



LIBRARY
Michigan State
University

This is to certify that the

thesis entitled

**MACROMOLECULAR SIZE DETERMINATION BY
SCANNING FORCE MICROSCOPY**

presented by

Martha Gilchrist

has been accepted towards fulfillment
of the requirements for

M. S. degree in Chemistry

Major professor

Date 4/25/96

**PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.**

DATE DUE	DATE DUE	DATE DUE
NOV 02 1998		

MSU Is An Affirmative Action/Equal Opportunity Institution

c:\circ\d\date.due.pm3-p.1

**MACROMOLECULAR SIZE DETERMINATION BY SCANNING FORCE
MICROSCOPY**

By

Martha Gilchrist

A THESIS

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

MASTER OF SCIENCE

Department of Chemistry

1996

ABSTRACT

MACROMOLECULAR SIZE DETERMINATION BY SCANNING FORCE MICROSCOPY

By

Martha Gilchrist

A fast and precise method of determining protein size has been developed using Scanning Force Microscopy (SFM). The technique, termed Random Adsorption Molecular Sizing (RAMS), has provided nanometer resolution for examinations of macromolecules under conditions in which the sample does not require the extensive fixation protocols normally associated with other high resolution techniques such as TEM or SEM. A model is proposed which allows accurate determination of the three dimensional size of molecules randomly oriented on a flat surface. This makes possible studies on changes in quaternary structure due to point mutations and ligand-induced oligomerization. Using RAMS, the molecular size for the enzyme ADP-glucose pyrophosphorylase (ADPgpp) is determined, and the enzyme is shown to undergo a change from tetramer to octamer caused by a single point mutation. RAMS is also used to study ligand-induced oligomerization in two molecules, Streptavidin and Wheat Germ Agglutinin.

ACKNOWLEDGEMENTS

For such an interdisciplinary work as this, I have many people to thank. First, thank you to Professor Melvin Schindler, who has been my main link to Biochemistry and has furnished many ideas for this project. He has been key in the editing of the paper, to be submitted to the journal *Biochemistry*, on which this thesis is based. Professor Jack Preiss has been a valuable source of advice on protein structure in general and his protein, the ADP glucose pyrophosphorylase, in particular. The help of Dr. Min Wu in Preiss's lab was invaluable when I was doing the electrophoresis. He gave me both instruction and space at his lab bench. I am very thankful to Professor John Wang, who pointed out some important structural features of Con A. Thank you to Joseph Leykam, whose generosity in teaching me to use the HPLC and allowing me to use his equipment has made this project run more smoothly. Finally, the enthusiasm and advice from Professor Marcos Dantus has been the driving force in developing this project, and I truly appreciate his patience and encouragement.

I am grateful to the funding agencies which provided support for me for several semesters and money with which to purchase the Scanning Force Microscope. This research was partially funded by the REF-Protein Structure, Function and Design Award (MSU), the Center for Fundamental Materials Research (MSU) and a Camille and Henry Dreyfus New Faculty Award to Professor Dantus.

TABLE OF CONTENTS

LIST OF TABLES.....	iii
LIST OF FIGURES.....	iv
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 IMAGING THE MOLECULAR DIMENSIONS AND OLIGOMERIZATION OF SINGLE PROTEIN MOLECULES BY RANDOM ADSORPTION MOLECULAR SIZING (RAMS) MICROSCOPY.....	9
CHAPTER 3 LIGAND BINDING STUDIES.....	48
CHAPTER 4 CONCLUSIONS.....	66
APPENDIX A: SOLID FILMS OF PROTEIN.....	74
APPENDIX B: VISUAL BASIC PROGRAM TO AUTOMATICALLY MEASURE AND TABULATE THE HEIGHT OF MOLECULES IN AN SFM IMAGE.....	78
APPENDIX C: VISUAL BASIC PROGRAM TO AUTOMATICALLY OPEN SEVERAL TEXT FILES AND PROCESS THE DATA THEN CLOSE THE FILES.....	87
APPENDIX D: C++ PROGRAM TO CONVERT SFM IMAGE INTO AN X, Y, ARRAY SUITABLE TO READ INTO EXEL.....	88
APPENDIX E: MATHEMATICA ROUTINE TO READ A FILE OF TABULATED RAMS DATA AND FIT THE DATA TO GAUSSIAN CURVES.....	90

LIST OF TABLES

Table 1: Summary of sizes of colloidal gold particles by TEM and SFM.....18

**Table 2: Summary of the protein dimensions derived from RAMS microscopy and x-ray
crystallography.....47**

LIST OF FIGURES

Figure 1: Schematic representation of the main working parts of the SFM.....	7
Figure 2: Diagram of tip convolution effect and error due to misalignment of tip and protein.....	8
Figure 3: Three orientations of a molecule the shape of Con A arranged on a mica lattice.....	30
Figure 4: 2.5 by 2.5 μm image of WGA on mica.....	31
Figure 5: Correlation between elongation and the probability that the two tallest dimensions are vertical..	32
Figure 6: RAMS on 4.9, 9.0, and 18.1 nm colloidal gold fitted to Gaussians.....	33
Figure 7: RAMS on Ferritin.....	34
Figure 8: RAMS on WGA fitted to (A) monomer and (B) dimer dimensions.....	35
Figure 9: RAMS on gluteraldehyde treated WGA fitted to (A) monomer and (B) 55% monomer and 45% dimer.....	36
Figure 10: RAMS on Con A fitted to (A) monomer (B) dimer.....	37
Figure 11: RAMS on Con A fitted to (A) tetramer (B) 13% monomer, 67% dimer, 20% tetramer.....	38
Figure 12: HPLC on (A) 3 $\mu\text{g/ml}$ Con A (B) 5 mg/ml FITC-sCon A.....	39
Figure 13: RAMS on Protein A fitted to (A) monomer (B) part A plus 24% from 24 and 40 $\text{\AA}$	40
Figure 14: RAMS on Streptavidin fitted to monomer.....	41
Figure 15: RAMS on gluteraldehyde treated Streptavidin fitted to (A) monomer (B) 66% monomer 34% dimer.....	42
Figure 16: Non-denaturing gel on ADPgpp wild type, G336D, and <i>Anabaena</i>	43
Figure 17: SFM data on ADPgpp wild type, G336D, and <i>Anabaena</i>	43
Figure 18: RAMS data on ADPgpp wild type.....	44
Figure 19: RAMS data on ADPgpp G336D mutant.....	45
Figure 20: RAMS data on ADPgpp from <i>Anabaena</i> fitted to (A) monomer (B) 68% monomer 31% tetramer.....	46
Figure 21: Pre-dilution mono-GLcNAc treated WGA fitted to (A) monomer (B) dimer.....	56

Figure 22: Pre-dilution mono-GLcNac treated WGA fitted to 50% monomer, 50% dimer.....	58
Figure 23: Post-dilution mono-GLcNac treated WGA fitted to (A) monomer (B) dimer.....	59
Figure 24: Pre-dilution di-GLcNac treated WGA fitted to 68% monomer, 32% dimer.....	60
Figure 25: Post-dilution di-GLcNac treated WGA fitted to (A) monomer (B) dimer.....	61
Figure 26: Pre-dilution tri-GLcNac treated WGA fitted to 81% monomer 19% dimer.....	62
Figure 27: Post-dilution tri-GlcNac treated WGA fitted to (A) monomer (B) dimer.....	63
Figure 28: Pre-dilution biotin treated Streptavidin fitted to tetramer.....	64
Figure 29: Post-dilution biotin treated Streptavidin fitted to (A) monomer (B) dimer.....	65
Figure 30: Post-dilution biotin treated Streptavidin fitted to (A) tetramer (B) 24% monomer 76% dimer...	66
Figure 31: Two-dimensional array of Streptavidin.....	77
Figure 32: Two-dimensional array of ADPgpp wild type.....	77

CHAPTER 1: INTRODUCTION

Protein aggregation has been shown to play a key role in regulating enzyme activity for many proteins including Ca^{2+} -ATP-ase^{1,2} and ADP-glucose pyrophosphorylase,³ and ligand-induced oligomerization has been suggested to cause enhanced binding for a large number of DNA binding proteins such as p53,⁴ Par A,⁵ Mu B,⁶ Rec A,⁷ and many others. To understand the function of such proteins it is critical to have a good understanding of their molecular size and oligomerization state under physiologically relevant conditions when the molecules are subjected to treatment with appropriate ligands. Unfortunately, few techniques are presently available with the necessary sensitivity and resolution to measure molecular dimensions and subunit organization under conditions that can maintain biological activity. Here we present a new technique which can quickly determine the size and oligomerization state for virtually any water-soluble protein molecule.

This research has focused on the development of techniques aimed at obtaining measurements of biological macromolecular structures by Scanning Probe Microscopy (SPM). The primary experiment was to develop a method for determining the size of a protein molecule using Scanning Force Microscopy (SFM). The method developed can give a reliable three dimensional size for virtually any water soluble protein only a few minutes after sample preparation. In addition to general structural studies, this method can be used to study ligand-induced conformational changes in a variety of molecules. When applied to study mutant varieties of a protein this method has also furnished information about the relative stability of the quaternary structure of the mutants, because if quaternary structure is intact this will affect the apparent size of the protein. Also, a model has been devised which correctly predicts the dimensions and orientation of protein molecules when they are randomly oriented on the mica surface.

SPM uses a sharp probe in close proximity to the sample to obtain high resolution images at a given x, y, position. The signal may be electrical conductance, as in scanning

tunneling microscopy (STM),⁸ attractive or repulsive forces, as in SFM,⁹ or absorption or transmission of light, as in near-field scanning optical microscopy (NSOM).¹⁰ SFM does not require the sample to be conductive as in STM or frozen, dehydrated, and coated with metal as in electron microscopy (EM). SPM is therefore an ideal tool for studying biological macromolecules and structures,^{11,12} which may be disrupted or decompose when subjected to metal coating or vacuum drying. The resolution that may be obtained by SFM is similar to that obtained by EM, on the angstrom level. Using SFM we have been able to image, with no special treatment, biological structures that still hold their solvation water.

SCANNING FORCE MICROSCOPY

Basic Operation. Scanning force microscopy works by measuring the interaction between a surface and a sharp tip on a cantilever. The cantilevers with integrated tips are commercially available from several companies; the ones used in these experiments were either silicon nitride with pyramidal tips having an aspect ratio of about 1 and a radius of curvature as small as 50 nm, or silicon with higher aspect ratio tips and a radius of curvature as small as 10 nm. In Figure 1 the main working parts of the SFM are pictured. These parts are the piezoelectric scanner, a, which moves the sample, b, underneath the tip, c, which is mounted on a flexible cantilever spring, d. The cantilever is deflected by the interaction forces between the tip and the sample. A diode laser, e, shines a beam onto the top of the cantilever which reflects the beam into a position sensitive photodiode, f. The proportion of the beam in each half of the photodiode is determined by the angle of the cantilever deflected by the sample. The position sensitive photodiode monitors changes in height by detecting changes in the direction of reflection of the diode laser reflecting off the top of the cantilever. The signal from the position sensitive photodiode is fed into a feedback loop to regulate the height of the sample, so that as the

sample is scanned constant force is maintained between the sample and the tip. The amount that the sample has to be moved to maintain the tip at constant height and constant force at a given x , y , position is recorded as the z value of the topographic scan. If the feedback loop is operating properly the deflection of the cantilever changes very little because the changes in topography are compensated for by expansion or contraction of the scanner tube. This technique is called constant-force contact SFM.

Alternatively, the cantilever is vibrated with an amplitude of about 30 Å at its resonant frequency (360 kHz for the 2 µm thick, 200 µm long triangular silicon nitride cantilevers) about 20 Å from the surface being scanned. As the tip approaches the sample, the resonant frequency of the cantilever changes because of the attractive Van der Waals forces between the tip and surface and this change in frequency is detected by the photodiode. The feedback system compensates for the change in distance between the tip and sample, and the amount of vertical movement necessary to return the resonant frequency to its previous value is registered as the change in height of the sample at a given x , y , position. The tip is typically maintained at a distance of 15-100Å above the surface. This mode of operation is called non-contact SFM, and is usually less disruptive to a surface because the forces between tip and sample are smaller. If a small amount of debris is picked up by the tip during scanning, this will disrupt the resonance frequency in non-contact mode and prevent the acquiring of images until the debris is removed or the resonance frequency is reset. Also, if the sample has a thin water layer on it, the capillary forces may be strong enough to pull the tip into contact with the sample making non-contact imaging impossible. Contact mode has the disadvantage of having the tip actually touching the sample, introducing the possibility of damaging a soft sample. Lateral forces in contact mode may be high, sweeping insecurely fastened objects in front of the tip as it is scanned. However, most published research using SFM is done in contact mode because non-contact mode tends to be unstable and more easily disrupted.

Tapping mode is a recently introduced mode of operation^{13,14} which combines non-contact methodology with some additional measure of stability due to intermittent contact between the tip and the surface. This mode of operation eliminates most of the lateral force induced by contact imaging, but still allows the possibility of sample damage due to vertical compression. It is still less stable than normal contact imaging due to the resonance frequency dependence of the imaging, but more stable than non-contact imaging because the forces measured in acquiring the image are much larger.

