sml1p (both for the
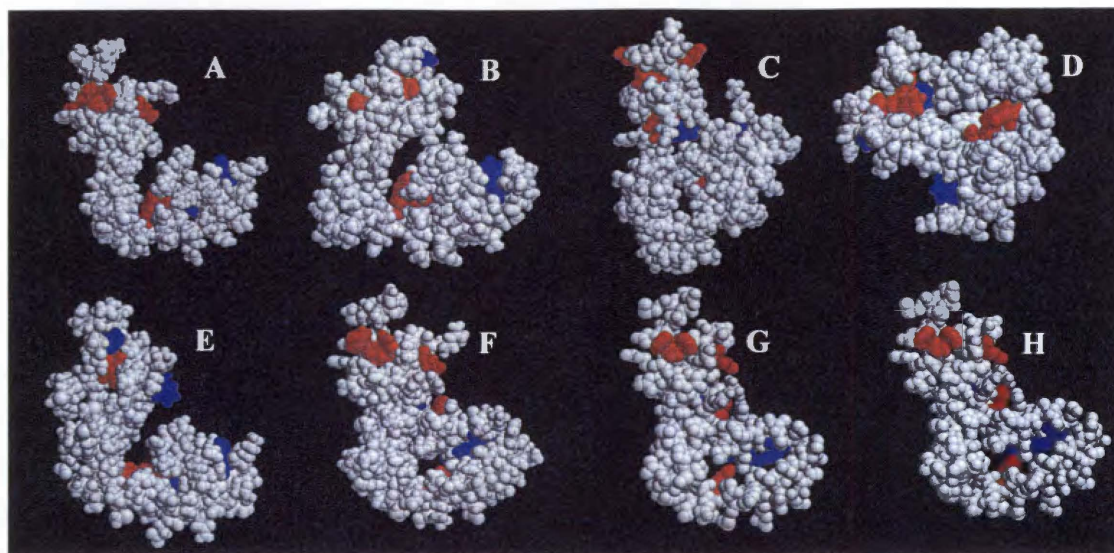
Figure 20: Oxidation Sites Plotted on Computational Models of Sml1p

Spacefill representations of each of the eight computational models discussed. Amino acids that are oxidized only in the denatured state (and therefore buried in the native structure) are colored blue. Amino acids that are oxidized in the native state (and therefore are accessible in the native structure) are colored red. 3A-Rosetta model 1; 3B-Rosetta model 2; 3C-Rosetta model 3; 3D-Rosetta model 4; 3E-Rosetta model 5; 3F-partial NMR Sml1p model; 3G-Sml1p rotamer 1 model; 3H-Sml1p rotamer 2 model

denatured and native experiments). Given that the model is correct, the product of the side chain area and the amino acid rate constant would estimate the relative reactivity.

The comparison of each of the computational models is shown in Figure 21. In the case of a correct model, one should be able to draw a line at some value along the ordinate below which lie all (area x rate) products for amino acids which were only oxidized in the denatured state (and thus are "buried" in the native structure), and above which lie all (area x rate) products for amino acids which were oxidized in the native state (and thus are "exposed" in the native structure). The ordinate value at which this line is drawn will depend upon the overall extent of oxidation and the sensitivity at which the oxidized peptides are detected; therefore, this value will differ from experiment to experiment. The important factor is not the value of this discriminating line, but rather the comparative value of the (area x rate) product for different amino acids within the same protein in the same experiment. All assignments of "buried" and "exposed" are relative to other amino acids within the same protein (i.e. when F10 is assigned as "exposed", the actual meaning is "more exposed than any buried residue"). Analysis of the computational models presented yielded a discriminating line for one computational model, as discussed below. Based on an analysis of the fit of each computational model with the solvent accessibility data gathered previously, the accuracy of each model can be critically evaluated against known biophysical properties of the native structure.

The first of the top five models generated by Rosetta is shown in Figure 20A. In examination of the surface mapping constraints, several residues look reasonable. Tyr9 and Tyr11 are both highly solvent exposed, which is supported by the surface mapping data. In addition, this model buries Pro22, Phe33, and Phe106, which are all known to be
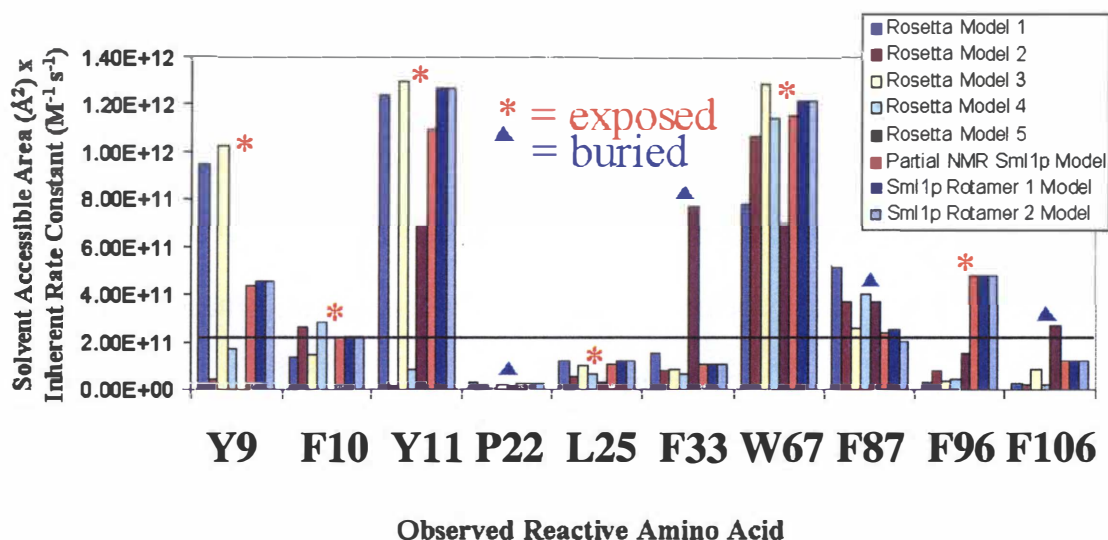
Figure 21: Quantitative Analysis of Sites of Oxidation of C14S Sml1p

The product of the solvent accessible reactive area and the respective free amino acid rate constant (given in Table 1) for each of the ten reactive, non-methionine residues in C14S Sml1p. Residues that were oxidized in the native state are marked with a red asterisk (exposed), and residues that were oxidized *only* in the denatured state are marked with a blue triangle (buried). An empirical discriminatory threshold separating oxidized and unoxidized residues is indicated by a horizontal line at 2.1×10^{11}. This line effectively separates exposed vs. buried residues for the Sml1p Rotamer 2 model.

buried in the native structure. However, this model has Phe10 fairly buried, and Phe96 is nearly completely buried. Both of these residues were shown to be solvent accessible in the native structure. Also, this model exposes Phe106, which is known to be buried in the native structure. These discrepancies eliminate Rosetta model 1 as a potential correct structure for Sml1p.

The second model generated by Rosetta is shown in Figure 20B. Comparison of this model with surface mapping constraints is even less promising than the first model. Both Tyr9 and Tyr11 are almost completely buried, as is Phe96. All of these residues should be solvent accessible. Trp67 is almost completely solvent accessible, which agrees with our data as well as with other tryptophan fluorescence data (150). This model does bury Pro22, Phe33, and Phe106, which is in agreement with the data. However, the buried aromatic residues near the N-terminus, as well as the buried Phe96 and the exposed Phe87, all eliminate Rosetta model 2 from consideration as a correct structure of Sml1p.

The third Rosetta model (Figure 20C) looks considerably better than the first two. Tyr9 and Tyr11 are both highly solvent exposed, as is Trp67. Pro22, Phe33 and Phe106 are all buried, and Phe87 is more buried in this model than in either of the previous ones. However, Phe10 is still relatively buried; its (area x rate) product is lower than that of Phe87, and yet Phe10 is known to be solvent accessible while Phe87 is known to be buried. In addition, this model still has Phe96 buried, which is at odds with the experimental surface mapping constraints. Therefore, Rosetta model 3 must also be rejected as a potential structure, although it does agree better than any of the other Rosetta models thus far.

The fourth model generated by Rosetta, Figure 20D, reveals immediate difficulties at the N-terminus. Tyr9 and Tyr11 are both buried in this model, as is Phe96. All of these residues were shown experimentally to be solvent accessible in the native structure. Phe10 is more solvent accessible in this structure than in any of the others so far, Trp67 is considerably exposed, and Phe33 and Phe014 are properly buried. However, Phe96 is incorrectly buried in this model, and Phe87 is too exposed. All of these disagreements between Rosetta model 4 and the experimental surface mapping data exclude this model from further consideration.

The fifth and final model generated by Rosetta, shown in Figure 20E, has incorrectly buried Tyr9 and Phe10 completely. Tyr11 and Trp67 are somewhat exposed in this model, and Phe96, while still buried, is more exposed than in any of the other Rosetta models. However, Phe33 is highly solvent accessible in this model, as is Phe87 and Phe106, when all three are known to be buried. Therefore, Rosetta model 5 must also be excluded from consideration as a correct structure for Sml1p.

All the five models generated by Rosetta for Sml1p have two α-helices (one near the N-terminus and one near the center of the sequence) that are consistent with the NMR data (177). However, each of the Rosetta models also has another α-helix near the C-terminus, contrary to the NMR data. Therefore, a hybrid model was constructed using segments consistent with NMR data (150). This model, shown as Figure 20F, was examined utilizing the solvent accessibility data generated here. This model looks considerably better than any of the models generated by Rosetta. Tyr9 and Phe10 are both moderately exposed, with Tyr11 highly solvent accessible. Phe33 has been properly buried in this model, and Trp67 is almost completely solvent exposed, in agreement with

91

our data and with the previous fluorescence quenching and anisotropy data (150). Phe106 is also correctly buried; however, the most dramatic change is in Phe96, which has become approximately three times more solvent exposed than in the most favorable Rosetta model. A minor difficulty still arises; the (area x rate) product of the exposed Phe10 is slightly lower than the (area x rate) product of the buried Phe87. This problem could be readily resolved by either exposing Phe10 or by burying Phe87; however, even small local changes may impact the solvent accessibilities of several other residues.

In order to improve the fit of the model with the experimental surface mapping data, different rotamers of selected side chains from the partial NMR Sml1p model were analyzed and compared with the surface mapping data. Sml1p rotamer 1 model (Figure 20G) features a lower-energy rotamer for the Arg26 side chain. The overall solvent accessibility profile is similar to the parent model, with a few minor differences. Tyr9 and Phe10 are both slightly more solvent accessible, and Tyr11 is much more solvent accessible. Trp67 is also slightly more solvent accessible. Unfortunately, Phe87 is also slightly more solvent accessible than the parent model, and as such its (area x rate) product is still slightly higher than the exposed Phe10.

The Sml1p rotamer 2 model (Figure 20H), which in addition to the altered Arg26 side chain rotamer also uses a lower energy rotamer for the Phe87 side chain, has an absolutely identical solvent accessibility profile as the Sml1p rotamer 1 model, with one important exception. The solvent accessibility of Phe87 has dropped slightly, lowering its (area x rate) product below that of the exposed Phe10. This model (with the exception of Leu25, discussed below) has no conflicts with the solvent accessibility data. A discriminating line can be drawn at an (area x rate) product of 2.1×10^{11} $Å^2$ M^{-1} s^{-1}; below

this line lie the (area x rate) products of all residues known to be buried, and the (area x rate) products of all residues known to be exposed lie above this line (with the exception of Leu25). However, the differences between the partial NMR Sml1p model, the Sml1p rotamer 1 model, and the Sml1p rotamer 2 model are very small, and the three cannot be confidently discriminated by solvent accessibility data alone. Therefore, while we cannot say with certainty that any of the three models are correct or incorrect, our results and calculations suggest that the Sml1p rotamer 2 model is a likely candidate for the native structure of Sml1p.

In the discussions of all of the computational models above, one apparent conflict in each model has not yet been addressed—the (area x rate) product of Leu25 is low in each model, and yet the data shows that the amino acid is oxidized in both the native and denatured state. The main difficulty is that, even if Leu25 were to be completely exposed, the (area x rate) product would only be approximately $2.1 \times 10^{11} \, \text{Å}^2 \, \text{M}^{-1} \, \text{s}^{-1}$, comparable to that of Phe10. In addition, the data generated from the denatured C14S Sml1p showed that both Leu25 and Pro22 were oxidized, while other leucines and prolines were not. Together, these data suggest that there may be characteristics in the primary sequence near these two residues that increase their rate of reactivity *or their ease of detection*. The fact that these two amino acids were detected in the denatured experiment while none of the other aliphatic amino acids were detected, even though these amino acids have comparable reactivities in a denatured protein, suggests that this peptide was detected with unusually high sensitivity and fragmented efficiently in the LC-MS/MS experiment. In addition, previous results have shown that the rate of oxidation varies directly with solvent accessibility; no neighbor effects have been

93

reported (101, 102, 133). Figure 22A shows the total ion chromatogram for the oxidized and unoxidized singly-charged 16-26 peptide, which contains both Pro22 and Leu25. The ratio of the base peak abundance of the oxidized 16-26 peptide to the unoxidized 16-26 peptide (which serves as a rough estimate of the rate of oxidation) is 0.0292. The low ratio is due to a very low abundance of the oxidized 16-26 peptide, which also is evident in the very noisy oxidized peptide selected ion chromatogram. In contrast, a peptide with an exposed phenylalanine (peptide 94-100) shows a base peak oxidized:unoxidized ratio of 0.4278. This peptide's selected ion chromatogram shows clear peaks representing the unoxidized and the oxidized peptide (Figure 22B); this ratio is similar for all other oxidized/unoxidized peptide pairs detected. Therefore, we have chosen to disregard Pro22 and Leu25, and hinge our analyses on the aromatic amino acids, all of which have similar reactivities (105, 106), and all of which showed reasonable ratios of oxidized/unoxidized peptide in the selected ion chromatograms (Figure 22B).