Capabilities and Limitations of SFM. Using SFM, one can measure height with a resolution of better than one angstrom, and on a flat crystalline surface, lateral dimensions can be measured with a resolution of better than 3 Å. The best resolution is obtained on flat crystalline samples because in these cases the image is acquired using only the very endmost atom or atoms of the tip. On a surface with particles on the order of the size of the probe, the image of the particle is convoluted with a reverse image of the probe. Since the tip is measuring the surface by scanning in contact, the finite dimensions of the tip become part of the image.^{15,16,17,18} A simple algorithm has been proposed to separate out tip convolution effects when imaging spherical particles with known size. Dimensions can be calculated as

$$d = w^2/4h \quad (1)$$

where d is the tip diameter and w is the apparent width measured for sample height h .¹⁸ Once the tip diameter and geometry are known, the image can be deconvoluted and probe effects separated from the true image of the sample. In these kinds of images, the sharper the tip compared to the sample size, the less the error caused by the probe convolution. At this time, the sharpest tip commercially available has a radius of curvature of 10 nm, a

value bigger than the diameter of most protein molecules. The error introduced by the tip convolution can widen the horizontal measurements of 50 Å diameter spheres by as much as 950 Å using a tip with radius of 500 Å. Fortunately, this tip convolution does not distort measurements of vertical dimensions. We have designed our protein sizing experiment based on vertical measurements only. Figure 2 illustrates how horizontal measurements are broadened due to this tip convolution effect.

Vertical Probe Tip Corrections. Height measurements of small spherical particles with SFM are slightly reduced when images are taken at low resolution. This measurement effect occurs because the end of the tip, which can be modeled as a sphere, is not necessarily at the apex of the protein. For example, in an image 2.5 microns square which is 256 by 256 pixels, the end of the tip could be as far as 48 Å from the molecule if it is located exactly between two scan lines. This effect can be reduced by using tips with a larger radius or having a larger number of pixels in a scan. In Figure 2, the idealized shapes of the end of the tip and a spherical protein molecule are illustrated. Because of the finite number of points which can be recorded, the apex of the protein will usually not be directly over the end of the tip, therefore the measured height of the particle will be less than the actual height of the particle. The discrepancy, e , can be calculated as follows:

$$e = (R + r) - ((R + r)^2 - X^2)^{1/2} \quad (2)$$

where R is the radius of curvature of the tip, r is the radius of a spherical protein, and X is the horizontal distance between the center of the tip and the center of the protein when this distance is smallest. For a very fine tip with 100 Å radius of curvature, a protein with a diameter of 40 Å, and taking an image with 256 by 256 points and 2.5 micrometer square, the height deviation can be as much as 10 Å. However, since the molecules are randomly

distributed on the surface, the effect on a set of measurements will be significantly less. For 500 random horizontal distances between 0 and 48 Å from the apex of the protein to the end of the tip, the average error is only 3.5 Å. The typical radius of the tips used in these experiments was 750 Å, as calculated from width and height measurements for spherical particles.¹⁸ The average height error expected for our measurements due to tip/protein misalignment is therefore 0.5 Å.

Because of the large tip convolution effect, SFM is most useful when used to image the height of very small particles spread on a surface, the surface topography of regular crystalline surfaces or the three dimensional structure of particles and surfaces with features much larger than the dimensions of the probe, or to measure forces between surfaces.^{19,20,21,22,23} SFM has developed into a very useful tool for investigating biological samples, having been used to image regular 2-D crystalline arrays of protein at molecular resolution,^{24,25} measure adhesion forces and elasticity of lysozyme on mica,²⁶ and observe the changes in a cell upon infection by a virus.²⁷ Also, there has been extensive work in imaging isolated DNA molecules^{28,29,30,31} and individual protein molecules.^{32,33,34} Work in our group has concentrated on imaging large molecular structures and has branched into two areas: developing a technique for fast protein sizing and measuring ligand-induced conformational changes in protein molecules.

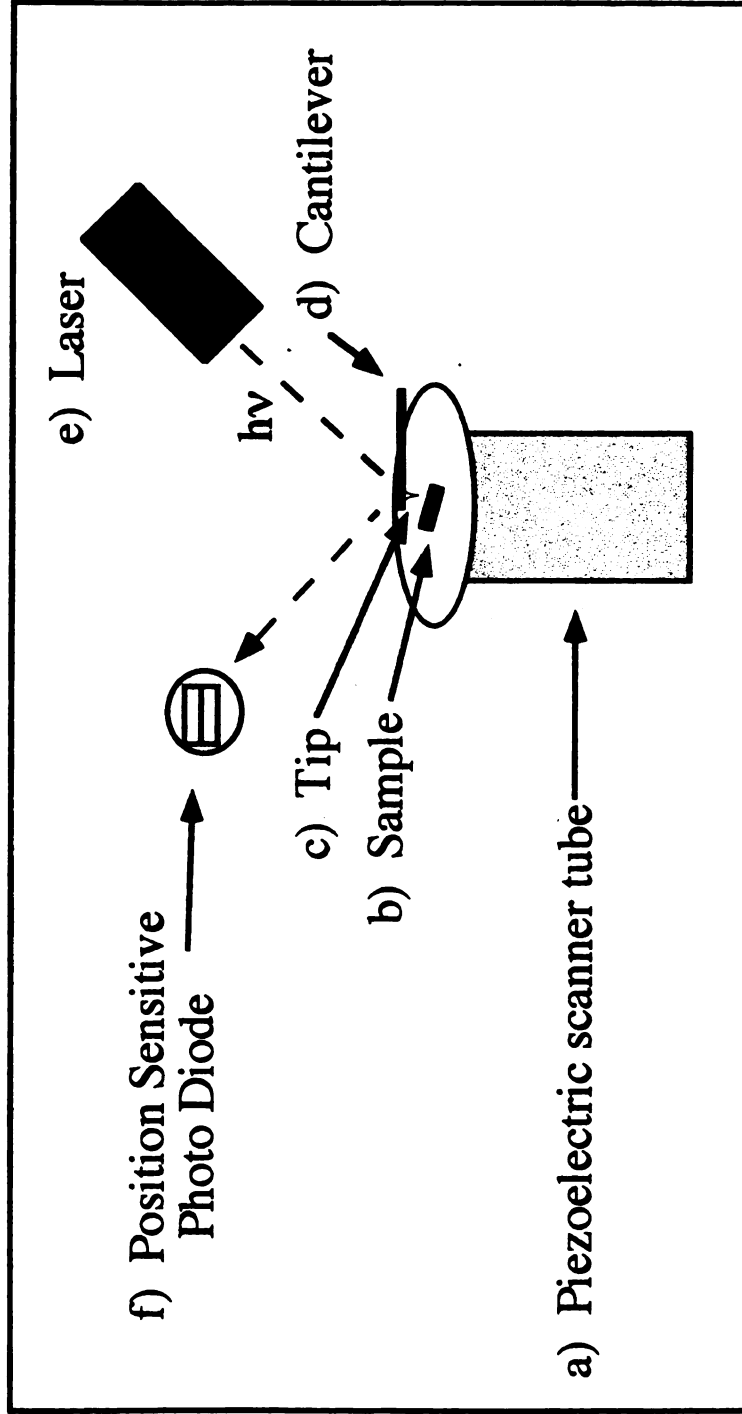


Figure 1: Schematic representation of the main working parts of the SFM

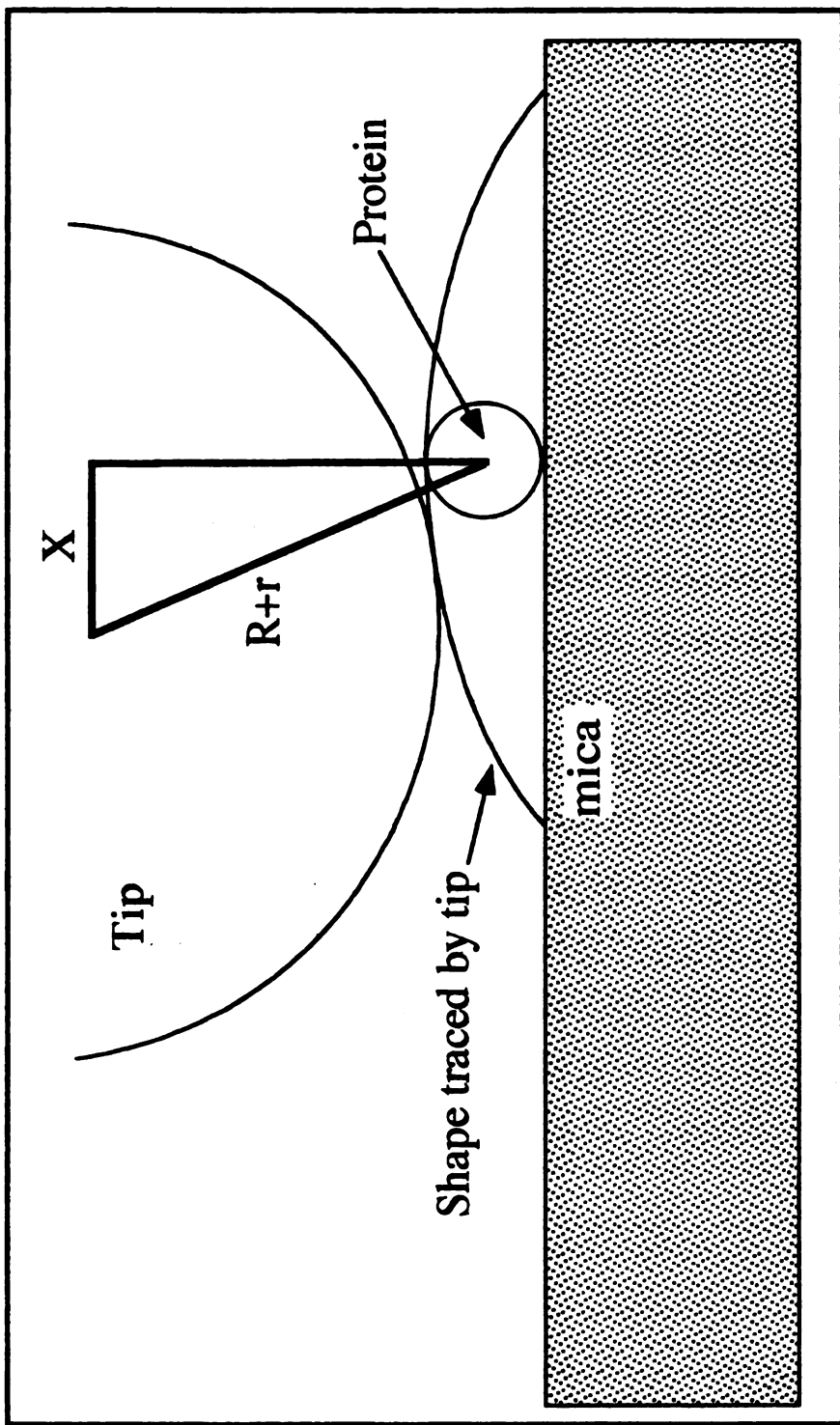


Figure 2: Diagram of tip convolution effect and error due to misalignment of tip and protein

CHAPTER 2: IMAGING THE MOLECULAR DIMENSIONS AND OLIGOMERIZATION OF SINGLE PROTEIN MOLECULES BY RANDOM ADSORPTION MOLECULAR SIZING (RAMS)³⁵ MICROSCOPY*

INTRODUCTION

A variety of analytical techniques have been developed that provide information on the size and shape of proteins. Chromatography, electrophoresis, transmission and scanning electron microscopy (TEM and SEM), light scattering, nuclear magnetic resonance (NMR) and equilibrium ultra-centrifugation are all commonly employed methods to determine the molecular size and shape of proteins.³⁶ Although of considerable importance in the characterization of protein structure, these techniques do not have the ability to provide molecular dimensions of large macromolecules at high resolution, with the exception of TEM and SEM, which require harsh fixation and dehydration protocols, and NMR, which is most effective with small proteins. Such measurements are essential for the characterization of the changes in protein conformation or state of aggregation that have been associated with ligand binding, protein-protein interactions, or activator induced oligomerization. At present, the only technique that can provide angstrom resolution for the direct measurement of molecular dimensions is electron diffraction microscopy.³⁷ To perform such measurements, however, crystals must be prepared from protein solutions. A particular difficulty with this method is that in the absence of a generalized crystal growth protocol for oligomeric and membrane proteins, the utility of this approach has been limited to predominantly soluble low molecular weight proteins.³⁸ In addition, co-crystallization of proteins with ligands is a more challenging task.

To circumvent the need of crystals for diffraction based approaches while still maintaining the capacity for dimensional analysis at high resolution, we have developed an analytical technique that uses scanning force microscopy to measure the dimensions of

* The work presented in this chapter is based entirely on a publication to be submitted- see reference number 35.

individual protein molecules with angstrom resolution. Since its recent introduction, SFM has been extensively used for imaging biomolecules.¹¹ The technique has provided nanometer resolution for examinations of macromolecules under conditions in which the sample does not require the extensive fixation protocols normally associated with other high resolution techniques such as TEM or SEM.⁸

Here we extend the utility of SFM to measurements of molecular dimensions with angstrom resolution. A loss of resolution is normally introduced into SFM measurements as a result of probe tip geometry^{15,16,17,18} as shown in Figure 2, a schematic representation of the tip/protein interaction. To avoid this resolution loss, we have devised a measurement strategy that uniquely measures the height of protein molecules that have randomly adsorbed to the atomically flat surface of mica. Figure 3 exhibits a representative drawing of Concanavalin A dimers arranged in three orientations on a mica surface with an SFM tip positioned on top of one of the molecules. The Statistical Orientation Model (SOM), described in Theory and Data Analysis, assumes that the proteins may be represented as parallelepipeds, and predicts the orientation according to the relative dimensions of the parallelepiped. In Figure 3 the size of the molecules, the tip and the mica lattice are drawn to scale. The technique to determine molecular size presented here has been termed Random Adsorption Molecular Sizing (RAMS).³⁵ The use of the height of adsorbed protein molecules as the defining parameter for molecular dimension makes the RAMS technique minimally dependent on probe tip geometry and extends the range of the SFM from nanometer to angstrom resolution. The relative ease and speed with which these high resolution measurements of molecular dimension may now be performed should provide investigators with new opportunities to examine changes in protein conformation and oligomerization that depend on protein concentration and ligand or allosteric activator binding. A further advantage of RAMS for structure function studies of proteins is that the method provides a rapid screening protocol in which the molecular dimensions of large

numbers of recombinant and site-mutated proteins may now be determined in a minimum amount of time (15 minutes) with a minimum amount of material (~100 molecules).

MATERIALS AND METHODS

Materials. Colloidal gold particles were obtained from Sigma (St. Louis, MO) and used without further purification. Three sizes of gold particles were used for this experiment: $49 \pm 6 \text{ \AA}$ (3.8×10^{13} particles/ml), $90 \pm 12 \text{ \AA}$ (6.0×10^{12} particles/ml), and $181 \pm 8 \text{ \AA}$ (6.7×10^{11} particles/ml). Protein A, wheat germ agglutinin (WGA), and Streptavidin were obtained from Sigma (St. Louis, MO). Ferritin was obtained from Polysciences (Warrington, PA) as the cationized protein. Concanavalin A (Con A) was a gift from Professor John Wang, Department of Biochemistry, Michigan State University. ADP-glucose pyrophosphorylase (ADPgpp) enzymes were a gift from Professor Jack Preiss, Michigan State University. HEPES was from Sigma (St. Louis, MO) and the MgCl_2 was ACS grade from Columbus Chemical Industries (Columbus, OH). The water used in preparing solutions was purified with a MilliQ MilliPore water purification unit.

High Performance Liquid Chromatography. HPLC was performed on Con A and fluorescein derivatized succinylated Con A (FITC-sCon A). The Beckman SEC-2000 size exclusion column was used with an elution buffer of 40 mM Tris, 300 mM NaCl, and 300 mM glucose pH 6.8. The flow rate was 0.8 ml/minute and the protein was monitored spectrophotometrically at 214 nm as it eluted from the column. Molecular size standards included Trypsin, (27 kDa), and bovine serum albumin (68 kDa), and FITC-sCon A (51 kDa). The concentration of Con A and FITC-sCon A was 3 $\mu\text{g/ml}$ and 5 mg/ml respectively dissolved in 6 mM NaCl, 10 mM HEPES pH 7.5.

Non-denaturing gel electrophoresis. Non-denaturing gel electrophoresis was performed on the wild type and G336D mutant of the *E. coli* ADPgp enzyme as well as the enzyme from *Anabaena*. The electrophoresis was performed essentially as described by Ornstein and Davis,³⁹ but without exposing the protein to SDS or DTT in the buffers, and without heating the proteins. The protein was run on 8% and 10% polyacrylamide gels. Resolving gels were made 400 mM in Tris pH 8.8 and polymerized with 0.1% ammonium persulfate and 0.04% TEMED, and either 8% or 10% acrylamide. Stacking gels were 5% acrylamide and 130 mM in Tris pH 6.8.

Preparation of Samples for Height Measurements with SFM. Measurements of protein height required an atomically flat substrate. In this study, we used muscovite mica obtained from Ward's Natural Science Establishment, Inc. (Rochester, NY.). When freshly cleaved, mica has a net negative surface charge. Treating a freshly cleaved surface of mica with a 5 mM MgCl₂ solution results in the replacement of endogenous K⁺ ions with Mg²⁺ making the surface more positively charged. The positively charged surface has been demonstrated to enhance adsorption of macromolecules.⁴⁰ Ten μ l of an approximately 3 μ g per ml solution of the protein dissolved in 10 mM HEPES, pH 7.5, is deposited onto the mica followed by the addition of 10 μ l of 0.5% glutaraldehyde. Fifteen minutes later, the surface is gently rinsed twice with 200 μ l portions of MilliQ water and allowed to dry for several hours. This preparation yields a dispersed population of individual protein molecules when the mica surface is imaged by SFM. Figure 4 is a typical RAMS scan used to collect protein height measurements, a 2.5 μ m by 2.5 μ m image of WGA. All the images used in our study showed a random distribution of individual particles. The protein concentration was chosen to provide a large population of protein molecules that are adsorbed in different orientations. In the presence of a relative humidity of about 30%, there is a monolayer of

water maintained on the mica surface to hydrate the adsorbed proteins, as described by Hu et. al.⁴¹

Random Adsorption Molecular Sizing (RAMS) Microscopy.

The microscope used for this work was an Autoprobe CP scanning probe microscope (Park Scientific Instruments), with a five μm high resolution scanner. SFM data were recorded in contact mode under controlled ambient conditions (30% humidity and 21° C). Ultralever cantilevers (Park Scientific Instruments) having a spring constant of approximately 0.06 N/m and an integrated silicon tip with a radius of curvature of approximately 10 nm, according to the manufacturer, were used as probes. The scanner was calibrated for vertical measurements using Tobacco Mosaic Virus (diameter 18 nm) (American Type Culture Collection), and for horizontal measurements using a 1,000 x 1,000 x-y features per millimeter grating.

It was found that the protein molecules adhered to the mica well enough for measurements only when the humidity was less than 40%, probably because high humidity increases the attractive capillary forces between tip and sample.^{34,41,42} In the summer, when ambient humidity could reach 85%, the humidity was controlled by placing the microscope in a loosely sealed glove bag with a hygrometer (Omega model RHDP-1) and flushing the bag with a slow stream of dry nitrogen to maintain a humidity of about 30%. In the winter when ambient humidity was typically less than 40%, measurements were performed without the glove bag and in ambient air. A humidity of less than 20% was found to interfere with the process of tip approach to the sample probably because of static charge buildup between the tip and the charged mica surface.

Height measurements were performed using the minimum force required for stable imaging in order to avoid compression of the molecule which would produce force induced height artifacts in our measurements. Images were taken within the range of the attractive

capillary forces. In order to accomplish this an approach was performed at a high force setting, then the force was lowered to scan the image. Scan size for all images was 2.5 μm by 2.5 μm and 256 by 256 pixels, and scans were recorded at a rate of 2 lines per second.

The height of each molecule was determined by taking the difference in height between the top of the particle and the surrounding substrate. The SFM software permits the user to draw a line at a specific place in the image and measure height differences along the line. For all our measurements the line was drawn parallel to the fast scan direction and placed so that it intersected the highest point on the molecule, enabling a measurement of the height of each molecule. The roughness of the mica surface in the prepared samples was less than 2-3 $\text{\AA}$, a surface distortion that is probably the result of a small amount of residual buffer salts. Occasionally, large clumps of protein were formed instead of the desired dispersed field of individual molecules. Only the heights of isolated and individual particles were measured.

THEORY AND DATA ANALYSIS

Statistical Orientation Model (SOM). Most of the proteins imaged for this work were asymmetrical or oligomeric demonstrating a distribution of measured heights. In order to interpret these single molecule measurements, we devised a statistical orientation model (SOM) to represent the distribution of heights that could be obtained by molecules of specific dimensions binding randomly to the mica. The SOM assumes each protein can be enclosed by a rectangular box (see Figure 3). A favorable interaction between the negatively charged protein molecules and the positively charged mica surface would favor adsorption such that the largest axis (represented by largest surface area) of the protein is in contact with the mica. The surface display of protein orientations that are observed in the RAMS measurements can be easily modeled macroscopically by performing an experiment

dropping a large number of small rectangular wood boxes. Each box represents an individual protein molecule adsorbed with a particular orientation. The parallelepipeds (with increasing dimensions a , b , and c) were allowed to fall under the influence of gravity, and their orientation was tabulated. Comparing the height distribution profile for a group of proteins obtained by the RAMS technique, we found that protein orientation on the mica surface (in the absence of specific chemical recognition between the molecule and the surface) closely resembled the distribution of heights obtained after dropping scale model parallelepipeds onto a surface. A phenomenological relation was found for the probability of measuring a given side of the parallelepiped. For example, the probability of measuring a , the smallest dimension, was found to be proportional to the area of the base ($b \times c$) divided by the distance of that area to the center of mass ($a/2$) such that:

$$P(a) = 2(b \times c)/a \quad (1)$$

In order to test the SOM we found a relation between the product of the normalized probability of measuring side b and side c such that,

$$C \times B = (a/(b \times c))^2 \quad (2)$$

This formula allows us to compare the experimentally measured probabilities B and C to the predicted value calculated from the known values for a , b , and c . The correlation between these measurements is shown in Figure 5, a comparison between the experimentally measured values for C and B (the relative probability that c and b will be the measured height of the block when it is allowed to fall under the influence of gravity) and the calculated $a/(b \times c)^2$ which is a measure of the elongation of the molecule. The relationship predicted by the SOM agrees well with the actual experimental values observed

when rectangular blocks are allowed to fall under the influence of gravity, as shown by a comparison between the experimental points and the straight line with slope of 1.

Using SOM we are able to predict the height distribution from a randomly adsorbed protein given a set of approximate dimensions a , b , c . While there is no physical reason why protein adsorption should closely mimic the distribution of measured heights observed for the wood blocks we have found a very close correspondence. The SOM, therefore, provides a mathematical tool to interpret the height distribution profiles and obtain the relative amounts of each molecular species under analysis e.g. monomer, dimer, tetramer. Deviations from this model by adsorbed proteins imply a preferred orientation that is presumably related to recognition between a protein domain and the surface. The extent of deviation from the SOM may be useful in defining such preferred binding to ligand coated surfaces or ligand induced conformational changes leading to the preferred orientation.

Data analysis consisted of two different procedures, one to determine molecular dimensions and the other to assign tertiary and quaternary structure. The raw data was first tabulated and smoothed with a 5Å window. A combination of Gaussian curves was then used to fit the data, using Peakfit (Jandal Corp.). The center of the Gaussian curves determined the measured molecular dimensions while their area determined their prevalence. In order to assign a tertiary and quaternary structure for the macromolecule the obtained measurements were used to propose the three dimensions (a , b , c) of a parallelepiped which most closely resembled that macromolecule. Given the dimensions of a proposed monomer the sizes of the higher oligomers were predicted and those predictions were checked against the observed dimensions. Once the sizes of monomers, dimers and higher oligomers were determined we used the SOM to obtain the predicted height distribution of such species. Finally, we used a *Mathematica* routine (see Appendix E) to fit the percentage of each oligomer in the data keeping the SOM values fixed. In this way we

have been able to analyze mixtures of different oligomers of a protein and determine their size and concentration.

RESULTS

Measurements of the Height of Colloidal Gold Standards with RAMS. Monodispersed spherical colloidal gold particles were initially employed to demonstrate that the RAMS technique is capable of making sub-nanometer height measurements that are minimally influenced by tip geometry. The particles were chosen because their sizes were comparable to that of protein molecules. Fifty microliters of sample were deposited onto a 1 cm² piece of mica that was mounted on a stainless steel disk for imaging as described in materials and methods, then the sample was rinsed twice with 200 μ l aliquots of MilliQ water and allowed to dry overnight. This preparation resulted in a uniform distribution of gold particles on the mica surface. The results of height measurements on these particles are presented in Figure 6 and Table 1. There is good agreement between the manufacturer's stated size (determined with TEM) and the size measured by SFM. The similarity between standard deviations as calculated for TEM and SFM measurements suggest that the dominant source of error in these measurement may reside in the size distribution of the particles themselves rather than the measurement technique or tip geometry.

Table 1: Summary of sizes of colloidal gold particles by TEM and SFM.

Size Determined by TEM (nm)	Size Determined by SFM (nm)	Size by Gaussian fit to SFM Data (nm)	Number of Particles Measured
4.9 ± 0.6	5.2 ± 1.6	4.8±1.9	353
9.0 ± 1.2	9.9 ± 2.0	10.4±1.6	207
18.1 ± 1.2	17.8 ± 2.8	16.9±1.8	241 total
		22.8±2.6	

Ferritin. To validate the RAMS technique for the measurement of the macromolecular dimensions of proteins, we performed measurements on a diverse group of proteins for which high resolution structures had been previously determined by either electron microscopy or x-ray crystallography. The first protein employed using the RAMS technique was horse spleen Ferritin. Ferritin has been demonstrated to be a roughly spherical molecule comprised of 24 polypeptides (ratios of homologous H and L chains) and bound iron (Fe^{3+}). Ferritin has a molecular weight of 445,000 and it has been shown to have an outer diameter of 120 Å with a hollow center of about 80 Å diameter by electron microscopy⁴³ and x-ray crystallography.⁴⁴ The hollow center is used for iron storage. RAMS Ferritin isolated from horse spleen yielded a distribution of measured heights with a dominant peak at 112 Å (see Figure 7 and Table 2), in excellent agreement with the previously reported results for ferritin. Figure 7 is the RAMS data for Ferritin fitted with three Gaussians centered at 73 Å, 112 Å, and 130 Å, representing the iron core, the intact ferritin molecule, and a presumed Ferritin trimer, respectively. The relative areas under the curves are 20, 41 and 39% representing the iron core, monomer and trimer. The additional peaks in the height distribution indicate a diversity of ferritin populations that are related to the presence of the 80 Å iron oxide core and Ferritin trimers,⁴⁵ all of which have been observed with adsorbed Ferritin samples on mica grids using STEM.