Discussion

The regulator of ribonuclease reductase Sml1p is a highly important regulatory protein, regulating intracellular levels of dNTPs and thus important in maintaining genetic stability (185). Sml1p is known to interact with Rnr1p in a 1:1 stoichiometry (172), and this interaction is known to take place via the C-terminus (177). Significant amounts of work have been performed in order to elucidate the structure of Sml1p. Previous NMR work performed showed that Sml1p has two alpha helices comprising residues 6-16 and 63-82 (177), which were added to the Rosetta model.
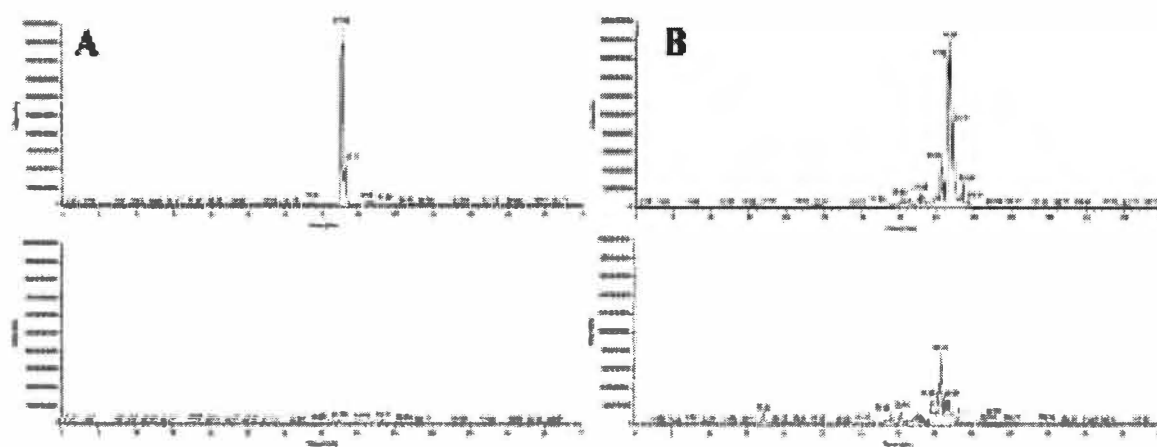
Figure 22: Selected Ion Chromatograms Showing Rates of Oxidation

Selected ion chromatograms (full scan MS only) from an LC-MS/MS run for oxidized and unoxidized versions of two peptides. A—**(top)** the unoxidized, singly-charged version of peptide 16-26; **(bottom)** the oxidized, singly-charged version of peptide 16-26. B—**(top)** the unoxidized, singly-charged version of peptide 94-100; **(bottom)** the oxidized, singly-charged version of peptide 94-100. For both A and B, both top and bottom panels are normalized to the same absolute signal intensity for direct comparison.

These helices were supported by gas phase collision of intact Sml1p, which showed that these regions were protected from fragmentation in the gas phase, suggesting secondary structure (151). The NMR work also suggested that the protein has no stable tertiary structure; however, the data generated here clearly demonstrate that a solvent-excluding tertiary structure exists for Sml1p.

Thorough random mutagenesis of *SML1* was also performed to determine which residues were essential for Sml1p activity. All of the mutants that were found to abolish the inhibition of ribonuclease reductase (Rnr1p) were found to be at the C-terminus. In addition, deletion of the N-terminus revealed that the C-terminal 54 amino acids were necessary and sufficient to maintain the inhibitory activity of Sml1p (177). Also, analysis of the unfolding profile of Sml1p and C14S Sml1p show identical unfolding transitions (150). Therefore, the substitution of a serine at position 14 for a cysteine should not substantially alter the structure or function of Sml1p.

It has been shown previously that Sml1p can form a stable dimer in solution, either covalently bound via a disulfide bridge or, in the case of C14S Sml1p, by noncovalent interactions (150). However, a previous report based on NMR spectroscopy did not report any dimerization (177). With the conflicting reports, any analysis of Sml1p solvent accessibility must take the possibility of dimerization into account. However, previous analysis of another noncovalently-bound structure, R67 dihydrofolate reductase, suggests that the photochemical oxidation technique may not probe all protein-protein interactions under the conditions and time-scales utilized here (unpublished results). In addition, the N-terminal 38 amino acids of C14S Sml1p are necessary for homodimerization (150). In the solvent accessibility analysis of C14S Sml1p, only P22

96

and F33 are buried in the N-terminal 38 amino acids, and neither of these amino acids were solely responsible for the elimination of a computational model. From the top five models generated by Rosetta, not one was excluded from consideration solely on the basis of buried residues, and it is unlikely that dimerization causes a residue to become exposed. Regardless, during analysis of computational models utilizing solvent accessibility constraints, it is important to note that residues that are found by surface mapping to be buried may in fact be buried by a protein-protein interaction, not by any tertiary structure.

As a general method, the application of surface mapping constraints to Rosetta models has great promise. Rosetta has performed very well in the recent CASP4 competition, successfully modeling large segments of most of the domains for which models were submitted. Rosetta was especially well suited for the analysis of Sml1p, as this protein would not thread correctly against any template structure, making modeling by protein threading impossible. As a pseudo-*ab initio* modeling algorithm, Rosetta does not require a folding template, and can attempt to predict new folds such as Sml1p. However, the modeling algorithm is imperfect, and utilized a number of artificial constraints as an attempt to produce valid models in a reasonable timeframe. Several of these procedures can be directly influenced by solvent accessibility data. For example, Rosetta utilizes an energy function that buries hydrophobic residues. In addition, a filter is used which eliminates structures based on solvent-accessible surface calculations (186). As can be determined from Table 1, many of the most highly oxidizable side chains are hydrophobic in nature, which allows the data to be used to successfully defy

the pseudo-*ab initio* calculations by exposing residues that the energy functions of Rosetta wish to bury.

While the approach taken with the analysis of Sml1p is a valid method for eliminating incorrect structures, better results would be achieved if the constraints were built into the modeling algorithm itself. Several model proteins have been analyzed by hydroxyl radical surface mapping (133, 134). Current efforts are being made to generate a constraining term for the protein threading algorithm PROSPECT, in an attempt to improve threading template selection and energy minimization (42). Once the solvent accessibility data can be inserted within a threading algorithm, a direct pipeline between surface mapping experiments and constrained computational structures can be established for pseudo-high throughput constrained structure prediction, a tremendous goal for systems biology.

CHAPTER 6—ANALYSIS OF THE STRUCTURE OF Aβ 1-40 AMYLOID FIBRILS BY PHOTOCHEMICAL OXIDATION AND MASS SPECTROMETRY

Introduction

The elucidation of the structures of amyloid fibrils is an intriguing and very important problem due not only to the fact that amyloid fibrils and related aggregates are implicated in over 20 diseases (187), but also as an alternative folding pathway as a β sheet-rich protein aggregate (188). It has also been shown that amyloid fibrils can actually serve a beneficial function, such as aiding in maintaining evolutionary variability (189); in addition, cellular machinery has been found responsible for creating and maintaining amyloid fibrils in *E. coli* (190). Amyloid fibrils are known to be fairly straight structures of approximately 80-120 Å in width with lengths up to several μm, with a substructure of bundled protofilaments of 20-30 Å (187).

The secondary structure of amyloid fibrils has been characterized by a variety of methods. X-ray fiber diffraction has given some important insight into the nature of the secondary structure of amyloid fibrils. It has been determined that the fibril consists of multiple β sheets perpendicular to the axis of the fibril (191). In addition, X-ray diffraction has suggested that the multiple β sheets are stacked, as opposed to polyglutamine fibrils, which may have an extended helical conformation (187, 192, 193). However, currently two models of amyloid fibril structure are favored; the hollow,

extended conformation β helical structure proposed by Perutz *et al* (192) and the parallel stacked β sheet helical structure proposed by Wille *et al* (194).

Perutz and colleagues recently published an analysis of various X-ray diffraction and electron microscopy data on a variety of fibril aggregate forms. Perutz *et al* argued that, due to the known β sheet composition of the fibrils as well as the lack of a 10 Å reflection in the polyglutamine diffraction pattern, that the polyglutamine fibril forms a cylindrical β sheet, with each peptide in an extended β strand conformation. The resulting structure, shown in Figure 23, would have an inner pore of about 11.8 Å diameter—too small for many dyes, but large enough for the diffusion of unbound water (192), and also large enough for the diffusion of hydrogen peroxide and hydroxyl radicals. Since half of the side chains of the fibril will be facing this inner, solvent accessible core, all of these amino acids should be available for oxidation by hydroxyl radicals.

In a study of the amyloid-like prion aggregate PrP^{Sc}, electron microscopy studies found a hexagonal crystal in fibril preparations of the infection prion (194). The authors contend that the hexagonal crystal represents the cross-section of the fibril; furthermore, due to evidence from spectroscopic studies of internal deletions of the PrP^{Sc} prion, the authors contend that the only known structural element that could make up the fibril is the parallel β helix (Figure 24). One major structural difference between this parallel β helix and the structural model presented for polyglutamine fibrils and proposed to also apply to Aβ amyloid fibrils (192) is that the parallel β helix model proposed by Wille *et al* would exclude solvent from the interior of the protofilament, while the model of Perutz would require a protofilament filled with unbound water (187).

100
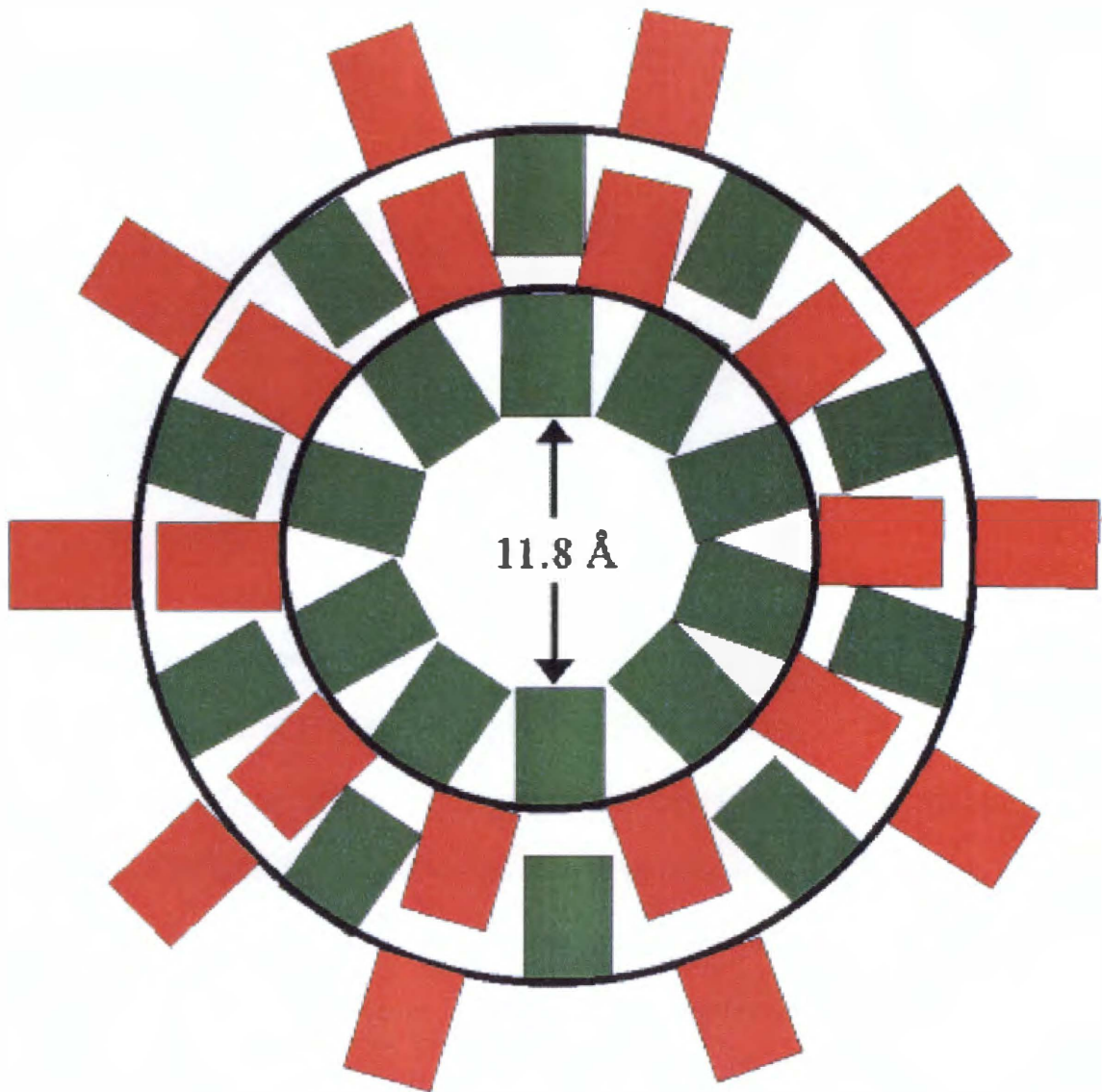
Figure 23: Amyloid Nanotube Model for Alzheimer's Aβ Peptide

A model adapted from Perutz *et al* of the structure of the Aβ fibril (192). Perutz *et al* stipulated a possible third concentric helix outside of the first two. Even given a 3Å layer of bound water within the nanotube, there is still a 6Å diameter of unbound water; this diameter is too small for many chemical dyes and labels, but small enough for water, hydrogen peroxide, and hydroxyl radicals.
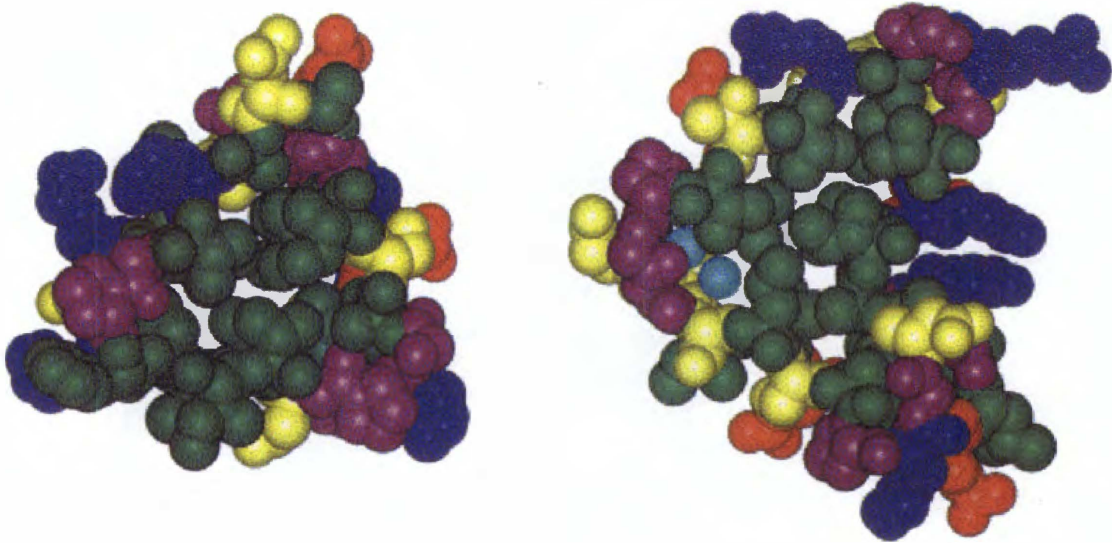
Figure 24: Spacefill Models of a Left-Handed and Right-Handed Stacked β Helix

Figure reproduced from Wetzel with permission (187). The left panel shows a top-down view of two rungs of a left-handed stacked β helix (PDB accession ID 1lxa); the right panel shows a top-down view of two rungs of a right-handed β helix (PDB accession ID 1dab) with amino acids colored to show chemical similarity (hydrophobics green, neutral hydrophilics yellow, acidics red, basics blue, Gly and Pro purple, and water light blue). Note that the central region of the helix is filled in with side chains.

Since these two leading models are differentiated by a solvent accessible core, the application of a technique that measures the solvent accessibility of a variety of amino acids would be useful to eliminate one of the two models. The hydroxyl radical labeling method is ideal for this problem. Not only does the radical label a wide variety of amino acid side chains, but the small size of the radical would allow it to penetrate into the hollow center of the protofilament if the model of Perutz *et al* is correct. Therefore, photochemical hydroxyl radical surface mapping was performed on the Aβ 1-40 peptide as well as the amyloid fibril generated by the Aβ 1-40 peptide. If the model of Perutz *et al* is correct, half of the peptide involved in the protofilament should be accessible to the solvent within the core of the nanotube (192), while if the model of Wille *et al* is correct, only the amino acids on the solvent accessible, outside face of the protofilament would be solvent accessible.

Results

In order to test the model of Perutz *et* al, the photochemical surface mapping technique previously developed was applied to the Aβ fibril. Aβ 1-40 peptide, shown in Figure 25, was photochemically oxidized as described in Chapter 2 in both the monomer and the fibril state. In order to determine the potential sites of oxidation, LC-MS/MS of the monomer was performed in triplicate. Table 8 shows the sites of oxidation and the number of replicates in which each site was identified. Oxidation at the amino acids L17, F19, F20, and M35 were detected in each replicate for the monomer, and were treated as the most reliable datapoints in fibril analysis. Oxidation at I32 and L34 were detected in

103

```
1     *    10     *    20     *    30     *    40
DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
```

Figure 25: Sequence of Aβ 1-40 Peptide

Colored red is peptide 6-16, which could not be detected in either the monomer or the fibril by LC-MS/MS or ES-FTMS. Colored blue are amino acids that were detected by LC-MS/MS in at least two of the three monomer oxidation replicates, but not in the fibril oxidation. Colored orange is methionine, which is detected as oxidized in both the monomer and fibril. No solvent accessibility data can be determined from the oxidation of methionine (101, 133, 134).

Table 8: Sites and Apparent Rates of Oxidation for Aβ 1-40 Monomer and Fibril

Peptide	[Oxidized]/[Unoxidized] Monomer	[Oxidized]/[Unoxidized] Fibril	Oxidized Residues
1-5	0.05*	0	F4
6-16	ND	ND	ND
17-28	0.08	0	F19, F20, L17
29-40	100% oxidized (0.06^)	100% oxidized	M35, I32, L34

ND—peptide not detected as oxidized or unoxidized

*--very poor signal to noise

^--ratio of methionine oxidation to I32 and L34 oxidation ([+ 1O] + [+ 2O]) / [+ 3O]

Underlined residues were identified in all three replicated by LC-MS/MS measurements; others were identified in two of three replicated except for F4, which was detected in one replicate and strongly implicated by ES-FTMS measurements.

two out of the three replicates, both at low signal to noise levels. Oxidation at F4 was detected in only one replicate at very low signal to noise levels; however, the oxidized peptide seems to be present at low abundance in all replicates of the LC-MS/MS experiment, just without a good MS/MS fragmentation spectrum. Peptide 1-5 did not ionize well in any experiment (either LC-MS/MS or ES-FTMS), causing even the unmodified peptide to be difficult to detect; therefore, detection of the oxidized peptide 1-5 was challenging in even the best cases. Peptide 6-16 could not be detected in the unoxidized or oxidized state in the monomer or the fibril. Peptide 6-16 has three histidines as well as a tyrosine present. Each histidine could have at least three major oxidation products, resulting in the signal from peptide 6-16 to be split into at least 27 peaks (ignoring the contribution from the tyrosine). Since the signal from one peptide is split into so many peaks, it is probable that the many peaks from this peptide are at too low signal to noise to detect by mass spectrometry. It is also possible, although improbable, that peptide 6-16 was not retained on the C18 SepPak used in cleaning up the sample after trypsin digestion; however, nothing about the peptide's sequence suggests a difficulty with solid phase extraction. It is known that at least H6 is solvent accessible in the fibril, and more recent mutagenesis experiments have suggested that H13 and H14 are also part of the solvent accessible tail (Wetzel, personal communication), suggesting that the oxidation of all three histidines is likely in both the monomer and the fibril. Oxidation of the Aβ 1-40 fibril was performed, and the rates of oxidation of each of the four peptides were compared with those of the monomer. The major oxidation sites of the monomer as estimated from the MS/MS fragmentation spectra were F19, F20, and M35, with very minor oxidation occurring at L17 and no relative rate estimates available

for F4. The rates and sites of oxidation of the fibril are shown in Table 8. The only oxidation site detected was M35. The methionine-containing peptide was not detected in the unoxidized state, with essentially the same oxidation rate for the dioxidized peptide (representing the methionine sulfone) as the monomer. As previously mentioned, oxidation of methionine can occur regardless of side chain solvent accessibility (101, 102); however, essentially no methionine oxidation occurs in the absence of the UV-induced hydrogen peroxide photolysis. Therefore, the measured oxidation of methionine at near the same rate as in the monomer shows that the oxidation reaction did occur as expected (as opposed by a true negative result). However, no other sites of oxidation were detected in the fibril.

Figure 26 shows the ES-FTMS relative quantitation of the [oxidized] / [unoxidized] ratio of peptide 1-5 in the monomer (Figure 26A) and the fibril (Figure 26B). As can be seen in the spectra, the signal to noise of peptide 1-5 is poor even in the native form, and the oxidized form is barely detectable in the monomer. The signal to noise is considerably worse in the fibril spectrum, probably due to the emergence of chemical noise from the TFA treatment required to dissociate the fibril into monomers after oxidation, as well as sample loss during the centrifugation and washing. Since the chemical noise is significantly higher in the fibril spectrum, we were unable to detect oxidized peptide 1-5. The lack of data at the N-terminus, while unfortunate, does not detract from the overall knowledge of Aβ 1-40 structure, as the N-terminal 12 residues have already been determined to be solvent accessible by partial proteolysis (195).
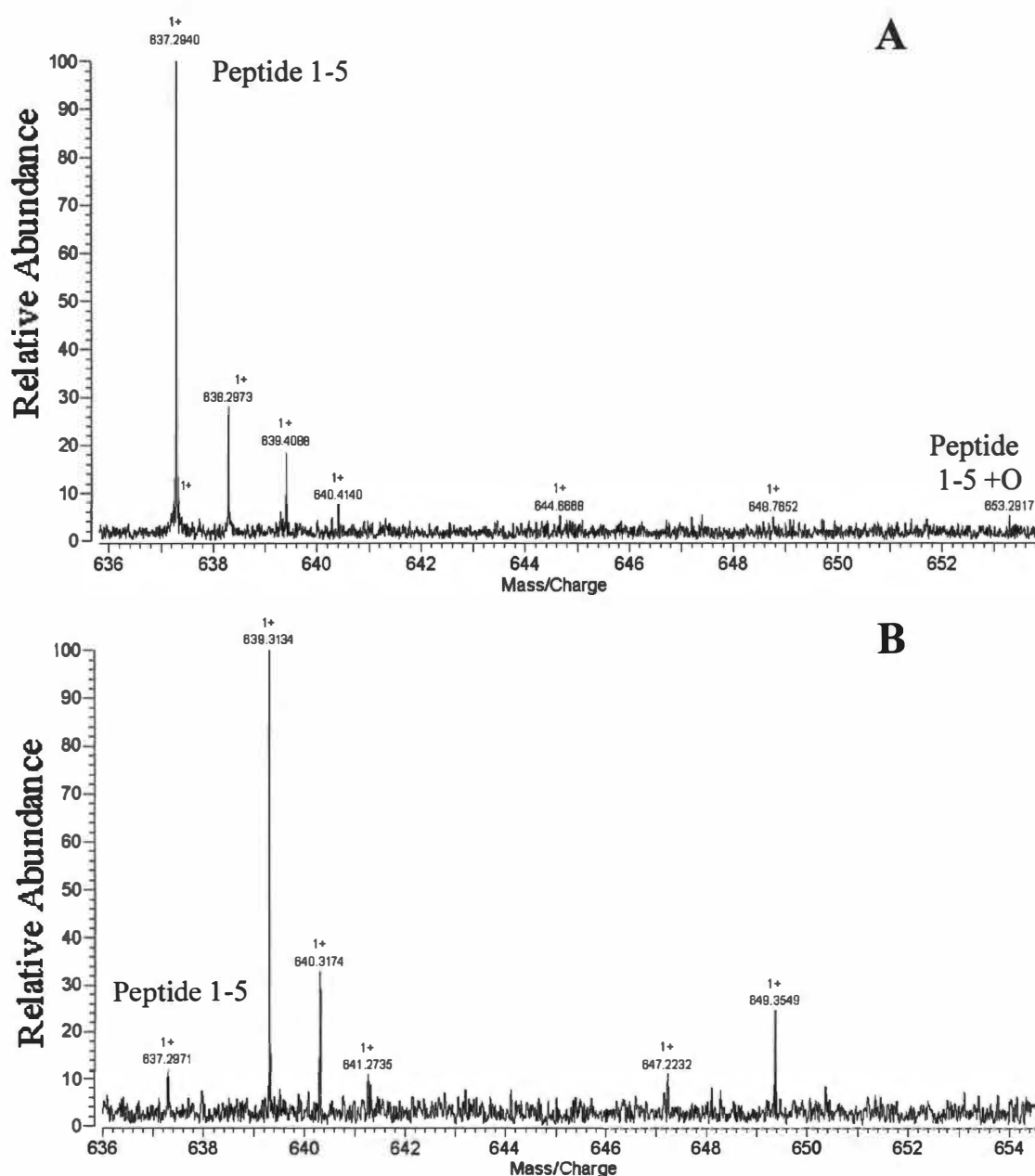
Figure 26: ES-FTMS of Peptide 1-5 Oxidation

Spectrum of peptide 1-5 oxidation in the **(A)** monomer and **(B)** fibril. The oxidized

peptide 1-5 can barely be detected in the monomer, but cannot be detected in the fibril,

possibly due to significantly worse chemical noise in the fibril spectrum. The only

oxidizable residue in peptide 1-5 is F4.

Figure 27 shows the ES-FTMS relative quantitation of the [oxidized] / [unoxidized] ratio of peptide 17-28 in the monomer (Figure 27A) and the fibril (Figure 27B). As the spectra show, the signal to noise of this peptide is very good, which allows even low oxidation levels to be detected reliably. Figure 27A shows that peptide 17-28 is oxidized to a detectable extent, with an [oxidized] / [unoxidized] ratio of 0.08 (Table 8), with oxidation occurring at F19, F20, and to a much lesser extent, L17. Figure 27B shows the oxidation of peptide 17-28 in the fibril. The signal from peptide 17-28 in the fibril still shows good signal to noise; however, no signal from an oxidized peptide can be detected. Even in the LC-MS/MS spectra, no evidence of oxidized peptide 17-28 can be found in the fibril. If only one of the two phenylalanines was exposed to solvent in only half of the peptide, as in the concentric continuous β sheet helix model of Perutz *et al* (192), we would expect an oxidation rate approximately 25% of that of the monomer. At the signal to noise levels obtained for this peptide in the fibril, an [oxidized]/[unoxidized] ratio of ~0.02 would be detectable; however, no such oxidation was seen. Therefore, it must be concluded that both phenylalanines are buried to a significant extent in the fibril.