Wheat Germ Agglutinin (WGA). Wheat Germ Agglutinin (WGA) is a plant lectin (i.e. a carbohydrate binding protein that can agglutinate erythrocytes and other types of cells) with affinity for N-Acetylglucosamine (GlcNac), and GlcNac oligomers.⁴⁶ The dimer has two highly specific,⁴⁷ independent binding sites per polypeptide chain.⁴⁸ WGA is composed of two identical subunits with molecular dimensions of the monomer approximately 22 x 38 x 48 Å each, and the subunits can join to form a dimer under physiological conditions with dimensions 41 x 46 x 61 Å.⁴⁹ Figure 8 shows RAMS data for wheat germ agglutinin, prepared as described in Materials and Methods. Part A is RAMS data fitted to dimensions corresponding to the monomer, 20 x 35 x 46 Å. The area under the curves is determined by the SOM ratios for these three measurements. As observed in Figure 7A and Table 2, SOM analysis demonstrates that WGA molecules are found almost exclusively as monomers. Part B is the data fitted to the dimensions corresponding to the dimer, 45 x 45 x 64 Å. The poor fit shown in B suggests that there is not a significant proportion of dimers in the molecules measured in this experiment. These observations suggest that under the conditions of the experiment, in which low concentrations of protein (5-10 µg/ml) are visualized, WGA dissociates into monomers. To further test the possibility of a concentration dependent dissociation of WGA dimer to monomer, WGA (1 mg/ml) was reacted with 2% glutaraldehyde to stabilize the dimer prior to dilution for SFM analysis. As observed in Figure 9, the RAMS distribution can no longer be attributed to pure monomer but has acquired a significant amount of dimer. Part A shows that the observed RAMS data is no longer fit by WGA monomers but is more closely fitted to a 55% monomer and 45% dimer combination, shown in part B. These observations are consistent with other reports of a concentration dependent dissociation of WGA into the monomer⁵⁰ and many reports in the literature describing the dissociation of oligomeric proteins under similar conditions of high protein dilution (1-10 µg protein /ml).^{1,51}

Concanavalin A (Con A). Concanavalin A (Con A) is a lectin, obtained from the jackbean (*Canavalia ensiformis*), that binds to mannose, glucose, and glycoconjugates containing these saccharides. Con A has been extensively used as a probe for glycoconjugates in both animal and plant cells, and can serve as a histochemical probe.^{52,53,54,55} The native protein is composed of four identical subunits each with molecular weight 25,500 Daltons⁵⁶ and molecular dimensions of approximately 33 x 39 x 57 Å.⁵⁷

Figure 10 shows the height distribution profile for Con A that has been fitted to (A) monomer, 28 x 36 x 57 Å, and (B) dimer, 28 x 46 x 68 Å, and Figure 11 shows the same data fitted to (A) tetramer 46 x 59 x 78 Å and (B) a mixture of 13% monomer, 67% dimer and 20% tetramer. Although Con A is normally described as a tetramer at pH 7.0,⁵⁶ it was previously shown by Gordon⁵⁸ that Con A can dissociate into subunits as a consequence of dilution and ionic strength. To determine whether Con A was a dimer under conditions similar to that employed for SFM measurements, we examined the elution profile of Con A (3 µg/ml) on high performance liquid chromatography (HPLC). As observed in Figure 12, the dominant molecular species at 3 µg/ml of Con A was the dimer (Fig. 12A). Further confirmation was obtained by a comparison with the elution profile for FITC-sCon A obtained from a 5 mg/ml solution of FITC-sCon A, known to be a dimer at pH 7.0^{59,60} (Figure 12B). Note the peak in both chromatographs at 8.6 minutes corresponding to Con A dimers. From these HPLC measurements we show that Con A is predominantly a dimer at the concentrations used in the RAMS experiments. Our experiments confirm the observations of Gordon⁵⁸ in which gel chromatography was used to demonstrate that the dimer/tetramer equilibrium of Con A was sensitive to changes in the concentration of Con A within the same ranges of concentration employed in the RAMS measurements.

Protein A. Protein A, a cell wall protein from *Staphylococcus aureus*, is a monomer with four nearly identical domains, A, B, C, and D, each consisting of 60 amino acids, and a fifth domain, the C-terminal, with approximately 150 residues.^{61,62} The ability of the protein to bind the Fc part of IgG from various species has motivated its use as a tool in immunochemical and cell-surface structural studies.⁶³ The crystal structure of fragment B is an oval 13 x 24 x 24 Å,⁶⁴ and the other homologous domains are expected to have similar dimensions. From hydrodynamic studies and sedimentation equilibrium analysis, the molecule was found to have a very extended shape.⁶⁵ It may therefore be concluded that the five domains are arranged in an extended array, giving the molecule dimensions of approximately 13 x 24 x 120 Å. Such an extended shape would be expected to result in a molecular orientation of Protein A on the mica surface in which adsorption predominantly occurs along the long molecular axis. The expected height measured by RAMS would therefore be 13 Å. As observed in the height distribution profile (Figure 13, Table 2), the dominant peak at 12 Å agrees very well with the 13 Å dimension derived from the crystal structure. Part A shows RAMS data fitted to dimensions corresponding to the monomer of Protein A excluding the C-terminal, 12 x 20 x 120 Å. Part B shows RAMS data fitted to the above measurements together with a 24% contribution from heights at 24 Å and 40 Å. These are attributed to measurements on the larger C-terminal domain of Protein A. The SOM fit confirms that the molecule, under the conditions of the experiment, is a monomer. The RAMS data reveals structures at 24 and 40 Å which amount to approximately 24% of the total area under the curve. These heights may be attributed to the C-terminal domain which contains 2.5 times more amino acids and is therefore expected to be larger than the other four domains. The data in Figure 13B has been fit to a combination of two structures, one with dimensions 12 x 20 x 120 Å and the other 24 x 40 x 120 Å.

Streptavidin. Streptavidin, a tetrameric component of the cell wall in *Streptomyces avidinii*,⁶⁶ has four sites of high affinity binding to biotin.⁶⁷ The Streptavidin tetramer has dimensions from the crystal structure of approximately 44 x 45 x 60 Å and is comprised of monomers with the molecular dimensions 18 x 24 x 47 Å.⁶⁸ The biological function of Streptavidin is poorly understood, but may involve an antibiotic role.⁶⁹ Our results again demonstrate the disaggregation of the tetramer to form predominantly a monomer (measured dimensions of 15 x 24 x 51 Å) with minor amounts of other oligomeric species. Figure 14 shows RAMS data for Streptavidin. The data has been fitted to dimensions corresponding to the monomer of Streptavidin, 16 x 27 x 49 Å, with the SOM ratios determining the areas under the curves. Notice that for Streptavidin the smallest dimension is more prevalent compared to the largest dimension than that predicted by SOM, probably because of selective protein/substrate interactions. It seems likely that the dissociation into monomer results from the high dilution of the protein solution analogously to the previously seen dissociation of subunits in Con A and WGA.

To test the possibility of a concentration dependent dissociation of Streptavidin from tetramer to monomer, a solution was made 2 mg/ml streptavidin in 10 mM HEPES pH 7.5, then made 0.3% in gluteraldehyde. It was incubated at room temperature for 45 minutes, then diluted to 0.002 mg/ml streptavidin and 0.0003% gluteraldehyde with 10 mM HEPES pH 7.5. The dilute solution was deposited on MgCl₂ treated mica and incubated at room temperature for 20 minutes, then rinsed twice with 200 µl aliquots of MilliQ water. The protein was analyzed with the SFM and sizes tabulated, and the results are presented in Figure 15. This treatment produces a distribution of 66% monomer, 34% dimer, and 0% tetramer when analyzed with the Gaussian fitting routine in Appendix E, indicating that by fixing the protein at a high concentration, some of the protein molecules are held together in the dimer form, but the native tetramer structure is still not retained. It

is possible, but not tested, that a higher gluteraldehyde concentration would cause complete retention of the tetramer structure.

ADP-glucose pyrophosphorylase (ADPgpp). ADP glucose pyrophosphorylase (ADPgpp) functions as the key regulatory enzyme in the biosynthesis of bacterial glycogen⁷⁰ by catalyzing the formation of ADP-glucose from glucose 1-phosphate and ATP. In *E. coli*, the enzyme is composed of four identical subunits,⁷¹ each with a molecular weight of about 50,000 kDa.⁷² A mutant ADPgpp (G336D) is expressed in *E. coli* K12 strain 618 in which a single amino acid replacement occurs at position 336, an aspartic acid is substituted for glycine.⁷¹ This mutant enzyme: a) stores about 33% more starch than the wild type, b) is more resistant to its inhibitor AMP, and c) is less dependent on the allosteric activator fructose 1, 6-bisphosphate (FbP). Characterization of a mutant enzyme from bacterial strain SG5-504 by sedimentation equilibrium centrifugation, polyacrylamide gel electrophoresis, and gel filtration column chromatography shows that in the presence of FbP, this enzyme aggregates, while native enzyme isolated from a wild type strain (AC70R1) does not exhibit this activator-induced aggregation.⁷² The data suggest that a relationship may exist between altered chemical activity in the ADPgpp mutants and changes in enzyme structure resulting in aggregation. Molecular sizing by non-denaturing gel electrophoresis and Scanning Force Microscopy was performed on wild type ADPgpp, the mutant ADPgpp containing a single amino acid substitution (G336D) and an ADPgpp from the cyanobacteria *Anabaena*. Like the *E. coli* enzyme, the enzyme isolated from *Anabaena* is composed of four identical subunits⁷³ with approximately the same molecular weight as the *E. coli* enzyme. There is a 33% sequence homology between the ADPgpp from the two species, so a similar size and gross structure is expected between these two molecules.⁷⁴

Here we demonstrate the use of RAMS³⁵ microscopy to rapidly determine the molecular size of a family of wild-type and mutant ADPgpp molecules. The data obtained from these measurements demonstrate that: a) a single amino acid replacement in the *E. coli* ADPgpp G336D mutant can induce aggregation of the protein, b) the ADPgpp enzyme from *Anabaena* shows different results indicating two oligomeric forms of the enzyme, the monomer and dimer, and c) RAMS provides a high resolution tool for performing comparative studies of structure/function relationships between wild type and mutant macromolecules. Such analyses can now be performed without recourse to harsh fixation techniques, high vacuums, crystallization, gel filtration, or ultracentrifugation. This analysis shows that the protein sizing technique presented here can be useful for a protein when the structure of the protein is not well understood.

As shown in the electrophoresis gel in Figure 16, G336D has a larger molecular weight compared to the wild type enzyme. The enzyme from *Anabaena* shows two bands, one at the lower molecular weight end of the broad *E. coli* enzyme band, and one much lower than the *E. coli* enzyme band. The larger molecular weight for the G336D enzyme must result from aggregation of molecules to form a higher molecular weight form. These findings are corroborated with the RAMS data.

For RAMS analysis the three enzymes were prepared on mica substrates as described in the Materials and Methods section, and the results are presented in Figure 17. To provide a more detailed analysis of the data and a more graphic means to compare size data between different proteins, the data are shown in a format similar to an electrophoresis gel, where a darker band indicates a larger quantity of protein at the size corresponding to the position on the graph. The size is plotted on a log scale for accurate comparison to the electrophoresis gel. This plot clearly shows the type and amount of each molecular species for each type of ADPgpp. It is observed that the single amino acid substitution in the ADPgpp G336D mutant results in a considerable increase in the measured size when

compared to the wild type enzyme. The large increase observed reflects aggregation of the mutant enzyme caused by the mutation imparting greater stability of the aggregate form.

Since this protein does not have a crystal structure solved, there is more potential for error in interpreting the RAMS data, however, we have obtained enough data to make reasonable predictions for molecular size for the monomer, tetramer and octamer of this enzyme. In Appendix A, two dimensional arrays of Streptavidin and ADPgpp protein molecules are described, formed when the protein concentration is higher than that required to produce isolated molecules on the surface. The array is 60 Å tall in the case of Streptavidin, and 100 Å tall in the case of ADPgpp. If an analogy is made between ADPgpp and Streptavidin, we could conclude that the largest dimension of the tetramer of ADPgpp is about 90 Å, because the two-dimensional arrays of ADPgpp are 100 Å tall and the Streptavidin two dimensional arrays suggest that these molecules align themselves with the longest axis as the height, and a slightly larger height of the two-dimensional array than the longest axis of the tetramer. The other two dimensions of the protein are left to be determined by the single particle RAMS measurements and by SOM. The molecular weight of the tetramer is about 200,000 Daltons,⁷² and the volume of the molecule can be estimated at 332 nm³ if one assumes the protein has a density equal to that of water. This would make dimensions of the tetramer of approximately 50 x 70 x 90 Å a possible set of three dimensions of the molecule that would be consistent with RAMS results seen for the wild type enzyme, shown in Figure 18. When the tetramer has this set of dimensions, the octamer could be formed by joining two tetramers with the largest face together forming a molecule with dimensions 75 by 90 by 95 Å. These dimensions agree well with the RAMS data for the G336D *E. coli* mutant enzyme (Figure 19), suggesting that the point mutation converting glycine to aspartate at position 336 causes aggregation of the enzyme from the normal tetramer to an octamer. For this *E. coli* enzyme and the enzyme from *Anabaena*, both of which in the native state are tetramers composed of identical subunits, the monomer

could have dimensions of about 25 x 35 x 90 Å. From the electrophoresis and the RAMS data, one can see that the enzyme from *Anabaena* is composed of a significant proportion of monomers and an additional proportion of a higher order oligomer. The fit of the RAMS data for the enzyme from *Anabaena* is shown in Figure 20. Figure 20A shows the data fitted to dimensions of the monomer alone, 25 x 35 x 90 Å, and figure 20B shows the RAMS data fitted to a combination of 68% monomer and 31% tetramer (dimensions 50 x 70 x 90 Å). The fit using these dimensions misses a peak in the data at about 63 Å. This could be caused by small structural differences between the *E. coli* enzyme, from which the tetramer dimensions were deduced, and the *Anabaena* enzyme. The two enzymes should be quite similar in structure since the sequences are highly homologous. Finally, it can be reasonably concluded based on RAMS and on electrophoresis that the *Anabaena* enzyme exists as a combination of monomer and tetramer.