The oxidization of peptide 29-40 as determined by ES-FTMS measurements is shown in Figure 28. Figure 28A shows the oxidation states of peptide 29-40 in the monomer. As shown in the figure, there is no detectable unmodified peptide. This is not uncommon for methionine-containing peptides. The peptide 29-40 + O peak represents the M35 sulfane as determined by LC-MS/MS fragmentation. The peptide 29-40 + 2O peak represents the M35 sulfone, again as determined by LC-MS/MS fragmentation. The peptide 29-40 +3O peak represents the M35 sulfone plus the oxidation of I32 or L34. In the ES-FTMS spectrum of the oxidized fibril (Figure 28B), the oxidation of M35 to both

109
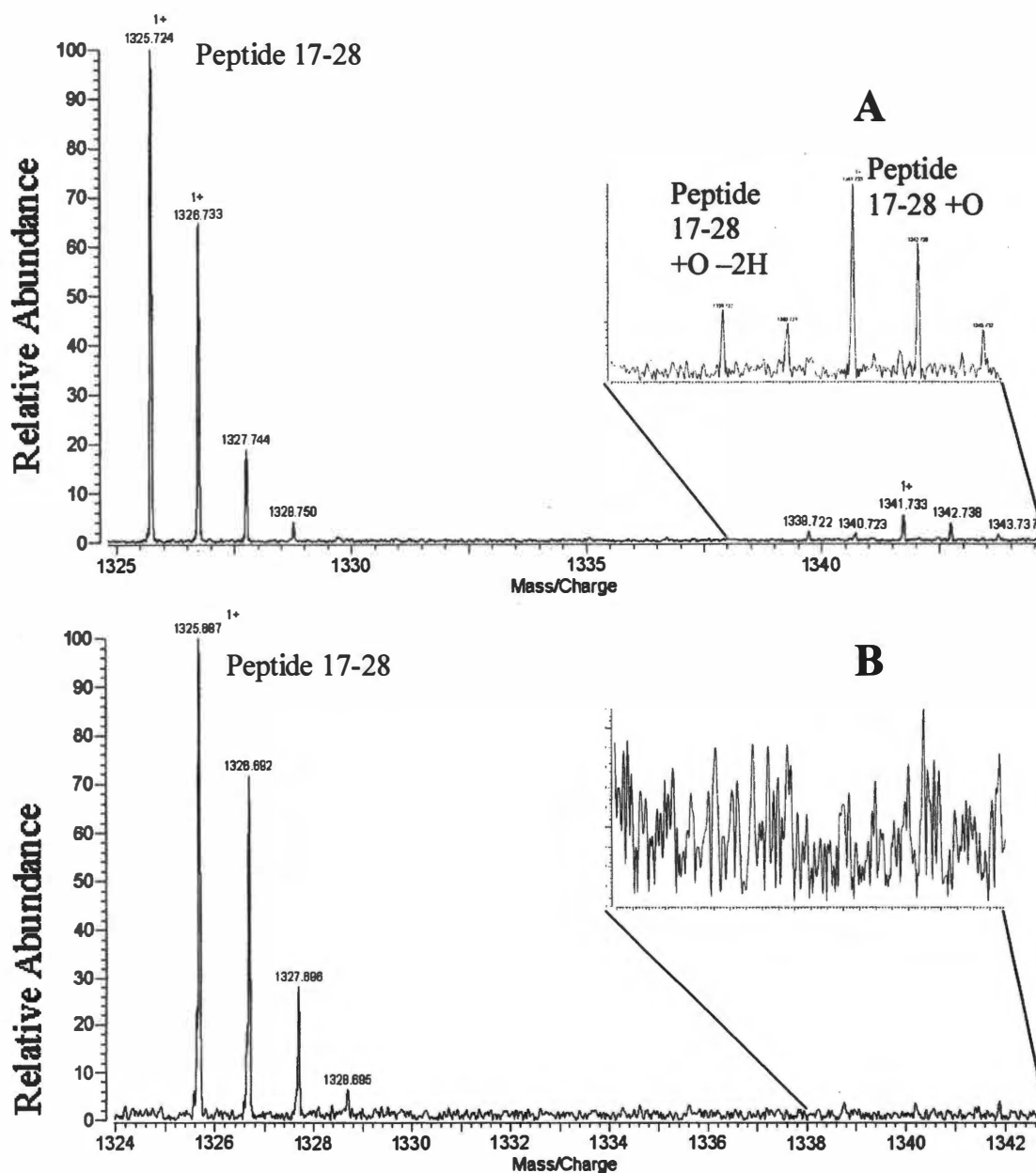
Figure 27: ES-FTMS of Peptide 17-28 Oxidation

Spectrum of peptide 17-28 oxidation in the **(A)** monomer and **(B)** fibril. The oxidized

peptide 17-28 can be detected in the monomer, but not in the fibril, even though the

signal to noise is excellent in both spectra. The major sites of oxidation in this peptide

are F19 and F20. The insert shows the m/z range of 1338-1344 expanded for detail.
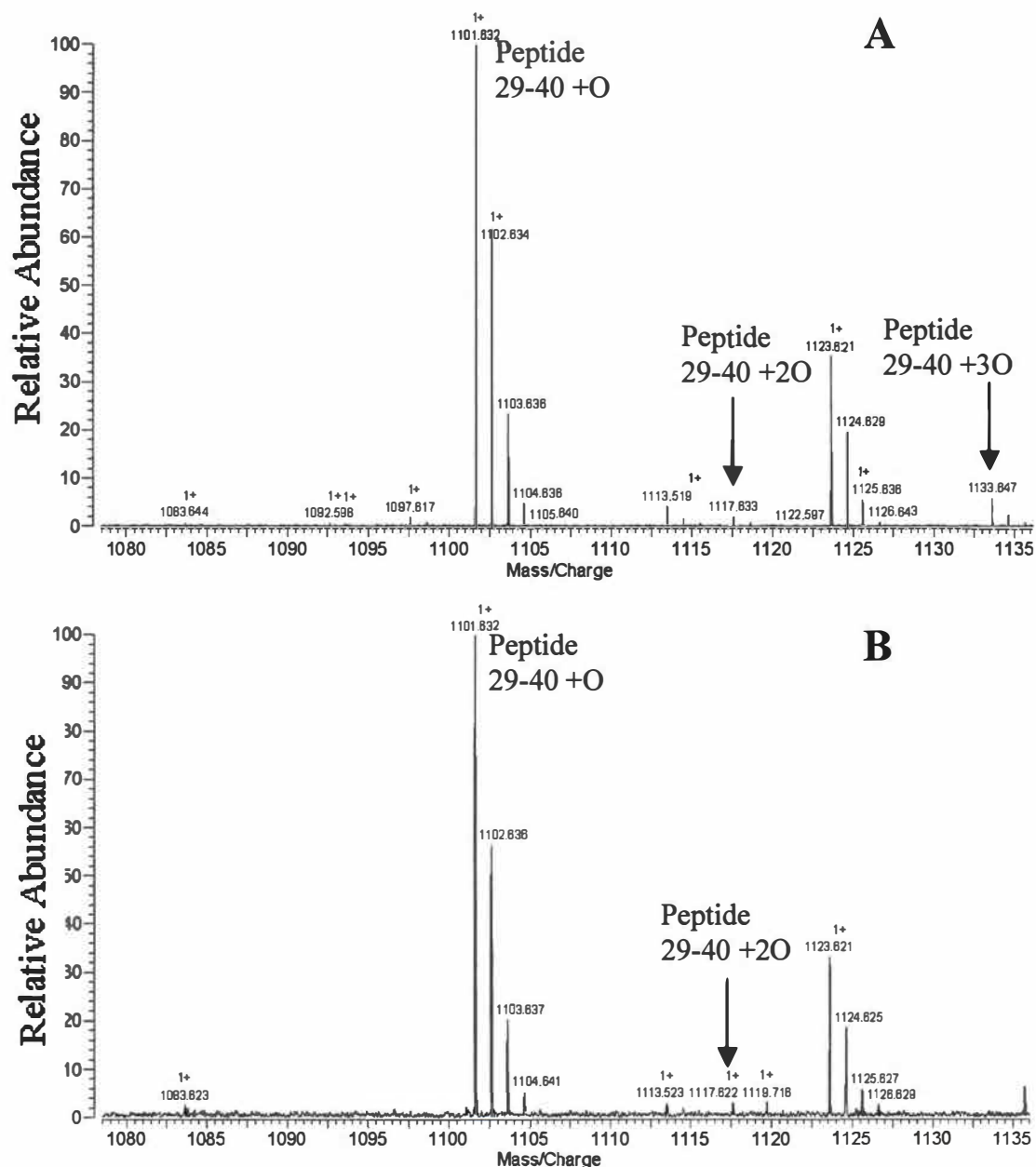
Figure 28: ES-FTMS of Peptide 29-40 Oxidation

Spectrum of peptide 29-40 oxidation in the **(A)** monomer and **(B)** fibril. The unoxidized

peptide 29-40 is not detected in the monomer or the fibril, even though the signal to noise

is excellent in both spectra. Peptide 29-40 + O and peptide 29-40 + 2O both represent

oxidation at M35; peptide 29-40 + 3O represents oxidation at M35 and I32 or L34.

111

the sulfane and the sulfone is present at essentially the same apparent rate (Table 8); however, the peptide 29-40 + 3O peak is not present. The signal to noise ratio of the peptide 29-40 + 3O peak in the monomer is large enough to be easily detected in the fibril if no change in the solvent accessibility of I32 and L34 occurred; however, no 29-40 + 3O peak can be detected in the ES-FTMS spectrum of the oxidized fibril. In addition, no evidence of oxidation of I32 or L34 can be detected in the LC-MS/MS spectra taken of the oxidized fibril. Therefore, the amount of overall hydroxyl radical production during the oxidation reaction appears to be comparable between the fibril reaction and the monomer reaction (due to the same apparent rate of oxidation of M35 into both the sulfane and the sulfone forms). In addition, the solvent accessibility of I32 and/or L34 must be decreased in the fibril as compared to the monomer. Since it cannot be determined with great accuracy to what extent each residue contributed to the oxidation of peptide 29-40 in the monomer, it is possible that the loss of solvent accessibility to one of the sites may be sufficient to lower the signal of the peptide 29-40 + 3O peak below the signal to noise ratio threshold necessary for detection in the oxidized fibril sample; however, both residues must be located on the same face of the β-strand, suggesting that a structure which buried one of them probably would bury the other as well.

Discussion

Photochemical oxidation surface mapping is an excellent tool for examining multiple amino acid side chains in wild type proteins with a single experiment. Amyloid fibrils are a very difficult protein to examine using traditional structural methods. The

112

fibrils will not crystallize, making X-ray crystallography unusable, and they are much too large for examination by solution-phase NMR spectroscopy. Much work has been done with X-ray diffraction (191, 193, 196, 197) and with solid-state NMR spectroscopy (198-207). The models proposed for the structure of the Aβ amyloid fibril are often in disagreement with one another. One such model recently proposed by Perutz *et al* (192) states that Aβ amyloid fibrils are composed of two or three concentric continuous β helices, with an unbound water-filled pore in the middle, leading to a nanotube-like structure. However, another model of amyloid structure recently proposed by Agard *et al* suggest a hexagonal array of cross-sectional slabs of peptide stacked upon one another, with a solvent-excluding central pore.

The results presented above show that F19 and F20, both of which are known to be part of the protected structure (presumably the Aβ core) of the fibril (195), both have solvent inaccessible side chains. In addition, I32 and/or L34 also have solvent inaccessible side chains. This is in agreement with other studies that suggest that I32 and L34 are involved with the β sheet structure and are not part of a free C-terminal tail (200); however, in the model of Perutz *et al*, these side chains would either be exposed to the outside of the continuous outer β sheet or exposed to the water-filled cavity of the nanotube structure. The center of this nanotube model is large enough to support the diffusion of both hydrogen peroxide and hydroxyl radicals (192); therefore, while a decrease in the apparent rate of oxidation of peptide 17-28 could be expected in the Perutz model, the complete lack of oxidation of the side chains should not occur.

A model which would support the burial of both F19 and F20 in a β sheet structure would be one with the strands perpendicular to the axis of the fibril, with the

side chains of the amino acids at the N- and C-termini of the β strand portion of the peptide exposed to solvent (Figure 29). While such a model is by no means the only possible fold that would protect the side chains of both F19 and F20 in a β sheet structure, the model published by Perutz *et al* (192) does not explain the photochemical oxidation surface mapping results presented here without significant alterations. Further experimentation to map other amino acid side chains not probed by this technique is required to establish a model for both the overall structure of the Aβ amyloid fibril as well as the side chain packing within this structure.
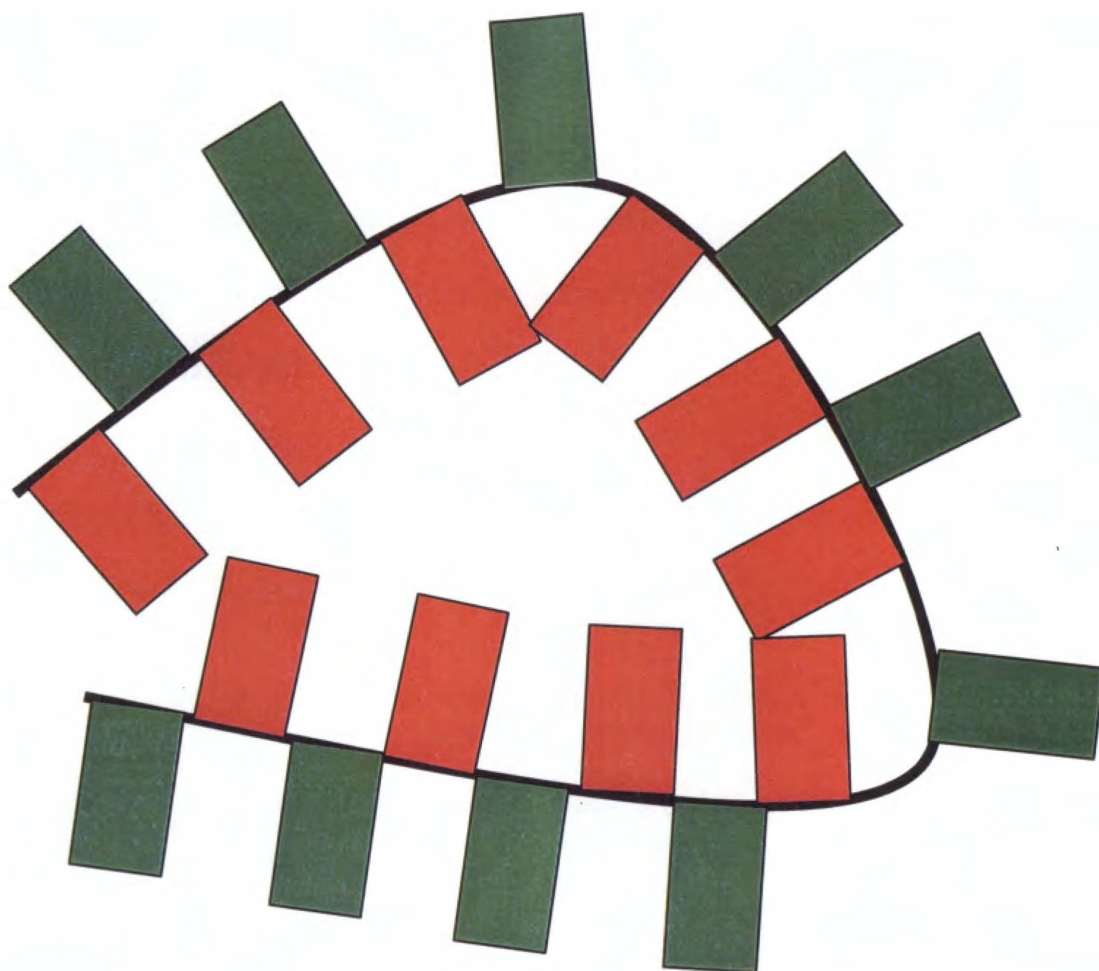
Figure 29: Model of Aβ 1-40 Stacked β Strand Structure

A rough theoretical model of the β strand portion of the Aβ 1-40 fibril. There is an unstructured N-terminus and possibly an unstructured C-terminal tail. The interior of the fibril is protected from solvent by the side chains at the ends of the β strand, with each strand stacking on top of another to form a helical β sheet. The side chains are packed too tightly together to allow penetration of solvent, therefore completely protecting the interior side chains from the probing hydroxyl radical. Protofilaments are packed together in such a way as to shield the outward-facing phenylalanine from the solvent.

115

CHAPTER 7—CONCLUSIONS

In the studies presented here, we have attempted to demonstrate the usefulness of the hydroxyl radical to generate surface constraints for protein modeling. We first set out to prove the suitableness of photochemical hydroxyl surface mapping for examining protein structure. In order to present the hydroxyl radical as a useful probe of protein surfaces, the four criteria set forth by Matheson *et al* should be met (76).

1) The labeling process itself should not alter the surface topography of the protein

2) The presence of the surface labeling compound(s) in solution before labeling should not alter the surface features of the protein

3) It should be demonstrated that residues known to be buried in the native protein are not labeled but can be labeled in the denatured protein

4) In order to correctly interpret the labeling patterns, it should be determined to what extent the labeling reagent is nonspecific

The hydroxyl radical has a complex chemistry with the variety of chemical groups on proteins that makes its applicability as a surface probe somewhat difficult to measure. Fortunately, the rich history of radiation biochemistry from the 1960's and 1970's has left us with an excellent understanding of the protein chemistry of hydroxyl radicals (105, 106). The rich body of literature has allowed the use of the hydroxyl radical as a chemical surface probe; despite the radical's complex protein chemistry, the rates of oxidation and the various products are well characterized, allowing the interpretation of the protein oxidation pattern and satisfying the fourth criterion of Matheson *et al*.

With the model protein lysozyme, it was shown that the labeling process does not alter the structure of the protein as measured by CD spectroscopy. In addition, it was demonstrated that highly chemically reactive residues buried in the structure of lysozyme are not labeled under native conditions, but are labeled under UV-denaturing conditions. These observations satisfy the remaining criteria of Matheson *et al*, leading to validation of the hydroxyl radical as a probe of protein surfaces.

The hydroxyl radical label, like all chemical surface probes, measures the aggregate average solvent accessibility. This is an important factor to keep in mind with all protein surface mapping studies. For instance, if a protein exists in more than one conformation, the apparent rate of oxidation for a specific amino acid will represent the weighted average solvent accessibility for the amino acid side chain over the time of the labeling process. This observation leads to several important implications for the analysis of structural data by this method. First of all, the experiments performed and presented in this work all had timescales of minutes; therefore, the structures measured were all averages of various conformations that the protein undergoes in solution. As we hypothesized based on the data from the one hour oxidation of intact lysozyme (Figure 11, Chapter 4), the presence of multiple structural conformations can result in a non-additive, "cooperative" oxidation profile. The reason behind the presence of two packets of oxidation is that one conformation has more highly reactive groups exposed than the other; therefore, the conformation with more reactive groups exposed will better compete for the hydroxyl radicals that the conformation with the reactive groups buried, and will therefore have higher oxidation states than the less reactive conformation.

The observation of multiple oxidation states for multiple conformations may be useful in examining conformational changes in proteins upon changes in pH, temperature, ionic strength, or ligand binding. By following the extent of oxidation while altering pH, ionic strength, or temperature, the transition point of the conformation change can be determined. In addition, in conformation changes that are very drastic without a gradual transition (such as rapid, drastic pH, ionic strength, or temperature driven changes or conformation changes such as single ligand binding), the multiple conformations can be simultaneously detected, either near the transition point for pH, ionic strength, or temperature driven conformations or at substoichiometric concentrations of ligand for ligand binding, by rapid oxidation under these conditions. Not only can the presence of the different conformations be determined, but the conformations can be generally described by examining the direction of the shift in the oxidation state. If upon ligand binding, a protein becomes more heavily oxidized under the same conditions, the protein can be said to become generally more "open", or unfolded. However, if the protein becomes less oxidized, the protein can be said to become generally more "closed", or tightly folded. Separate experiments can also be run to more carefully characterize the nature of the conformation change. Studies in this area are currently being pursued, examining the oxidation of β-lactoglobulin A at different pH.

In addition to the applicability in measuring conformation changes, photochemical hydroxyl radical surface mapping has great potential in mapping membrane proteins. Great interest has been presented for membrane proteins (208-213), due to their importance in signal transduction, as drug targets, as transporters, and their implication in disease due to misfolding. The hydroxyl radical has the excellent

118

characteristic of closely mimicking water in size and hydrophilicity; unlike many chemical labels, the radical can realistically claim to model solvent accessibility closely. This is highly important for determining the structure of small molecule transporters, especially aquaporins, which allow the transport of water (214-218). In addition to aquaporins, hydroxyl radical mapping could be an important tool for determining the locations of loops and tails in membrane proteins, therefore assisting in localizing the location of transmembrane domains. In addition to topology, radical mapping of intact cells may assist in determining the orientation of membrane proteins, both important questions in biology (211, 219-228).

Another area of interest for future research is that of mapping sites of protein-protein interaction interfaces. Initial work in this area was presented in Chapter 6, measuring the interaction between the Aβ 1-40 fibrils; however, additional work in mapping protein interaction sites of globular proteins is currently underway. The potential for developing models for mutagenesis studies by determining sites that become buried upon addition of an interacting partner is an exciting prospect for examining the topology of protein complexes. One difficulty lies in the fact that the protocol developed measured protein solvent accessibility over a period of minutes. In this time, protein complexes are often formed, then fall apart, and are reformed many times. Under most circumstances, this is not a problem; the technique measures the aggregate average solvent accessibility, and for stable complexes, the majority of the protein will be in the complexed form. However, prior work with R67 dihydrofolate reductase brought to light a specific problem with this technique when applied to certain complexes. The R67 dihydrofolate reductase complex is a homotetramer which exists in equilibrium with the

dimeric form in solution, with the tetramer much more prevalent. However, one essential residue in the dimer-dimer interface is a histidine. Upon hydroxyl radical surface mapping of the complex, it was found that the oxidation tended to degrade a portion of the sample into an irreversible dimeric form. The reason for this is, in the small portion of the protein that occurs in the dimeric form, the radical oxidized the solvent-exposed histidine. Upon oxidation, histidine undergoes a complex series of oxidation pathways, resulting in numerous products, most of which involve scission of the imidazole ring (Figure 4). For the small subpopulation of protein that exists in the dimeric form, the histidine is oxidized. Once this essential histidine is oxidized, the oxidized dimer can no longer form the tetramer. The inability of this dimer to form the tetramer removes it from the dimer-tetramer equilibrium, irreversibly populating the dimer state the longer the oxidation occurs.

In order to overcome this problem for examining sensitive protein-protein interaction interfaces, a flash labeling method needs to be developed. If the labeling occurs in less time than the half-life of the complex, then the vast majority of the oxidative damage to the dimer-dimer interface can be avoided. In order to realize this goal, a more stable labeling precursor needs to be developed. As shown in Figure 11, hydrogen peroxide at ambient conditions can lead to oxidation, albeit at a low rate. A flash labeling method would require immediate activation and quenching of the labeling reaction; therefore, the unstable hydrogen peroxide is unsuitable. Work performed in high intensity X-ray radiolysis of water to generate hydroxyl radicals can label on the order of microseconds (99-101, 104, 164-167, 184); however, the availability of high flux X-ray synchrotron sources greatly limit the availability of this technique to the research

120

community. High-flux gamma ray or fast neutron sources would also work for generating hydroxyl radicals (229); however, sources that generate enough flux to sufficiently probe protein surfaces on the order of microseconds or less would again place the technique beyond general utility to the research community. The only form of radiation that is readily available at high flux is UV and visible light. A probe that is more stable in solution while still readily reactive after exposure to an intense flash of light would be an excellent probe for general flash labeling; considerable effort should be directed towards identifying alternative reagents for flash labeling, perhaps using an alternative mechanism other than the hydroxyl radical reaction.

An area where this research may have considerable impact is the field of structural genomics. As mentioned previously, the advances in genomics, transcriptomics, and proteomics have far outstripped the modest advances in structural genomics (Figure 1). When the analytical techniques applied to each problem are examined, the explanation for this lag in structural genomics is obvious. Both genomics and transcriptomics are successful at high-throughput analysis because they are easily multiplexed and highly automated, and the purification of the sample is straightforward. Proteomics, while not as capable of multiplexing, is still highly automated and utilizes the superb resolution, accuracy, sensitivity, and dynamic range of mass spectrometry to examine complex mixtures in a single run without purification of individual proteins.

Structural genomics relies upon XRC and NMR spectroscopy. Neither technique can be readily automated due to difficulties in sample purification (each protein must be highly purified), sample preparation (proteins must form perfect crystals for XRC), and complex data analysis. Neither technique is amenable to all (or perhaps even most)

121

proteins, with a published success rate of ~20% from target to structure (31). With the enormous amount of effort required for each structure, the low success rate, and the unsuitability of the techniques for many proteins, it is no surprise that the "high-throughput" initiatives based on these techniques boast tens of structures per year (28) as opposed to the thousands of genes, transcriptional levels, and protein ID's determined on a daily or weekly basis by other, truly high-throughput techniques. In order to bring structural genomics into the realm of high-throughput biology, a different analytical technique must be applied. The application of the peroxide-mediated hydroxyl radical surface mapping technique is not suitable in and of itself for mapping entire proteomes due to the presence of transition metal-binding proteins. In the presence of transition metals, uncontrolled Fenton chemistry will occur, leading to oxidative degradation of metal-binding proteins. Therefore, in order to be useful for whole proteome measurements, either the transition metals must be removed from solution prior to measurement (meaning that the apo- structure of metalloproteins will be probed), or a more stable labeling reagent must be developed.

In addition to alternative labeling methods, another essential piece that must be in place before this technique can be applied in a high-throughput fashion is an automated data analysis pipeline. In the studies detailed here, TurboSEQUEST was utilized to screen tremendous amounts of LC-MS/MS data for the presence of oxidized peptides. TurboSEQUEST does a fairly good job identifying the presence and identity of oxidized peptides; however, the *in silico* fragmentation algorithm is very crude and inaccurate, and the algorithm will often identify chemical noise as actual peaks. In addition, TurboSEQUEST only allows one potential modification per search; therefore, one cannot

122

search for both the sulfoxide and the sulfone version of methionine oxidation in one search. The end result is that, while the algorithm can identify which peptides were oxidized by LC-MS/MS, TurboSEQUEST is very bad at determining the actual sites of oxidation from fragmentation spectra. Currently, all MS/MS spectra must be interpreted manually. A more detailed program for determining sites of modification based on MS/MS fragmentation spectra must be developed in order to improve the speed of analysis.

Another area that must be improved for the high-throughput application of surface constraints is the computational modeling algorithms. As a leading pseudo-*ab initio* modeling algorithm that relies upon full-atom representations for accuracy, Rosetta is an excellent algorithm for use with surface constraints. However, the current method utilized for evaluating computational models is inefficient. Currently, we generate the models and then screen them manually for agreement with our data. A much more efficient and more productive approach would be to utilize an energy function within the Rosetta algorithm itself to penalize structures that violate the constraints provided. This would greatly increase the likelihood that the proper structure is scored at or near the top of the list of models, while further automating the data analysis process. However, the source code for Rosetta is currently unavailable. Efforts to either obtain the source code or to foster a collaboration with the creator of Rosetta (186) should be made.

The principles learned from the experiments detailed here are invaluable lessons in what must be achieved for high-throughput structural genomics initiatives. While the applicability for photochemical hydroxyl radical surface mapping for targeted applications in proteins, protein-protein complexes, and membrane protein topology, as

123

well as more esoteric subjects such as viral particle surface mapping (230) is apparent and important, the implications for high-throughput analysis of protein structure is a more exciting yet difficult prospect. The value of multi-target surface labeling is apparent as a method for evaluating structural modeling. Further work must be performed in order to optimize both the experimental conditions and the data analysis; however, the foundation presented here shows that the goal of rapid, broadly applicable structural characterization without the need for extensive protein purification is realizable. The current technology as presented here enables the rapid probing of structures of proteins that were resistant to other, more direct analyses, granting important insight into their structure. As more broadly applicable labeling reagents are developed and more powerful data analysis methods developed, the role of protein surface mapping backed by the analytical power of mass spectrometry should play an increasingly important role in all facets of protein structure-function analyses.

Works Cited

1. Fleischmann RD, Adams MD, White O, et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269:** 496-512.

2. Fraser CM, Gocayne JD, White O, et al. (1995) The minimal gene complement of Mycoplasma genitalium. *Science* **270:** 397-403.

3. Dujon B. (1996) The yeast genome project: what did we learn? *Trends Genet* **12:** 263-270.

4. Bult CJ, White O, Olsen GJ, et al. (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* **273:** 1058-1073.

5. Ureta-Vidal A, Ettwiller L, Birney E. (2003) Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat Rev Genet* **4:** 251-262.

6. Edwards RA, Olsen GJ, Maloy SR. (2002) Comparative genomics of closely related salmonellae. *Trends Microbiol* **10:** 94-99.

7. Castresana J. (2001) Comparative genomics and bioenergetics. *Biochim Biophys Acta* **1506:** 147-162.

8. Payne DJ, Holmes DJ, Rosenberg M. (2001) Delivering novel targets and antibiotics from genomics. *Curr Opin Investig Drugs* **2:** 1028-1034.

9. McDevitt D, Payne DJ, Holmes DJ, Rosenberg M. (2002) Novel targets for the future development of antibacterial agents. *Symp Ser Soc Appl Microbiol*: 28S-34S.

10. Wernegreen JJ. (2002) Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet* **3:** 850-861.

11. Paulsen IT, Chen J, Nelson KE, Saier MH, Jr. (2001) Comparative genomics of microbial drug efflux systems. *J Mol Microbiol Biotechnol* **3:** 145-150.

12. Mushegian AR, Koonin EV. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci U S A* **93**: 10268-10273.

13. Koonin EV. (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu Rev Genomics Hum Genet* **1**: 99-116.

14. DeRisi JL, Iyer VR, Brown PO. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680-686.

15. Behr MA, Wilson MA, Gill WP, et al. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**: 1520-1523.

16. Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S. (2000) A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. *Proc Natl Acad Sci U S A* **97**: 14668-14673.