DISCUSSION

Analytical Methods for Measuring Molecular Dimensions and Oligomerization of Macromolecules. Investigations with receptor proteins in membranes, allosteric enzymes, and DNA binding proteins have all demonstrated a structural promiscuity for proteins in their ability to undergo functionally important conformational changes and oligomerization in response to changes in protein concentration, protein-protein interactions, ligand/effector binding and post-translational modification.⁴ To understand and measure such phenomenon has proven difficult, since the concentration range in which purified proteins are studied is usually far above that found in the intracellular environment. In many instances substrate or ligand induced changes in protein conformation that are functionally significant are masked as a consequence of the necessity for performing measurements at high concentrations of

protein. Recent measurements with a variety of oligomeric enzymes clearly show that their activities can be regulated by controlling their ratio of oligomers e.g. dimer to tetramer.^{1,2} In many cases membrane receptor activation requires ligand induced aggregation.⁷⁵ Enhanced binding/avidity and activity for a large number of diverse DNA binding proteins e.g. p53, Rec A, Par A, CAP has been suggested to occur through substrate induced oligomerization of DNA bound protein.^{4,5,7,76,77} Such observations suggest that new methods are necessary to determine the molecular dimensions of proteins at low abundance and under conditions of protein concentration that are normally encountered within the cell. These methods should more accurately reflect the changes in conformation and oligomerization that are suggested to be responsible for regulating biological activity.

Unfortunately, a limited number of analytical techniques are presently available with the requisite sensitivity and resolution to measure the molecular dimensions and organization of individual protein molecules under conditions that can maintain biological activity of proteins at low protein concentration. New approaches with NMR have provided the opportunity to measure the conformation of proteins in solution and to determine the arrangement of specific amino acids within the polypeptide chain. Oligomerization and ligand binding studies have also been performed using these methods.⁷⁸ Although these techniques are useful, it is presently too complex to interpret NMR data for molecular investigations of proteins larger than 40 kDa. A further limitation is the need to use relatively high concentrations of protein and expensive NMR instrumentation for these investigations. Electron microscopy (scanning (SEM) and transmission (TEM)) may be successfully employed to measure the molecular organization of large proteins and macromolecular structures.^{79,80} A significant disadvantage of such procedures, however, is that for most studies the preparation techniques involve dehydration and metal coating, both of which can damage protein structure and certainly destroy protein activity. In addition, electron microscopy is not particularly sensitive to

smaller proteins (<100 kDa). Polyacrylamide gel electrophoresis (denaturing and non-denaturing), gel and liquid chromatography have both been useful in providing measurements of molecular weight and protein oligomerization, but do not have the resolution to provide molecular dimensions and generally require significant concentrations of protein. Although sensitive to protein aggregation and protein shape, equilibrium sedimentation ultracentrifugation can not provide molecular dimensions and requires high concentrations of protein.

RAMS- A New Method for Measurements of Molecular Dimension and Protein Oligomerization with Angstrom Resolution. The RAMS technique provides a means to measure the molecular dimensions for the complete size range of monomeric and oligomeric proteins. This approach extends the nanometer resolution of SFM to that of angstroms by avoiding excessive image broadening of biological samples introduced by the lateral distortions introduced by geometry of the probe tip.^{15,16,17,18} Tip-protein interaction, mainly lateral and vertical pressure, can cause the molecule to appear shorter.¹⁹ With regard to protein-substrate interaction, there is an attractive interaction between the polar surface of the protein molecule and the charged mica surface, possibly deforming the molecule and causing it to flatten against the mica surface, however, study on lysozyme has shown that it retains its activity when adsorbed onto mica and imaged with SFM, so these deformations must not be very extensive.⁸¹ From our measurements we have observed 10% or less compression for the sizes 50Å or larger, and very little deformation in the smaller molecules.

By using a simple mathematical model, the statistical orientation model (SOM), to interpret the height distribution profiles, it is possible to convert the observed distribution of heights measured in a large population of randomly oriented molecules into 3 dimensional coordinates. A particularly important aspect of this model is that it provides a

means to determine the contribution of each molecular species, e.g. monomer, dimer, tetramer, to the height distribution obtained by RAMS. An examination of the preferred orientation of the molecules also provides the opportunity to predict the organization of individual subunits within the protein oligomer. Measurements with angstrom resolution may be performed on functional complexes. The sensitivity and resolution of the RAMS technique make it a powerful and unique tool for the determination of protein structure, protein oligomerization as a regulator of enzyme activity, DNA/RNA organization, gene regulation, chaperonin function, and the effects of post-translational modification on protein organization and activity.

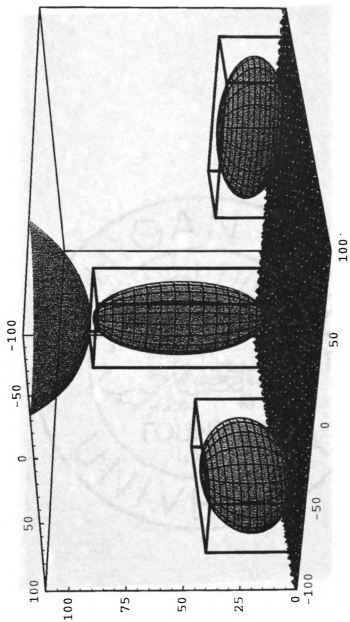


Figure 3: Three orientations of a molecule the shape of Con A arranged on a mica lattice.

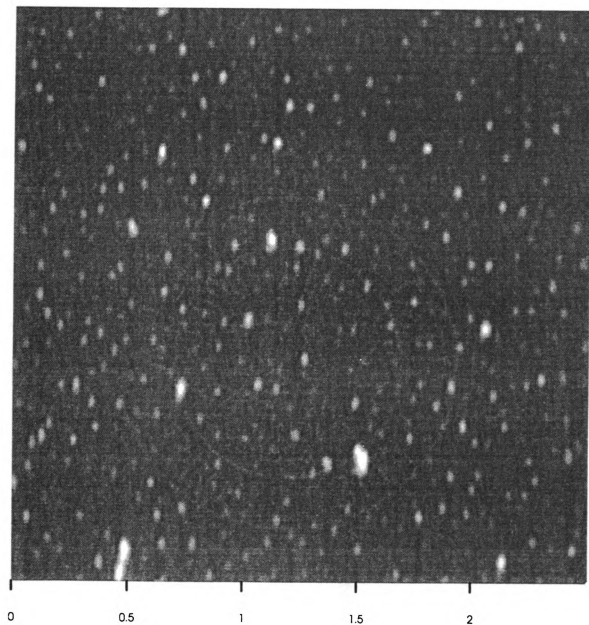


Figure 4: 2.5 by 2.5 μm image of WGA on mica.

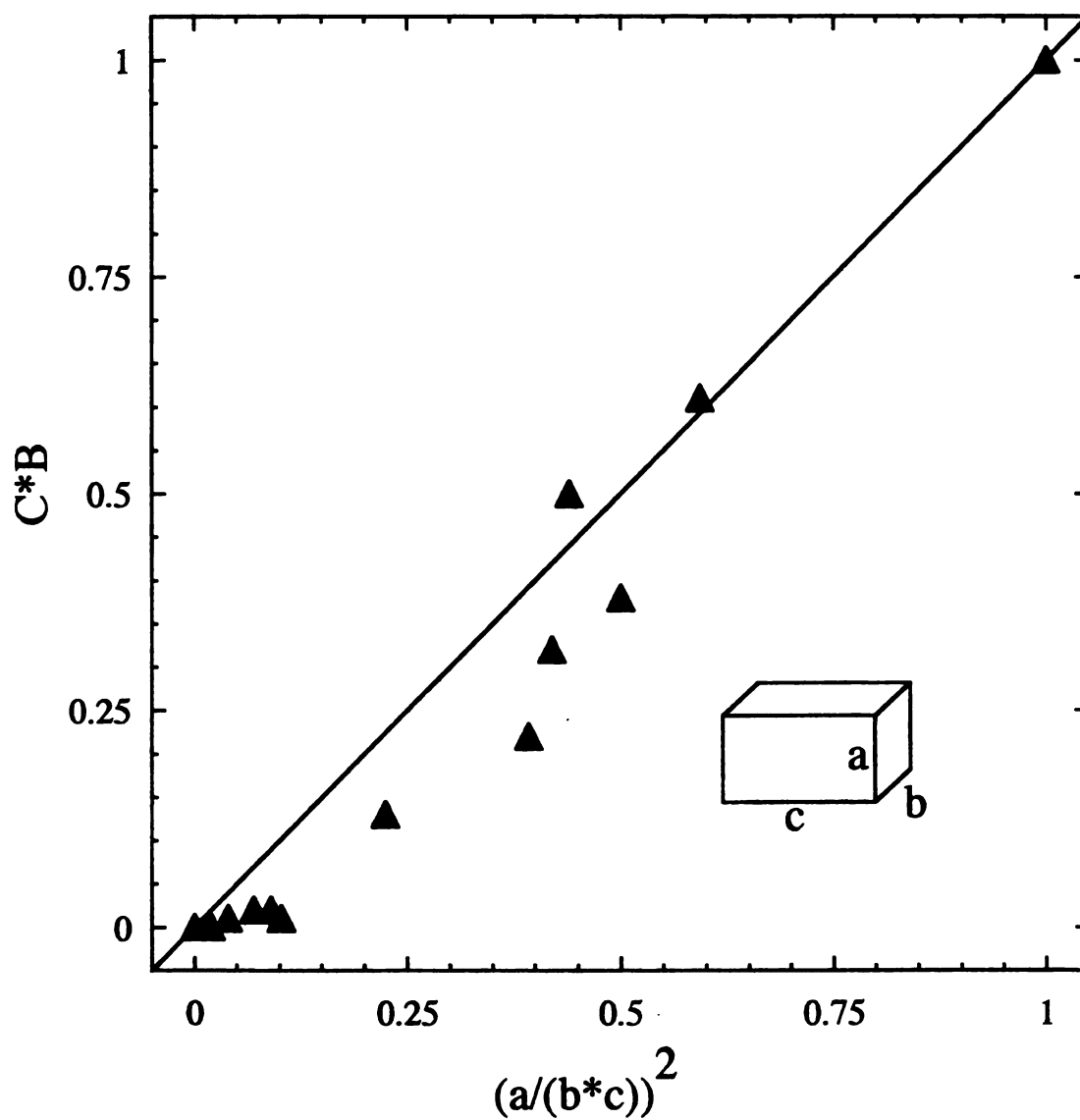


Figure 5: Correlation between elongation and the probability that the two tallest dimensions are vertical

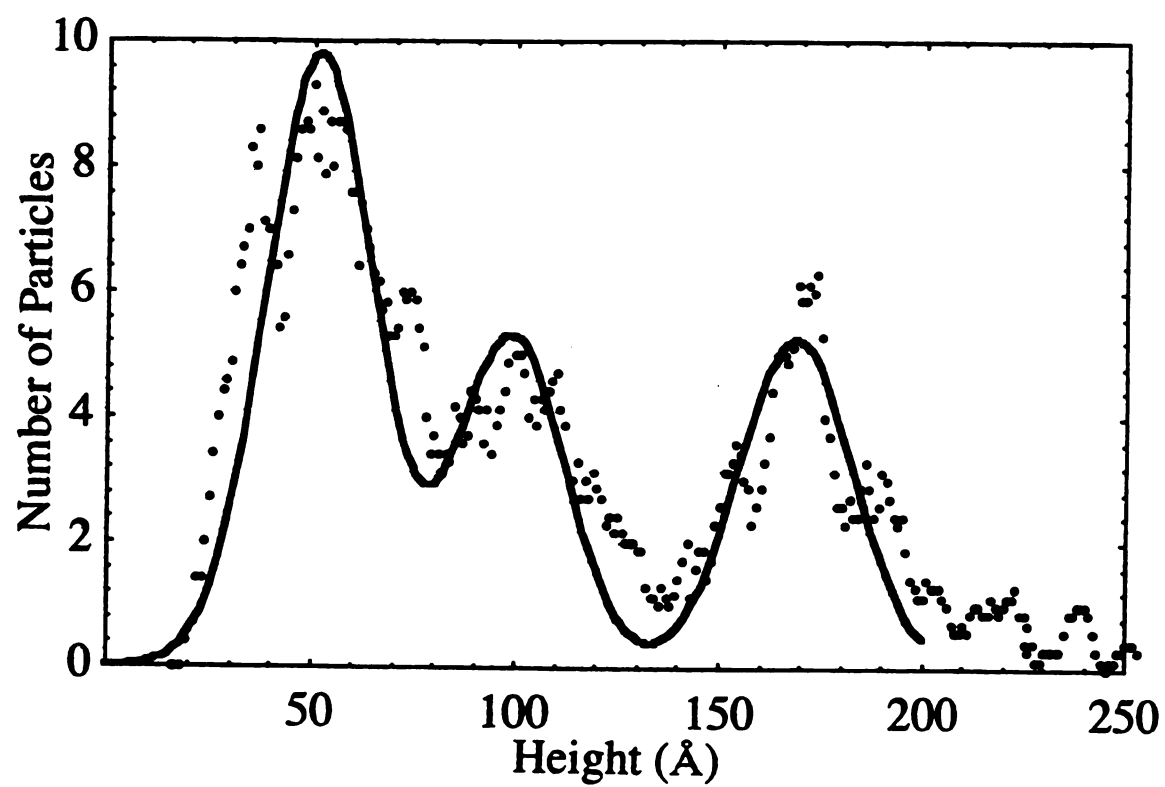


Figure 6: RAMS on 4.9, 9.0, and 18.1 nm colloidal gold fitted to Gaussians.