17. Kato-Maeda M, Rhee JT, Gingeras TR, et al. (2001) Comparing genomes within the species Mycobacterium tuberculosis. *Genome Res* **11**: 547-554.

18. Hakenbeck R, Balmelle N, Weber B, Gardes C, Keck W, de Saizieu A. (2001) Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of Streptococcus pneumoniae. *Infect Immun* **69**: 2477-2486.

19. Selinger DW, Cheung KJ, Mei R, et al. (2000) RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nat Biotechnol* **18**: 1262-1268.

20. Ye RW, Tao W, Bedzyk L, Young T, Chen M, Li L. (2000) Global gene expression profiles of Bacillus subtilis grown under anaerobic conditions. *J Bacteriol* **182**: 4458-4465.

21. James TC, Campbell S, Donnelly D, Bond U. (2003) Transcription profile of brewery yeast under fermentation conditions. *J Appl Microbiol* **94:** 432-448.

22. Smith RD. (2002) Trends in mass spectrometry instrumentation for proteomics. *Trends Biotechnol* **20:** S3-7.

23. Wolters DA, Washburn MP, Yates JR, 3rd. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73:** 5683-5690.

24. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17:** 994-999.

25. Smolka MB, Zhou H, Purkayastha S, Aebersold R. (2001) Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal Biochem* **297:** 25-31.

26. Harris RD, Nindl G, Balcavage WX, Weiner W, Johnson MT. (2003) Use of proteomics methodology to evaluate inflammatory protein expression in tendinitis. *Biomed Sci Instrum* **39:** 493-499.

27. Ong SE, Blagoev B, Kratchmarova I, et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1:** 376-386.

28. Chance MR, Bresnick AR, Burley SK, et al. (2002) Structural genomics: a pipeline for providing structures for the biologist. *Protein Sci* **11:** 723-738.

29. Tanford C, Bunville, L. G., Nozaki, Y. (1959) Reversible transformation of b-lactoglobulin at pH 7.5. *J Am Chem Soc* **81:** 4032-4036.

30. Qin BY, Bewley MC, Creamer LK, Baker HM, Baker EN, Jameson GB. (1998) Structural basis of the tanford transition of bovine beta-lactoglobulin. *Biochemistry* **37:** 14014-14023.

31. Vitkup D, Melamud E, Moult J, Sander C. (2001) Completeness in structural genomics. *Nat Struct Biol* **8:** 559-566.

32. Staunton D, Owen J, Campbell ID. (2003) NMR and structural genomics. *Acc Chem Res* **36:** 207-214.

33. Christendat D, Yee A, Dharamsi A, et al. (2000) Structural proteomics of an archaeon. *Nat Struct Biol* **7:** 903-909.

34. Yee A, Chang X, Pineda-Lucena A, et al. (2002) An NMR approach to structural proteomics. *Proc Natl Acad Sci U S A* **99:** 1825-1830.

35. Forster MJ. (2002) Molecular modelling in structural biology. *Micron* **33:** 365-384.

36. Fiser A, Feig M, Brooks CL, 3rd, Sali A. (2002) Evolution and physics in comparative protein structure modeling. *Acc Chem Res* **35:** 413-421.

37. Sanchez R, Sali A. (1997) Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins* **Suppl 1:** 50-58.

38. Martin AC, MacArthur MW, Thornton JM. (1997) Assessment of comparative modeling in CASP2. *Proteins* **Suppl 1:** 14-28.

39. Lesk AM, Lo Conte L, Hubbard TJ. (2001) Assessment of novel fold targets in CASP4: predictions of three-dimensional structures, secondary structures, and interresidue contacts. *Proteins* **Suppl 5:** 98-118.

40. Bowers PM, Strauss CE, Baker D. (2000) De novo protein structure determination using sparse NMR data. *J Biomol NMR* **18:** 311-318.

41. Xu Y, Xu, D., Crawford, O. H., Einstein, J. R. (2000) A computational method for NMR-constrained protein threading. *Journal of Computational Biology* **7:** 449-467.

42. Xu Y, Xu D. (2000) Protein threading using PROSPECT: design and evaluation. *Proteins* **40:** 343-354.

43. Schonbrun J, Wedemeyer WJ, Baker D. (2002) Protein structure prediction in 2002. *Curr Opin Struct Biol* **12:** 348-354.

44. Benkirane MM, Bon, D., Costagliola, S., Paolucci, F., Darbouret, B., Prince, P., Carayon, P. (1987) Monoclonal Antibody Mapping of the Antigenic Surface of Human Thyrotropin and Its Subunits. *Endocrinology* **121:** 1171-1177.

45. Meyer WJ, Johnston, R. E. (1993) Structural Rearrangement of Infecting Sindbis Virions at the Cell Surface: Mapping of Newly Accessible Epitopes. *Journal of Virology* **67:** 5117-5125.

46. Dong L, Hulsmeyer M, Durkop H, et al. (2003) Human CD30: Structural implications from epitope mapping and modeling studies. *J Mol Recognit* **16:** 28-36.

47. Leclerc E, Peretz D, Ball H, et al. (2003) Conformation of PrP(C) on the cell surface as probed by antibodies. *J Mol Biol* **326:** 475-483.

48. Thole HH, Maschler, I., Jungblut, P. W. (1995) Surface Mapping of the ligand-filled C-terminal half of the porcine estradiol receptor by restricted proteolysis. *European Journal of Biochemistry* **231:** 510-516.

130

49.   Nemoto T, Sato, N., Iwanari, H., Yamashita, H., Takagi, T. (1997) Domain structures and immunogenic regions of the 90-kDa heat-shock protein (HSP90). Probing with a library of anti-HSP90 monoclonal antibodies and limited proteolysis. *J Biol Chem* **272:** 26179-26187.

50.   Shields SJ, Oyeyemi O, Lightstone FC, Balhorn R. (2003) Mass spectrometry and non-covalent protein-ligand complexes: confirmation of binding sites and changes in tertiary structure. *J Am Soc Mass Spectrom* **14:** 460-470.

51.   D'Ambrosio C, Talamo F, Vitale RM, et al. (2003) Probing the dimeric structure of porcine aminoacylase 1 by mass spectrometric and modeling procedures. *Biochemistry* **42:** 4430-4443.

52.   Vinci F, Couprie J, Pucci P, Quemeneur E, Moutiez M. (2002) Description of the topographical changes associated to the different stages of the DsbA catalytic cycle. *Protein Sci* **11:** 1600-1612.

53.   Monti M, Principe S, Giorgetti S, et al. (2002) Topological investigation of amyloid fibrils obtained from beta2-microglobulin. *Protein Sci* **11:** 2362-2369.

54.   Moradian-Oldak J, Jimenez I, Maltby D, Fincham AG. (2001) Controlled proteolysis of amelogenins reveals exposure of both carboxy- and amino-terminal regions. *Biopolymers* **58:** 606-616.

55.   Niu W, Wu Y, Sui SF. (2000) Orientation of membrane-bound melittin studied by a combination of HPLC and liquid secondary ion mass spectrometry (LSIMS). *IUBMB Life* **50:** 215-219.

56. Phinney BS, Blackburn K, Brown DT. (2000) The surface conformation of Sindbis virus glycoproteins E1 and E2 at neutral and low pH, as determined by mass spectrometry-based mapping. *J Virol* **74**: 5667-5678.

57. O'Brien JA, Taylor JA, Bellamy AR. (2000) Probing the structure of rotavirus NSP4: a short sequence at the extreme C terminus mediates binding to the inner capsid particle. *J Virol* **74**: 5388-5394.

58. Orru S, Dal Piaz F, Casbarra A, et al. (1999) Conformational changes in the NS3 protease from hepatitis C virus strain Bk monitored by limited proteolysis and mass spectrometry. *Protein Sci* **8**: 1445-1454.

59. Scaloni A, Monti M, Acquaviva R, et al. (1999) Topology of the thyroid transcription factor 1 homeodomain-DNA complex. *Biochemistry* **38**: 64-72.

60. Burns JM, Gallo RC, DeVico AL, Lewis GK. (1998) A new monoclonal antibody, mAb 4A12, identifies a role for the glycosaminoglycan (GAG) binding domain of RANTES in the antiviral effect against HIV-1 and intracellular Ca2+ signaling. *J Exp Med* **188**: 1917-1927.

61. Gervasoni P, Staudenmann W, James P, Pluckthun A. (1998) Identification of the binding surface on beta-lactamase for GroEL by limited proteolysis and MALDI-mass spectrometry. *Biochemistry* **37**: 11660-11669.

62. Bothner B, Dong XF, Bibbs L, Johnson JE, Siuzdak G. (1998) Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J Biol Chem* **273**: 673-676.

63. Zappacosta F, Pessi A, Bianchi E, et al. (1996) Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: the case of Minibody. *Protein Sci* **5**: 802-813.

64. Hager-Braun C, Tomer KB. (2002) Characterization of the tertiary structure of soluble CD4 bound to glycosylated full-length HIVgp120 by chemical modification of arginine residues and mass spectrometric analysis. *Biochemistry* **41**: 1759-1766.

65. Kvaratskhelia M, Miller JT, Budihas SR, Pannell LK, Le Grice SF. (2002) Identification of specific HIV-1 reverse transcriptase contacts to the viral RNA:tRNA complex by mass spectrometry and a primary amine selective reagent. *Proc Natl Acad Sci U S A* **99**: 15988-15993.

66. Britto PJ, Knipling L, Wolff J. (2002) The local electrostatic environment determines cysteine reactivity of tubulin. *J Biol Chem* **277**: 29018-29027.

67. Leite JF, Cascio M. (2002) Probing the topology of the glycine receptor by chemical modification coupled to mass spectrometry. *Biochemistry* **41**: 6140-6148.

68. Apuy JL, Park ZY, Swartz PD, Dangott LJ, Russell DH, Baldwin TO. (2001) Pulsed-alkylation mass spectrometry for the study of protein folding and dynamics: development and application to the study of a folding/unfolding intermediate of bacterial luciferase. *Biochemistry* **40**: 15153-15163.

69. Linder MD, Morkunaite-Haimi S, Kinnunen PK, Bernardi P, Eriksson O. (2002) Ligand-selective modulation of the permeability transition pore by arginine

modification. Opposing effects of p-hydroxyphenylglyoxal and phenylglyoxal. *J Biol Chem* **277**: 937-942.

70. Gonzalez de Peredo A, Saint-Pierre C, Latour JM, Michaud-Soret I, Forest E. (2001) Conformational changes of the ferric uptake regulation protein upon metal activation and DNA binding; first evidence of structural homologies with the diphtheria toxin repressor. *J Mol Biol* **310**: 83-91.

71. Watson JT, Yang Y, Wu J. (2001) Capture and identification of folding intermediates of cystinyl proteins by cyanylation and mass spectrometry. *J Mol Graph Model* **19**: 119-128.

72. Ashfield JT, Meyers T, Lowne D, et al. (2000) Chemical modification of a variant of human MIP-1alpha; implications for dimer structure. *Protein Sci* **9**: 2047-2053.

73. Kalume DE, Stenflo J, Czerwiec E, et al. (2000) Structure determination of two conotoxins from Conus textile by a combination of matrix-assisted laser desorption/ionization time-of-flight and electrospray ionization mass spectrometry and biochemical methods. *J Mass Spectrom* **35**: 145-156.

74. Calvete JJ, Campanero-Rhodes MA, Raida M, Sanz L. (1999) Characterisation of the conformational and quaternary structure-dependent heparin-binding region of bovine seminal plasma protein PDC-109. *FEBS Lett* **444**: 260-264.

75. Staros JV, Richards, F. M. (1974) Photochemical Labeling of the Surface Proteins of Human Erythrocytes. *Biochemistry* **13**: 2720-2726.

76. Matheson RR, Jr., Van Wart HE, Burgess AW, Weinstein LI, Scheraga HA. (1977) Study of protein topography with flash photolytically generated

nonspecific surface-labeling reagents: surface labeling of ribonuclease A. *Biochemistry* **16:** 396-403.

77. Mas MT, Wang, J. K., Hargrave, P. A. (1980) Topography of Rhodopsin in Rod Outer Segment Disk Membranes. Photochemical Labeling with *N*-(4-Azido-2-nitrophenyl)-2-aminoethanesulfonate. *Biochemistry* **19:** 684-692.

78. DiPaola M, Maxfield, F. R. (1984) Conformational Changes in the Receptors for Epidermal Growth Factor and Asialoglycoproteins Induced by the Mildly Acidic pH Found in Endocytic Vesicles. *The Journal of Biological Chemistry* **259:** 9163-9171.

79. Yamada N, Suzuki E, Hirayama K. (2002) Identification of the interface of a large protein-protein complex using H/D exchange and Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* **16:** 293-299.

80. Smith DL, Deng Y, Zhang Z. (1997) Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J Mass Spectrom* **32:** 135-146.

81. Chen J, Smith DL. (2001) Amide hydrogen exchange shows that malate dehydrogenase is a folded monomer at pH 5. *Protein Sci* **10:** 1079-1083.

82. Engen JR. (2003) Analysis of protein complexes with hydrogen exchange and mass spectrometry. *Analyst* **128:** 623-628.

83. Simmons DA, Dunn SD, Konermann L. (2003) Conformational dynamics of partially denatured myoglobin studied by time-resolved electrospray mass spectrometry with online hydrogen-deuterium exchange. *Biochemistry* **42:** 9248.

84.   Nazabal A, Dos Reis S, Bonneu M, Saupe SJ, Schmitter JM. (2003)
       Conformational Transition Occurring upon Amyloid Aggregation of the HET-s
       Prion Protein of Podospora anserina Analyzed by Hydrogen/Deuterium Exchange
       and Mass Spectrometry. *Biochemistry* **42:** 8852-8861.

85.   Englander JJ, Del Mar C, Li W, et al. (2003) Protein structure change studied by
       hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc
       Natl Acad Sci U S A* **100:** 7057-7062.

86.   Nazabal A, Laguerre M, Schmitter JM, Vaillier J, Chaignepain S, Velours J.
       (2003) Hydrogen/deuterium exchange on yeast ATPase supramolecular protein
       complex analyzed at high sensitivity by MALDI mass spectrometry. *J Am Soc
       Mass Spectrom* **14:** 471-481.

87.   Zhu MM, Rempel DL, Du Z, Gross ML. (2003) Quantification of protein-ligand
       interactions by mass spectrometry, titration, and H/D exchange: PLIMSTEX. *J
       Am Chem Soc* **125:** 5252-5253.

88.   Powell KD, Fitzgerald MC. (2003) Accuracy and precision of a new H/D
       exchange- and mass spectrometry-based technique for measuring the
       thermodynamic properties of protein-peptide complexes. *Biochemistry* **42:** 4962-
       4970.

89.   Kipping M, Schierhorn A. (2003) Improving hydrogen/deuterium exchange mass
       spectrometry by reduction of the back-exchange effect. *J Mass Spectrom* **38:** 271-
       276.

90.   Alomirah H, Alli I, Konishi Y. (2003) Charge state distribution and
       hydrogen/deuterium exchange of alpha-lactalbumin and beta-lactoglobulin

preparations by electrospray ionization mass spectrometry. *J Agric Food Chem* **51:** 2049-2057.

91.   Lam TT, Lanman JK, Emmett MR, Hendrickson CL, Marshall AG, Prevelige PE. (2002) Mapping of protein:protein contact surfaces by hydrogen/deuterium exchange, followed by on-line high-performance liquid chromatography-electrospray ionization Fourier-transform ion-cyclotron-resonance mass analysis. *J Chromatogr A* **982:** 85-95.

92.   Wang F, Tang X. (1996) Conformational heterogeneity of stability of apomyoglobin studied by hydrogen/deuterium exchange and electrospray ionization mass spectrometry. *Biochemistry* **35:** 4069-4078.

93.   Heyduk E, Heyduk T. (1994) Mapping protein domains involved in macromolecular interactions: a novel protein footprinting approach. *Biochemistry* **33:** 9643-9650.

94.   Greiner DP, Hughes, K. A., Gunasekera, A. H., Meares, C. F. (1996) Binding of the σ^{70} protein to the core subunits of *Escherichia coli* RNA polymerase, studied by iron-EDTA protein footprinting. *Proceedings of the National Academy of Science, USA* **93:** 71-75.

95.   Baichoo N, Heyduk, T. (1997) Mapping conformational changes in a protein: application of a protein footprinting technique to a cAMP-induced conformational changes in cAMP receptor protein. *Biochemistry* **36:** 10830-10836.

96.   Cheng X, Shaltiel, S., Taylor, S.S. (1998) Mapping substrate-induced conformational changes in cAMP-dependent protein kinase by protein footprinting. *Biochemistry* **37:** 14005-14013.

137

97.   Wei RR, Richardson, J.P. (2001) Identification of an RNA-binding site in the ATP binding domain of Escherichia coli Rho by H_2O_2/Fe-EDTA cleavage protection studies. *J Biol Chem* **276:** 28380-28387.

98.   Heyduk T, Baichoo N, Heyduk E. (2001) Hydroxyl radical footprinting of proteins using metal ion complexes. *Met Ions Biol Syst* **38:** 255-287.

99.   Maleknia SD, Ralston CY, Brenowitz MD, Downard KM, Chance MR. (2001) Determination of macromolecular folding and structure by synchrotron x-ray radiolysis techniques. *Anal Biochem* **289:** 103-115.

100.   Maleknia SD, Brenowitz M, Chance MR. (1999) Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry. *Anal Chem* **71:** 3965-3973.

101.   Kiselar JG, Maleknia SD, Sullivan M, Downard KM, Chance MR. (2002) Hydroxyl radical probe of protein surfaces using synchrotron X-ray radiolysis and mass spectrometry. *Int J Radiat Biol* **78:** 101-114.

102.   Maleknia SD, Kiselar JG, Downard KM. (2002) Hydroxyl radical probe of the surface of lysozyme by synchrotron radiolysis and mass spectrometry. *Rapid Commun Mass Spectrom* **16:** 53-61.

103.   Maleknia SD, Downard KM. (2001) Unfolding of apomyoglobin helices by synchrotron radiolysis and mass spectrometry. *Eur J Biochem* **268:** 5578-5588.

104.   Maleknia SD, Chance MR, Downard KM. (1999) Electrospray-assisted modification of proteins: a radical probe of protein structure. *Rapid Commun Mass Spectrom* **13:** 2352-2358.

105.	Garrison WM. (1987) Reaction-Mechanisms in the Radiolysis of Peptides, Polypeptides, and Proteins. *Chem Rev* **87**: 381-398.

106.	Buxton GV, Greenstock CL, Helman WP, Ross AB. (1988) Critical-Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen-Atoms and Hydroxyl Radicals (.Oh/.O-) in Aqueous-Solution. *J Phys Chem Ref Data* **17**: 513-886.

107.	Tcherkasskaya O, Davidson EA, Uversky VN. (2003) Biophysical constraints for protein structure prediction. *J Proteome Res* **2**: 37-42.

108.	Xu Y, Xu D, Crawford OH, et al. (1999) Protein threading by PROSPECT: a prediction experiment in CASP3. *Protein Eng* **12**: 899-907.

109.	Watts A, Burnett IJ, Glaubitz C, et al. (1999) Membrane protein structure determination by solid state NMR. *Nat Prod Rep* **16**: 419-423.

110.	Leeuwenburgh C, Hansen P, Shaish A, Holloszy JO, Heinecke JW. (1998) Markers of protein oxidation by hydroxyl radical and reactive nitrogen species in tissues of aging rats. *Am J Physiol* **274**: R453-461.

111.	Leeuwenburgh C, Wagner P, Holloszy JO, Sohal RS, Heinecke JW. (1997) Caloric restriction attenuates dityrosine cross-linking of cardiac and skeletal muscle proteins in aging mice. *Arch Biochem Biophys* **346**: 74-80.

112.	Garrison WM, Jayko, M. E., Bennett, W. (1962) Radiation-induced oxidation of protein in aqueous solution. *Radiat. Res.* **16**: 483-502.

113.	Garrison WM, Kland-English, M., Sokol, H. A., Jayko, M. E. (1970) Radiolytic degradation of the peptide main chain in dilute aqueous solution containing oxygen. *J. Phys. Chem.* **74**: 4506-4509.

114. Hawkins CL, Davies MJ. (2001) Generation and propagation of radical reactions on proteins. *Biochim Biophys Acta* **1504:** 196-219.

115. Joshi A, Rustgi, S., Moss, H., Riesz, P. (1978) E.S.R. of spin-trapped radicals in aqueous solutions of peptides. Reactions of the hydroxyl radical. *Int J Radiat Biol* **33:** 205-229.

116. Liebster J, Kopoldova, J. (1966) Radiation chemical reactions in aqueous oxygenated and oxygen-free solutions of aliphatic dipeptides and tripeptides. *J Radiat Res* **27:** 162-173.

117. Sokol HA, Bennett-Corniea, W., Garrison, W. M. (1965) A marked effect of conformation in the radiolysis of poly-a-L-glutamic acid in aqueous solution. *J Am Chem Soc* **87:** 1391-1392.

118. Bennett-Corniea W, Sokol, H. A., Garrison, W. M. (1970) Reductive deamination in the radiolysis of oligopeptides in aqueous solution and in the solid state. *Radiat. Res.* **43:** 257-265.

119. Dizdaroglu M, Gajewski E, Simic MG. (1982) Mechanisms of Radiation-Induced Racemization and Dimerization of Aliphatic Peptides. *Radiat Res* **91:** 296-296.

120. Dizdaroglu M, Simic MG. (1983) Isolation and Characterization of Radiation-Induced Aliphatic Peptide Dimers. *International Journal of Radiation Biology* **44:** 231-239.

121. Garrison WM, Sokol, H. A., Bennett-Corniea, W. (1973) Radiation chemistry of glycylglycine in oxygen-free systems. *Radiat. Res.* **53:** 376-384.

122. Brodskaya GA, Sharpatyi, V. A. (1967) Radiolysis of aqueous tyrosine solutions. II. Dependence on the medium pH. *Russ J Phys Chem* **41:** 2850-2854.

123. Dizdaroglu M, Simic, M. G. (1981). In: Rodgers MAJ, Powers, E. L. (ed.) *Oxygen and Oxy-Radicals in Chemistry and Biology.* Academic, New York.

124. Balakrishnan I, Reddy, M. P. (1970) Mechanism of reaction of hydroxyl radicals with benzene in the g radiolysis of the aerated aqueous benzene system. *J Phys Chem* **74**: 850-855.

125. Winchester RV, Lynn, K. R. (1970) X- and gamma-radiolysis of some tryptophan dipeptides. *Int J Radiat Biol* **17**: 541-548.

126. Armstrong RC, Swallow, A. J. (1969) Pulse- and gamma-radiolysis of aqueous solutions of tryptophan. *J Radiat Res* **40**: 563-579.

127. Kopoldova J, Hrncir, S. (1977) Gamma-radiolysis of aqueous solution of histidine. *Z. Naturforsch. C.: Biosci.* **32C**: 482-487.

128. Samuni A, Neta, P. (1973) Hydroxy radical reaction with phosphate esters and the mechanism of phosphate cleavage. *J Phys Chem* **77**: 2425-2429.

129. Rao PS, Simic, M., Hayon, E. (1975) Pulse radiolysis study of imidazole and histidine in water. *J Phys Chem* **79**: 1260-1263.

130. Mittal JP, Hayon, E. (1974) Interaction of hydrated electrons with phenylalanine and related compounds. *J Phys Chem* **78**: 1790-1794.

131. Kopoldova J, Liebster, J., Gross, E. (1967) Radiation chemical reactions in aqueous solutions of methionine and its peptides. *Radiat. Res.* **30**: 261-274.

132. Gajewski E, Dizdaroglu, M., Krutzsch, H. C., Simic, M. G. (1984) Hydroxyl radical-induced crosslinks of methionine peptides. *Int J Radiat Biol* **46**: 47-55.

133. Sharp JS, Becker, J. M., and Hettich, R. L. (in review) Rapid Analysis of Protein Solvent Accessible Surfaces by Photochemical Oxidation and Mass Spectrometry.

134. Sharp JS, Becker JM, Hettich RL. (2003) Protein surface mapping by chemical oxidation: Structural analysis by mass spectrometry. *Anal Biochem* **313:** 216-225.

135. Barton JPP, J. E. (1970) Radiolysis of oxygenated cysteine solutions at neutral pH. The role of RSSR and O2-. *int J Radiat Phys Chem* **2:** 159-166.

136. Al-Thannon AA, Barton, J. P., Packer, J. E., Sims, R. J., Trumbore, C. N., Winchester, R. V. (1974) Radiolysis of aqueous solutions of cysteine in the presence of oxygen. *int J Radiat Phys Chem* **6:** 233-247.

137. Bonifacic M, Schaefer, K., Moeckel, H., Asmus, K. D. (1975) Primary steps in the reactions of organic disulfides with hydroxyl radicals in aqueous solution. *J Phys Chem* **79:** 1496-1502.

138. Bonifacic M, Moeckel, H., Bahnemann, D., Asmus, K. D. (1975) Formation of positive ions and other primary species in the oxidation of sulfides by hydroxyl radicals. *J Chem Soc Perkin Trans 2:* 675-685.

139. Bonifacic M, Asmus, K. D. (1976) Free radical oxidation of organic disulfides. *J Phys Chem* **80:** 2426-2430.

140. Senko MW, Hendrickson, C. L., Emmett, M. R., Shi, S. D.-H., Marshall, A. G. (1997) External Accumulation of Ions for Enhanced Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **8:** 970-976.

141. McIver RT, Jr., Hunter, R. L., Bowers, W. D. (1985) Coupling a quadrupole mass spectrometer and a Fourier transform mass spectrometer. *Int J Mass Spectrom Ion Processes* **64:** 67-77.

142. Horn DMZ, R.A. and McLafferty, F.W. (2000) Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci.* **97:** 10313-10317.

143. Gauthier JW, Trautman, T.R., Jacobson, D.B. (1991) Sustained off-resonance irradiation for CAD involving FTMS. CAD technique that emulates infrared multiphoton dissociation. *Anal Chim Acta* **246:** 211-225.

144. Ducret A, Van Oostveen I, Eng JK, Yates JR, 3rd, Aebersold R. (1998) High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci* **7:** 706-719.

145. Roepstorff P, Fohlman J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **11:** 601.

146. Schwalbe H, Grimshaw SB, Spencer A, et al. (2001) A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci* **10:** 677-688.

147. Kuwata K, Hoshino M, Forge V, Era S, Batt CA, Goto Y. (1999) Solution structure and dynamics of bovine beta-lactoglobulin A. *Protein Sci* **8:** 2541-2545.

148. Evans SV, Brayer GD. (1990) High-resolution study of the three-dimensional structure of horse heart metmyoglobin. *J Mol Biol* **213:** 885-897.

149. Fraczkiewicz R, Braun W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem* **19:** 319-333.

150. Gupta V, Peterson, C.B., Dice, L., Uchiki, T., Guo, J., Xu, Y., Hettich, R., and Dealwis, C. (In review) Sml1p Is a Dimer in Solution: Characterization of Denaturation and Renaturation of Recombinant Sml1p.

151. Uchiki T, Hettich R, Gupta V, Dealwis C. (2002) Characterization of monomeric and dimeric forms of recombinant Sml1p-histag protein by electrospray mass spectrometry. *Anal Biochem* **301**: 35-48.

152. Bystroff C, Thorsson V, Baker D. (2000) HMMSTR: a hidden Markov model for local sequence-structure correlations in proteins. *J Mol Biol* **301**: 173-190.

153. Sali A, Blundell TL. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**: 779-815.

154. Guex N, Peitsch MC. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**: 2714-2723.

155. Zagorski MG, Yang J, Shao H, Ma K, Zeng H, Hong A. (1999) Methodological and chemical factors affecting amyloid beta peptide amyloidogenicity. *Methods Enzymol* **309**: 189-204.

156. LeVine H, 3rd. (1999) Quantification of beta-sheet amyloid fibril structures with thioflavin T. *Methods Enzymol* **309**: 274-284.

157. Alayash AI, Patel RP, Cashon RE. (2001) Redox reactions of hemoglobin and myoglobin: biological and toxicological implications. *Antioxid Redox Signal* **3**: 313-327.

158. Eliezer D, Wright PE. (1996) Is apomyoglobin a molten globule? Structural characterization by NMR. *J Mol Biol* **263**: 531-538.

159. Eliezer D, Yao J, Dyson HJ, Wright PE. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat Struct Biol* **5**: 148-155.