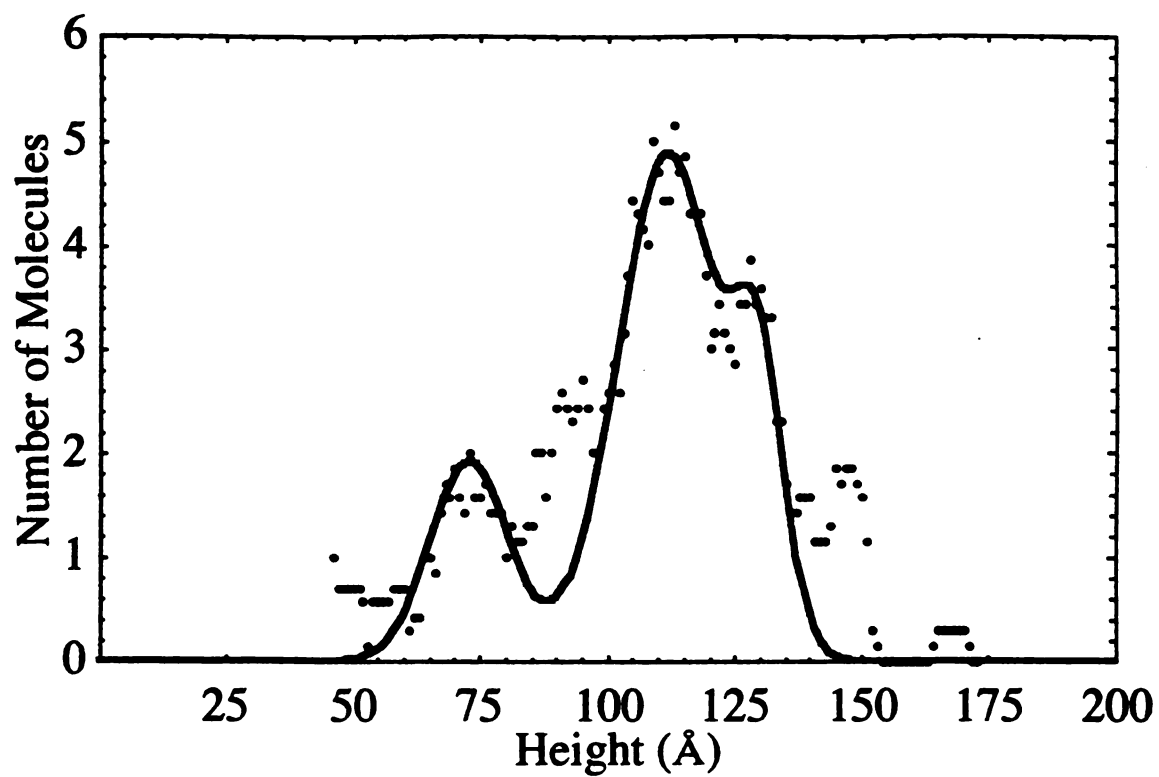


Figure 7: RAMS on Ferritin.

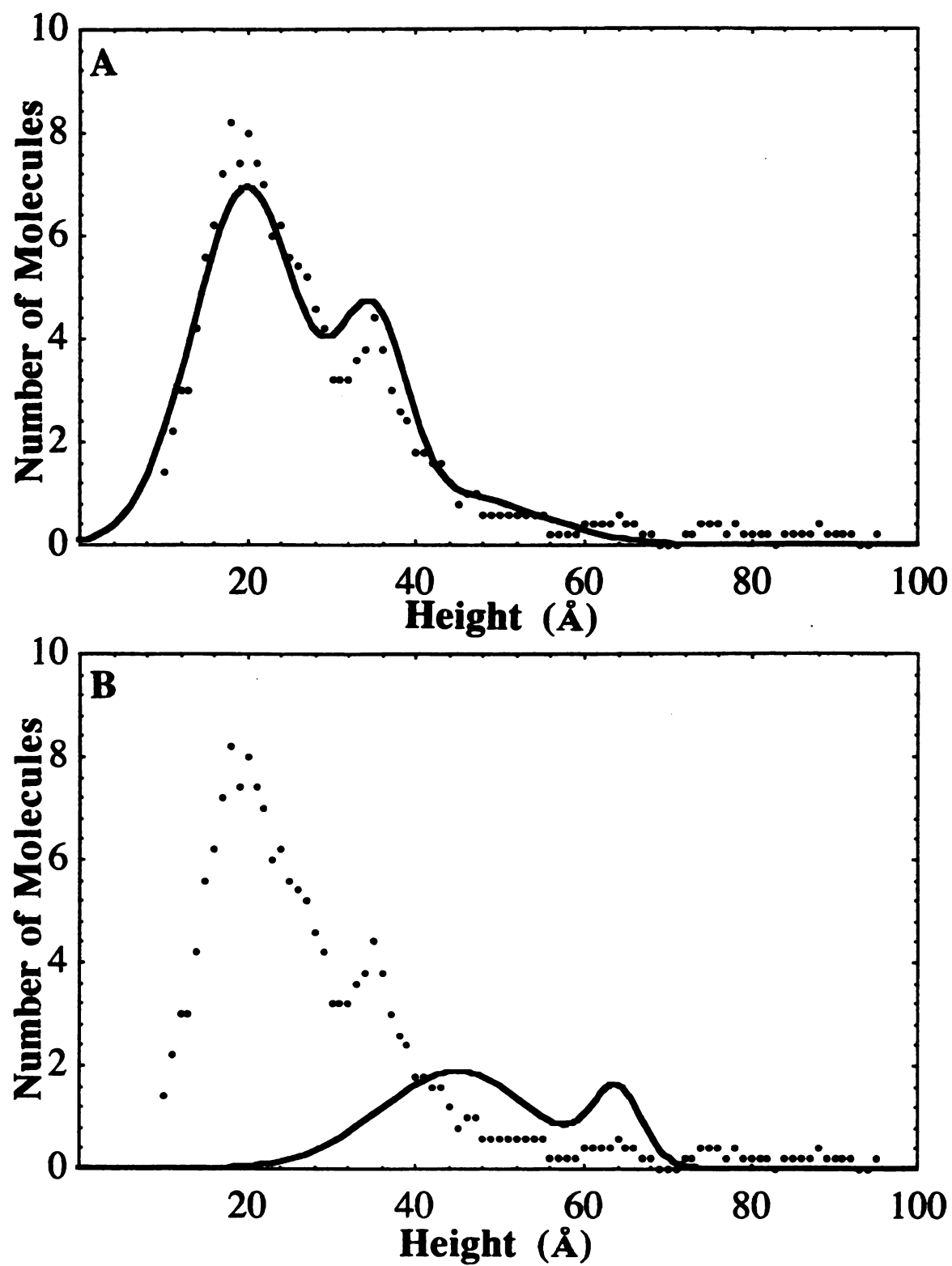


Figure 8: RAMS on WGA fitted to (A) monomer and (B) dimer dimensions.

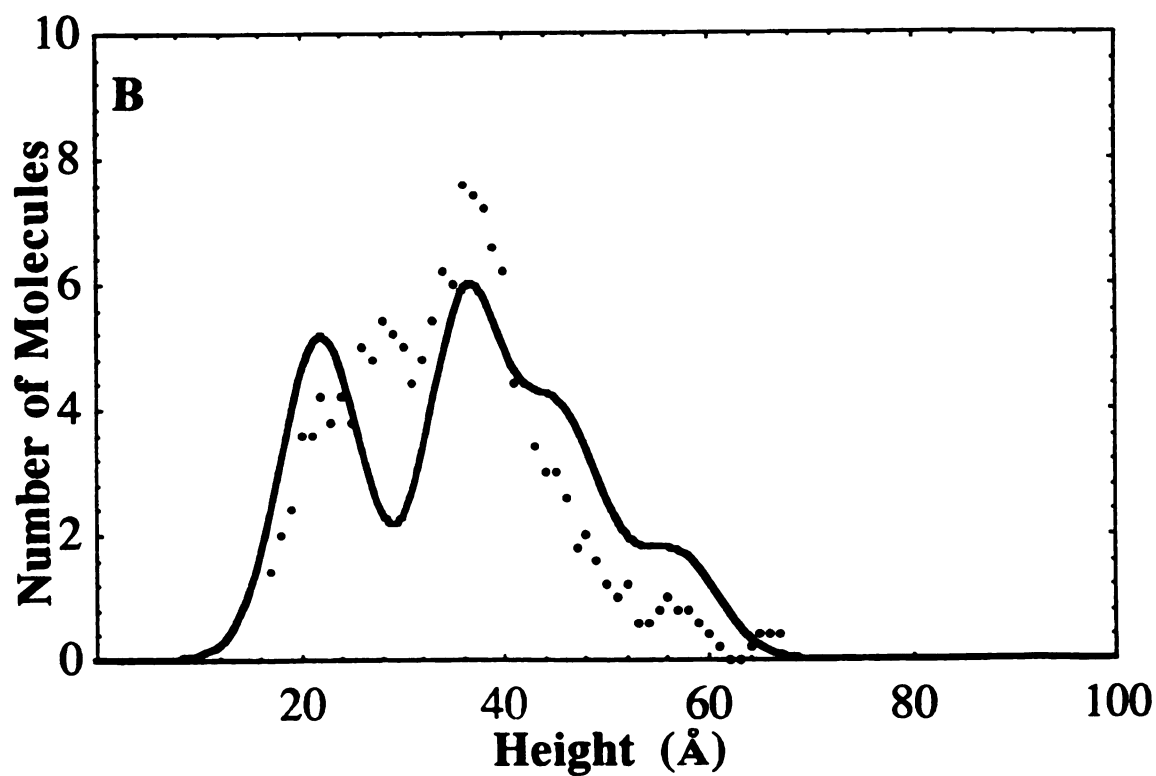
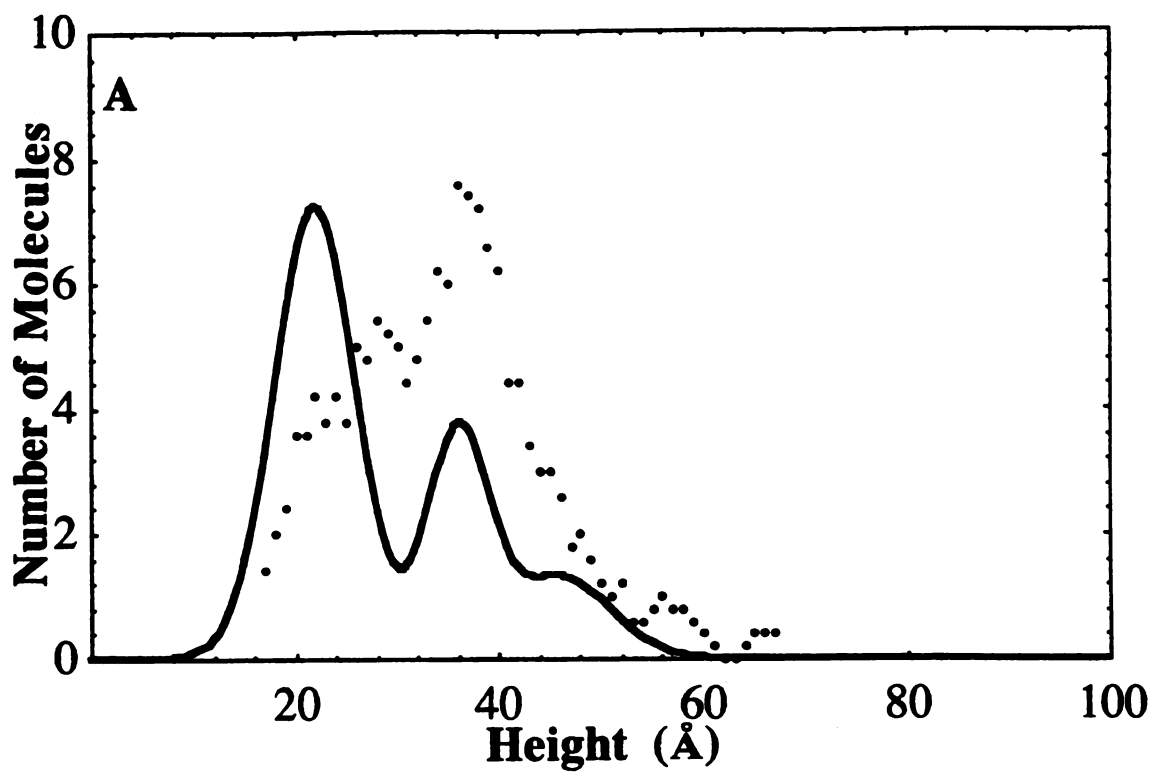


Figure 9: RAMS on WGA fitted to (A) monomer and (B) 55% monomer and 45% dimer.