160. Lecomte JT, Kao YH, Cocco MJ. (1996) The native state of apomyoglobin described by proton NMR spectroscopy: the A-B-G-H interface of wild-type sperm whale apomyoglobin. *Proteins* **25**: 267-285.

161. Lecomte JT, Sukits SF, Bhattacharya S, Falzone CJ. (1999) Conformational properties of native sperm whale apomyoglobin in solution. *Protein Sci* **8**: 1484-1491.

162. Baichoo N, Heyduk T. (1997) Mapping conformational changes in a protein: application of a protein footprinting technique to cAMP-induced conformational changes in cAMP receptor protein. *Biochemistry* **36**: 10830-10836.

163. Fenyo D, Zhang, W., Chait, B.T., Beavis, R.C. (1996) Internet-based analytical chemistry resources: A model project. *Anal Chem* **68**: 721A-726A.

164. Chance MR. (2001) Unfolding of apomyoglobin examined by synchrotron footprinting. *Biochem Biophys Res Commun* **287**: 614-621.

165. Goldsmith SC, Guan JQ, Almo S, Chance M. (2001) Synchrotron protein footprinting: a technique to investigate protein-protein interactions. *J Biomol Struct Dyn* **19**: 405-418.

166. Guan JQ, Vorobiev S, Almo SC, Chance MR. (2002) Mapping the G-actin binding surface of cofilin using synchrotron protein footprinting. *Biochemistry* **41**: 5765-5775.

167. Kiselar JG, Janmey PA, Almo SC, Chance MR. (2003) Visualizing the Ca2+-dependent activation of gelsolin by using synchrotron footprinting. *Proc Natl Acad Sci U S A* **100:** 3942-3947.

168. Neves-Petersen MT, Gryczynski Z, Lakowicz J, et al. (2002) High probability of disrupting a disulphide bridge mediated by an endogenous excited tryptophan residue. *Protein Sci* **11:** 588-600.

169. Young AC, Tilton RF, Dewan JC. (1994) Thermal expansion of hen egg-white lysozyme. Comparison of the 1.9 A resolution structures of the tetragonal form of the enzyme at 100 K and 298 K. *J Mol Biol* **235:** 302-317.

170. Brownlow S, Morais Cabral JH, Cooper R, et al. (1997) Bovine beta-lactoglobulin at 1.8 A resolution--still an enigmatic lipocalin. *Structure* **5:** 481-495.

171. Zhao X, Muller EG, Rothstein R. (1998) A suppressor of two essential checkpoint genes identifies a novel protein that negatively affects dNTP pools. *Mol Cell* **2:** 329-340.

172. Chabes A, Domkin V, Thelander L. (1999) Yeast Sml1, a protein inhibitor of ribonucleotide reductase. *J Biol Chem* **274:** 36679-36683.

173. Zhao X, Chabes A, Domkin V, Thelander L, Rothstein R. (2001) The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. *Embo J* **20:** 3544-3553.

174. Ouspenski, II, Elledge SJ, Brinkley BR. (1999) New yeast genes important for chromosome integrity and segregation identified by dosage effects on genome stability. *Nucleic Acids Res* **27:** 3001-3008.

175. Weinert TA, Kiser GL, Hartwell LH. (1994) Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. *Genes Dev* **8**: 652-665.

176. Weinert TA, Hartwell LH. (1993) Cell cycle arrest of cdc mutants and specificity of the RAD9 checkpoint. *Genetics* **134**: 63-80.

177. Zhao X, Georgieva B, Chabes A, et al. (2000) Mutational and structural analyses of the ribonucleotide reductase inhibitor Sml1 define its Rnr1 interaction domain whose inactivation allows suppression of mec1 and rad53 lethality. *Mol Cell Biol* **20**: 9076-9083.

178. Moult J, Hubbard T, Bryant SH, Fidelis K, Pedersen JT. (1997) Critical assessment of methods of protein structure prediction (CASP): round II. *Proteins* **Suppl 1:** 2-6.

179. Moult J, Hubbard T, Fidelis K, Pedersen JT. (1999) Critical assessment of methods of protein structure prediction (CASP): round III. *Proteins* **Suppl 3:** 2-6.

180. Moult J, Fidelis K, Zemla A, Hubbard T. (2001) Critical assessment of methods of protein structure prediction (CASP): round IV. *Proteins* **Suppl 5:** 2-7.

181. Bernstein FC, Koetzle TF, Williams GJ, et al. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* **112**: 535-542.

182. Baker D, Sali A. (2001) Protein structure prediction and structural genomics. *Science* **294**: 93-96.

183. Xu Y, Xu D, Crawford OH, Einstein JR. (2000) A computational method for NMR-constrained protein threading. *J Comput Biol* **7**: 449-467.

147

184.   Rashidzadeh H, Khrapunov S, Chance MR, Brenowitz M. (2003) Solution structure and interdomain interactions of the Saccharomyces cerevisiae "TATA binding protein" (TBP) probed by radiolytic protein footprinting. *Biochemistry* **42:** 3655-3665.

185.   Reichard P. (1988) Interactions between deoxyribonucleotide and DNA synthesis. *Annu Rev Biochem* **57:** 349-374.

186.   Bonneau R, Tsai J, Ruczinski I, et al. (2001) Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins* **Suppl 5:** 119-126.

187.   Wetzel R. (2002) Ideas of order for amyloid fibril structure. *Structure* **10:** 1031-1036.

188.   Betts S, King, J. (1999) There's a right way and a wrong way: in vivo and in vitro folding, misfolding, and subunit assembly of the P22 tailspike. *Structure* **7:** R131-R139.

189.   True HL, Lindquist SL. (2000) A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407:** 477-483.

190.   Chapman MR, Robinson LS, Pinkner JS, et al. (2002) Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science* **295:** 851-855.

191.   Sunde M, Blake C. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem* **50:** 123-159.

192.   Perutz MF, Finch JT, Berriman J, Lesk A. (2002) Amyloid fibers are water-filled nanotubes. *P Natl Acad Sci USA* **99:** 5591-5595.

193.   Fraser PE, Mclachlan DR, Surewicz WK, et al. (1994) Conformation and Fibrillogenesis of Alzheimer a-Beta Peptides with Selected Substitution of Charged Residues. *J Mol Biol* **244:** 64-73.

194.   Wille H, Michelitsch MD, Guenebaut V, et al. (2002) Structural studies of the scrapie prion protein by electron crystallography. *P Natl Acad Sci USA* **99:** 3563-3568.

195.   Kheterpal I, Williams A, Murphy C, Bledsoe B, Wetzel R. (2001) Structural features of the Abeta amyloid fibril elucidated by limited proteolysis. *Biochemistry* **40:** 11757-11767.

196.   Serpell LC, Blake CC, Fraser PE. (2000) Molecular structure of a fibrillar Alzheimer's A beta fragment. *Biochemistry* **39:** 13269-13275.

197.   Malinchik SB, Inouye H, Szumowski KE, Kirschner DA. (1998) Structural analysis of Alzheimer's beta(1-40) amyloid: protofilament assembly of tubular fibrils. *Biophys J* **74:** 537-545.

198.   Antzutkin ON, Balbach JJ, Tycko R. (2003) Site-Specific Identification of Non-beta-Strand Conformations in Alzheimer's beta-Amyloid Fibrils by Solid-State NMR. *Biophys J* **84:** 3326-3335.

199.   Antzutkin ON, Leapman RD, Balbach JJ, Tycko R. (2002) Supramolecular structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry* **41:** 15436-15450.

200. Petkova AT, Ishii Y, Balbach JJ, et al. (2002) A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci US A* **99:** 16742-16747.

201. Balbach JJ, Petkova AT, Oyler NA, et al. (2002) Supramolecular structure in full-length Alzheimer's beta-amyloid fibrils: evidence for a parallel beta-sheet organization from solid-state nuclear magnetic resonance. *Biophys J* **83:** 1205-1216.

202. Balbach JJ, Ishii Y, Antzutkin ON, et al. (2000) Amyloid fibril formation by A beta 16-22, a seven-residue fragment of the Alzheimer's beta-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* **39:** 13748-13759.

203. Antzutkin ON, Balbach JJ, Leapman RD, Rizzo NW, Reed J, Tycko R. (2000) Multiple quantum solid-state NMR indicates a parallel, not antiparallel, organization of beta-sheets in Alzheimer's beta-amyloid fibrils. *Proc Natl Acad Sci U S A* **97:** 13045-13050.

204. Tycko R. (2000) Solid-state NMR as a probe of amyloid fibril structure. *Curr Opin Chem Biol* **4:** 500-506.

205. Spencer RG, Halverson KJ, Auger M, McDermott AE, Griffin RG, Lansbury PT, Jr. (1991) An unusual peptide conformation may precipitate amyloid formation in Alzheimer's disease: application of solid-state NMR to the determination of protein secondary structure. *Biochemistry* **30:** 10382-10387.

206. Lansbury PT, Jr., Costa PR, Griffiths JM, et al. (1995) Structural model for the beta-amyloid fibril based on interstrand alignment of an antiparallel-sheet comprising a C-terminal peptide. *Nat Struct Biol* **2:** 990-998.

207. Benzinger TL, Gregory DM, Burkoth TS, et al. (1998) Propagating structure of Alzheimer's beta-amyloid(10-35) is parallel beta-sheet with residues in exact register. *Proc Natl Acad Sci U S A* **95:** 13407-13412.

208. Clemmons DR, Maile LA. (2003) Minireview: Integral membrane proteins that function coordinately with the insulin-like growth factor I receptor to regulate intracellular signaling. *Endocrinology* **144:** 1664-1670.

209. Goder V, Spiess M. (2001) Topogenesis of membrane proteins: determinants and dynamics. *FEBS Lett* **504:** 87-93.

210. Koebnik R, Locher KP, Van Gelder P. (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol* **37:** 239-253.

211. van Geest M, Lolkema JS. (2000) Membrane topology and insertion of membrane proteins: search for topogenic signals. *Microbiol Mol Biol Rev* **64:** 13-33.

212. Robinson PJ. (1997) Signal transduction via GPI-anchored membrane proteins. *Adv Exp Med Biol* **419:** 365-370.

213. Stanley KK. (1996) Regulation of targeting signals in membrane proteins. [review]. *Mol Membr Biol* **13:** 19-27.

214. Agre P, King LS, Yasui M, et al. (2002) Aquaporin water channels--from atomic structure to clinical medicine. *J Physiol* **542:** 3-16.

215. Mitra AK. (2001) Three-Dimensional Organization of the aquaporin water channel: what can structure tell us about function? *Vitam Horm* **62:** 133-166.

216. Engel A, Fujiyoshi Y, Agre P. (2000) The importance of aquaporin water channel protein structures. *Embo J* **19:** 800-806.

217. Verkman AS, Mitra AK. (2000) Structure and function of aquaporin water channels. *Am J Physiol Renal Physiol* **278:** F13-28.

218. Sansom MS, Kerr ID, Law R, Davison L, Tieleman DP. (1998) Modelling the packing of transmembrane helices: application to aquaporin-1. *Biochem Soc Trans* **26:** 509-515.

219. Ott CM, Lingappa VR. (2002) Integral membrane protein biosynthesis: why topology is hard to predict. *J Cell Sci* **115:** 2003-2009.

220. VanDongen HM, VanDongen AM. (1999) Determination of membrane topology of glutamate receptors. *Methods Mol Biol* **128:** 155-166.

221. Jung H. (1998) Topology and function of the Na+/proline transporter of Escherichia coli, a member of the Na+/solute cotransporter family. *Biochim Biophys Acta* **1365:** 60-64.

222. Arias HR. (1997) Topology of ligand binding sites on the nicotinic acetylcholine receptor. *Brain Res Brain Res Rev* **25:** 133-191.

223. Turk E, Wright EM. (1997) Membrane topology motifs in the SGLT cotransporter family. *J Membr Biol* **159:** 1-20.

224. Hemrika W, Wever R. (1997) A new model for the membrane topology of glucose-6-phosphatase: the enzyme involved in von Gierke disease. *FEBS Lett* **409:** 317-319.

225.   Tusnady GE, Bakos E, Varadi A, Sarkadi B. (1997) Membrane topology distinguishes a subfamily of the ATP-binding cassette (ABC) transporters. *FEBS Lett* **402:** 1-3.

226.   Stokes DL, Taylor WR, Green NM. (1994) Structure, transmembrane topology and helix packing of P-type ion pumps. *FEBS Lett* **346:** 32-38.

227.   Hughes TE. (1994) Transmembrane topology of the glutamate receptors. A tale of novel twists and turns. *J Mol Neurosci* **5:** 211-217.

228.   Lingappa VR. (1991) Control of protein topology at the endoplasmic reticulum. *Cell Biophys* **19:** 1-15.

229.   Franchet-Beuzit J, Spotheim-Maurizot M, Sabattier R, Blazy-Baudras B, Charlier M. (1993) Radiolytic footprinting. Beta rays, gamma photons, and fast neutrons probe DNA-protein interactions. *Biochemistry* **32:** 2104-2110.

230.   Gowen B, Bamford JK, Bamford DH, Fuller SD. (2003) The tailless icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. *J Virol* **77:** 7863-7871.

VITA

Joshua S. Sharp began his collegiate education at the University of Tennessee—Knoxville in August, 1995, majoring in microbiology. He began his practical laboratory experience working under Dr. W. Stuart Riggsby in 1996 on the phylogeny of asexual yeast. He continued his undergraduate work experience in the laboratory of Dr. Jeffrey Becker, working in molecular genetics of yeast as applied to the G-protein coupled receptor Ste2p. Joshua Sharp graduated *cum laude* with a B.S. in Microbiology in May, 1999. He began his graduate education at Rutgers University in August, 1999 working in the laboratory of Dr. Stephen Anderson on heterologous expression of the β-amyloid precursor protein. He transferred to the Graduate School of Genome Science and Technology at the University of Tennessee and Oak Ridge National Laboratory in January, 2000, performing graduate work in developing methods for the rapid characterization of protein surfaces by free radical labeling and mass spectrometry. Upon completion of his graduate program in October, 2003, he accepted a postdoctoral position as a mass spectrometrist working on development and application of mass spectrometry-based methods for protein structure characterization under the supervision of Dr. Kenneth Tomer of the Laboratory of Structural Biology of the National Institute of Environmental Health Science in Research Triangle Park, North Carolina.