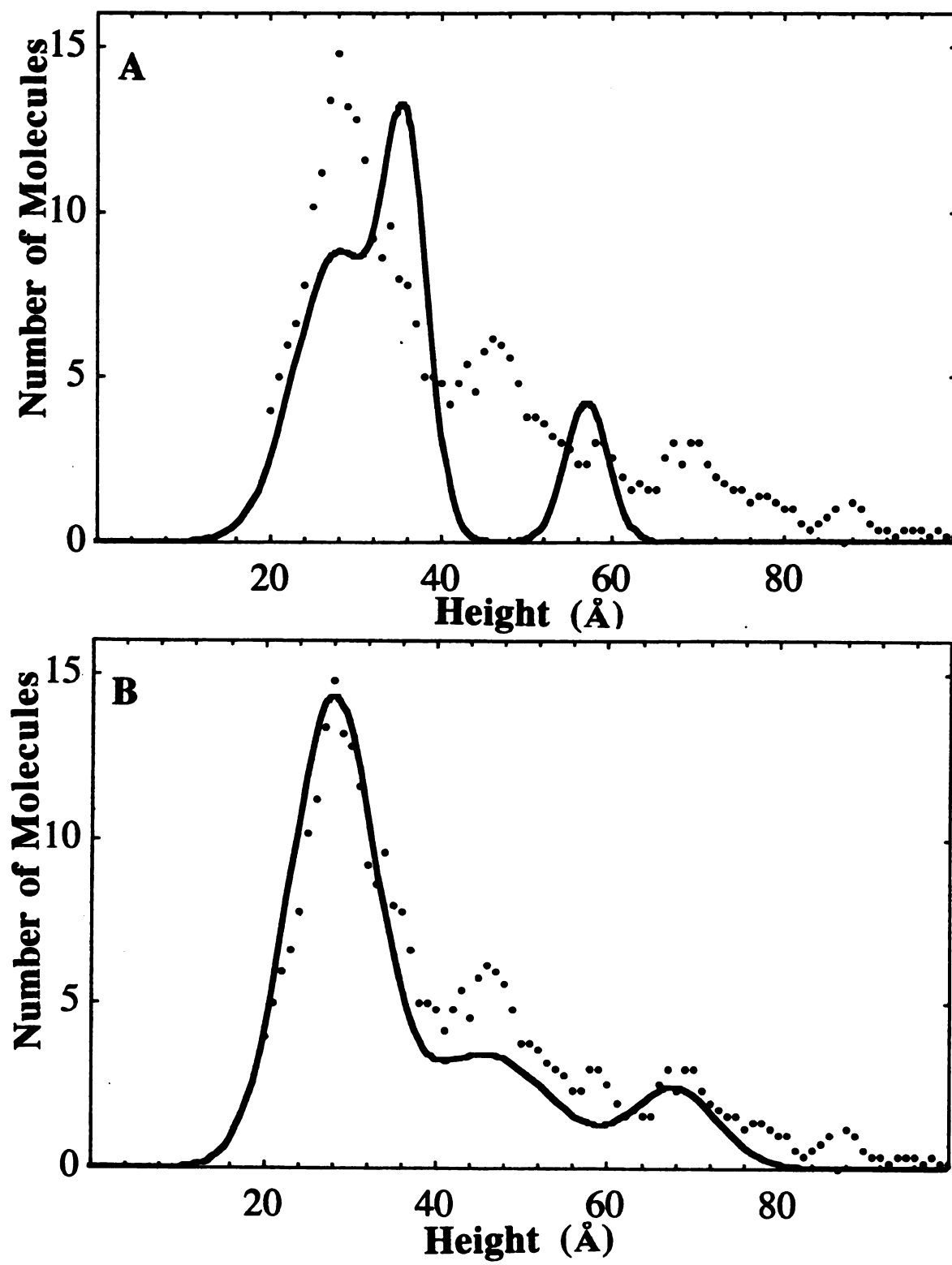


Figure 10: RAMS on Con A fitted to (A) monomer (B) dimer

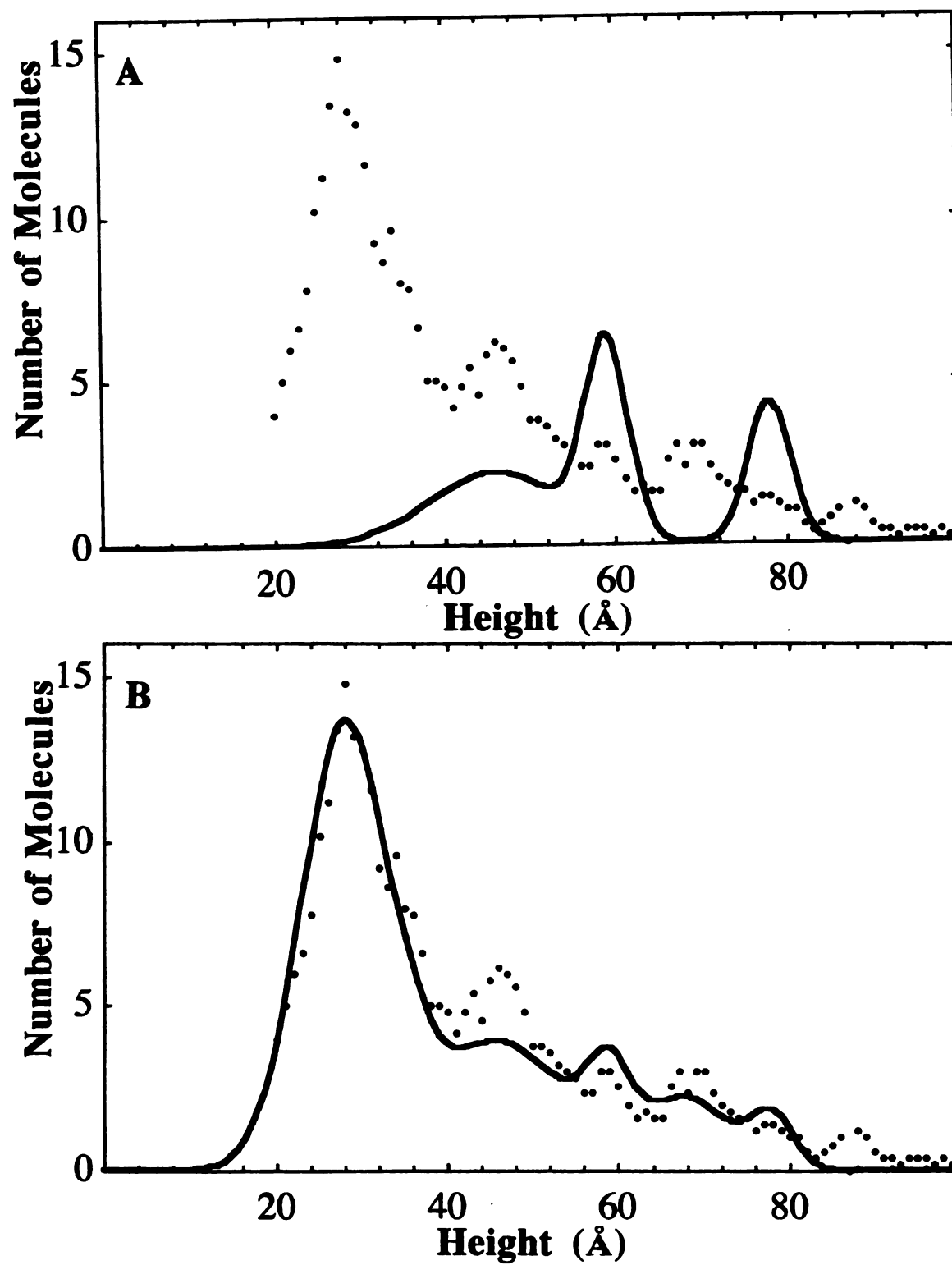


Figure 11: RAMS on Con A fitted to (A) tetramer (B) 13% monomer, 67% dimer, 20% tetramer.

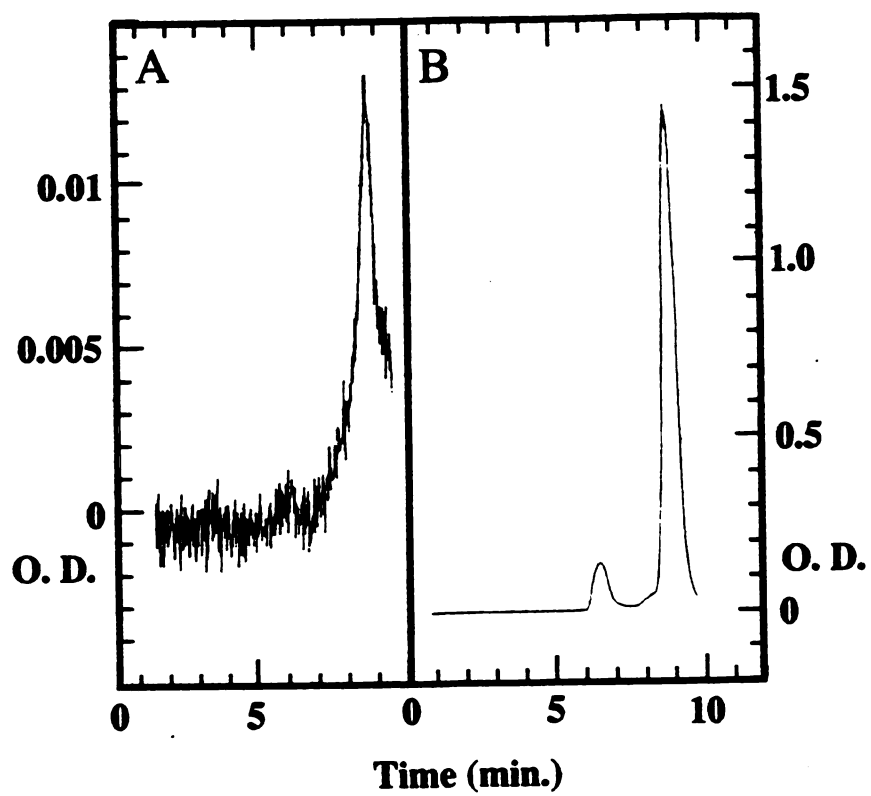


Figure 12: HPLC on (A) 3 µg/ml Con A (B) 5 mg/ml FITC-sCon A.

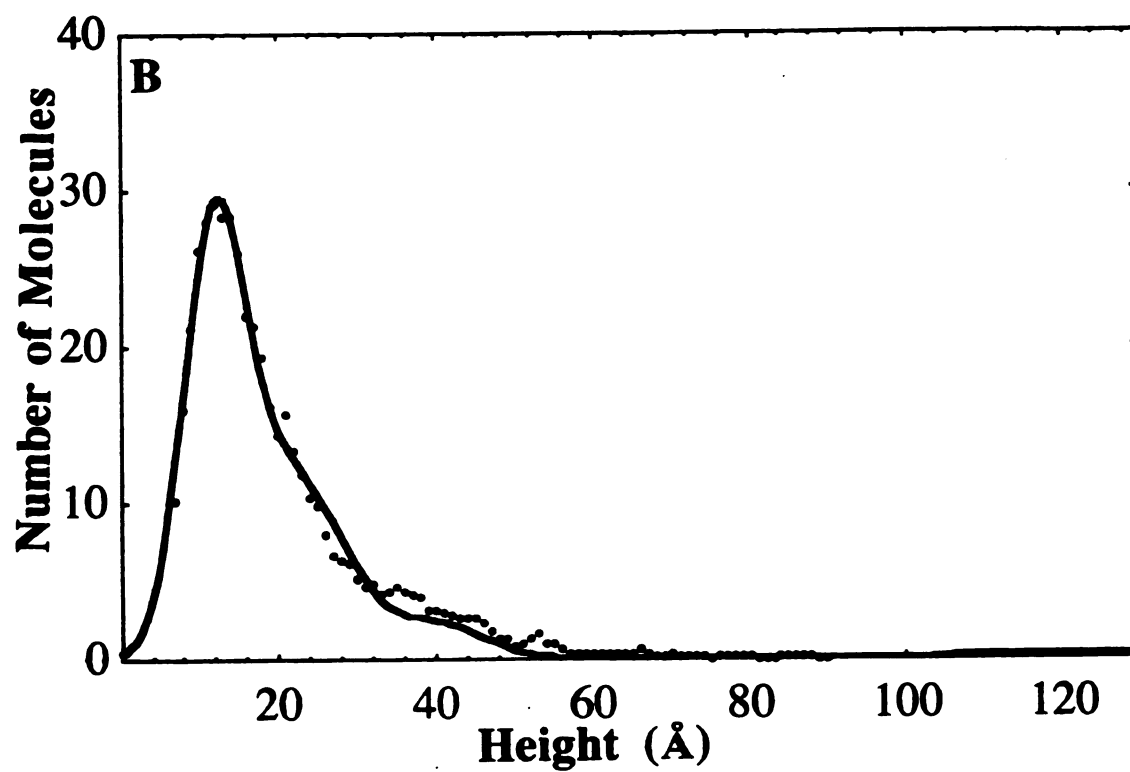
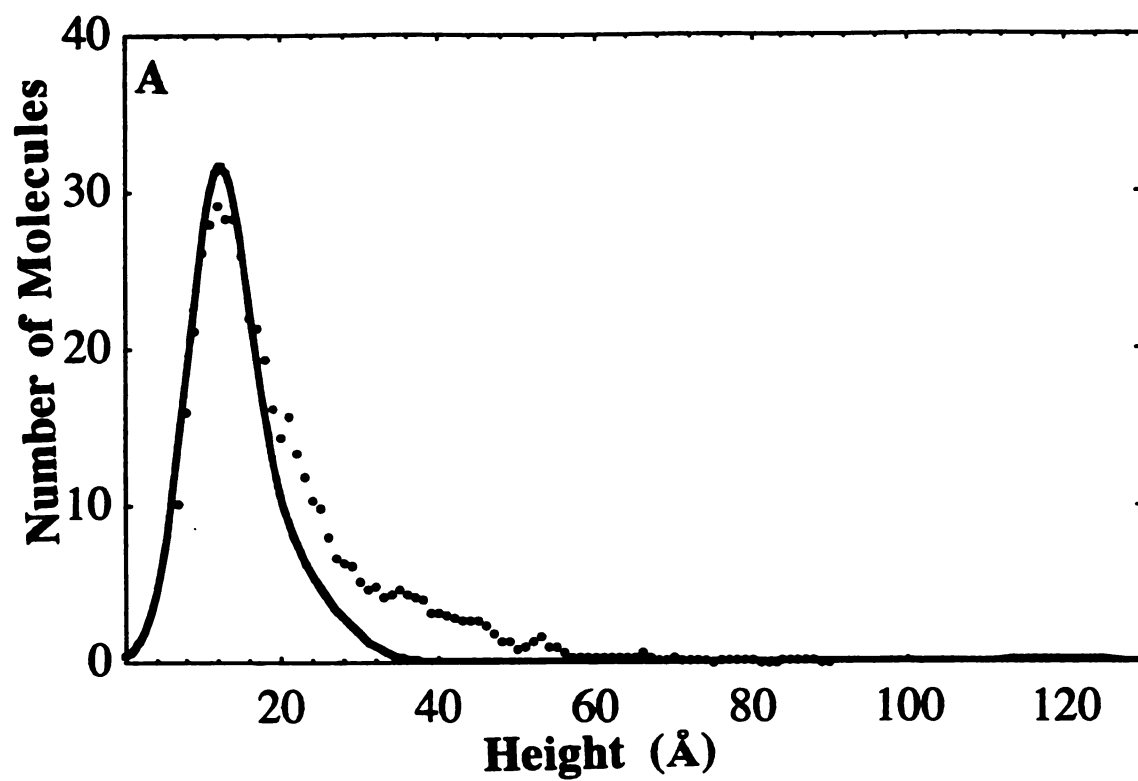


Figure 13: RAMS on Protein A fitted to (A) monomer (B) part A plus 24% contribution from 24 and 40 Å.

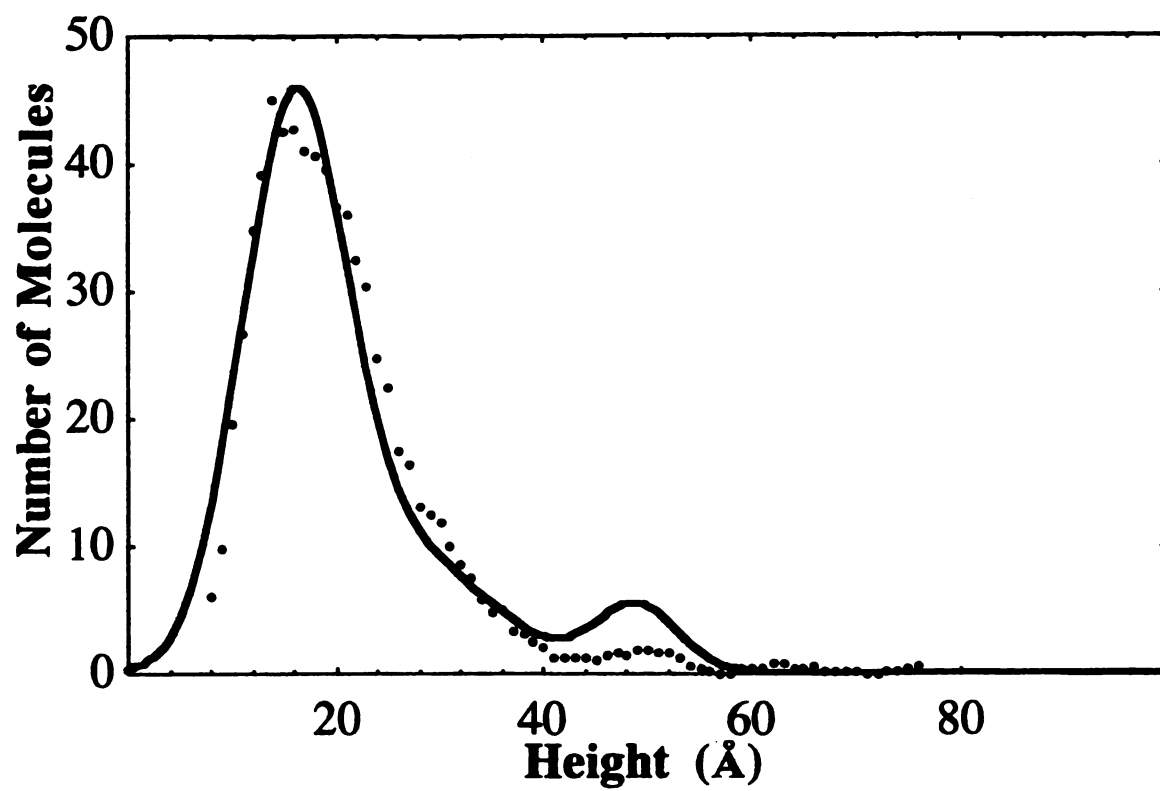


Figure 14: RAMS on Streptavidin fitted to monomer.

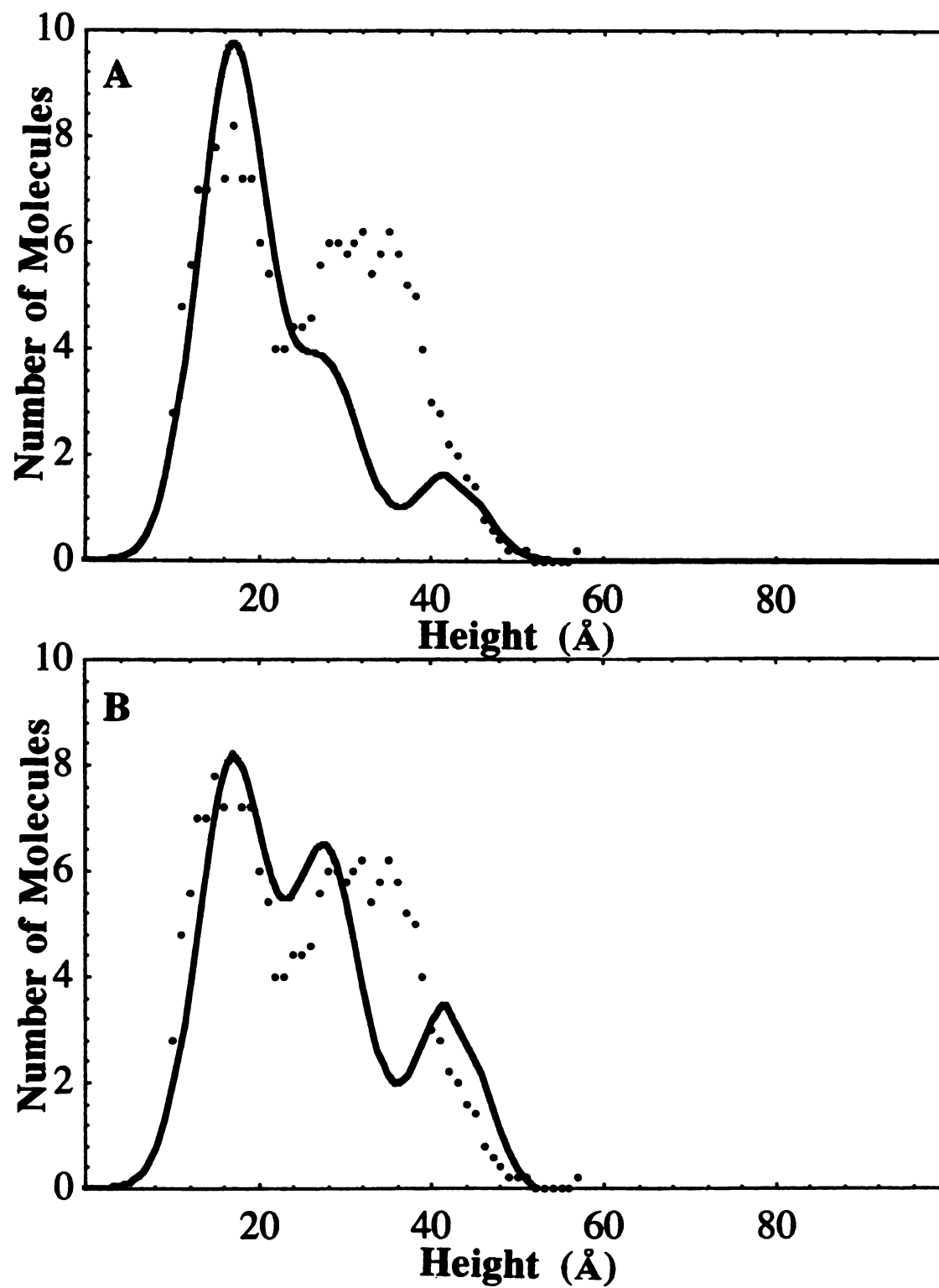


Figure 15: RAMS on glutaraldehyde treated Streptavidin fitted to (A) monomer (B) 66% monomer 34% dimer.

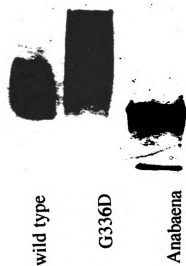


Figure 16: Non-denaturing gel on ADPgpp wild type, G336D, and Anabaena.

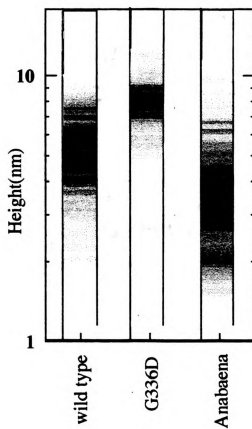


Figure 17: SFM data on ADPgpp wild type, G336D, and Anabaena.

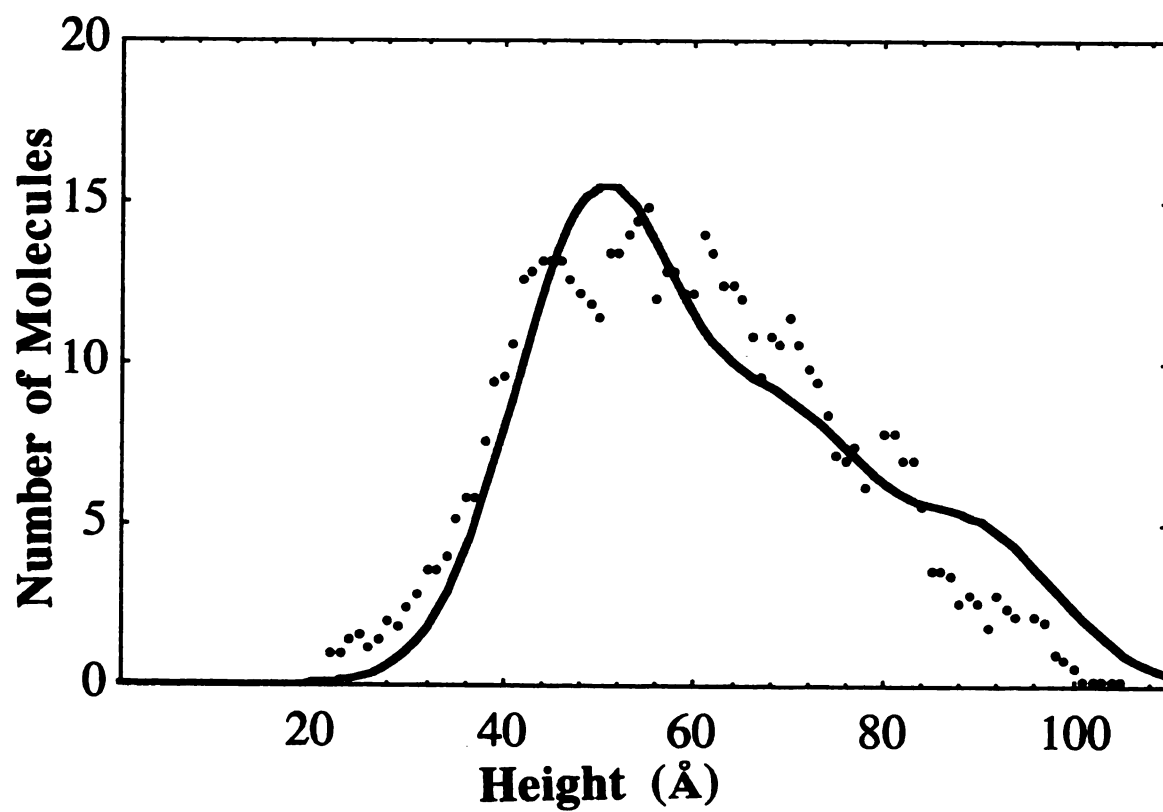


Figure 18: RAMS data on ADPgpp wild type.

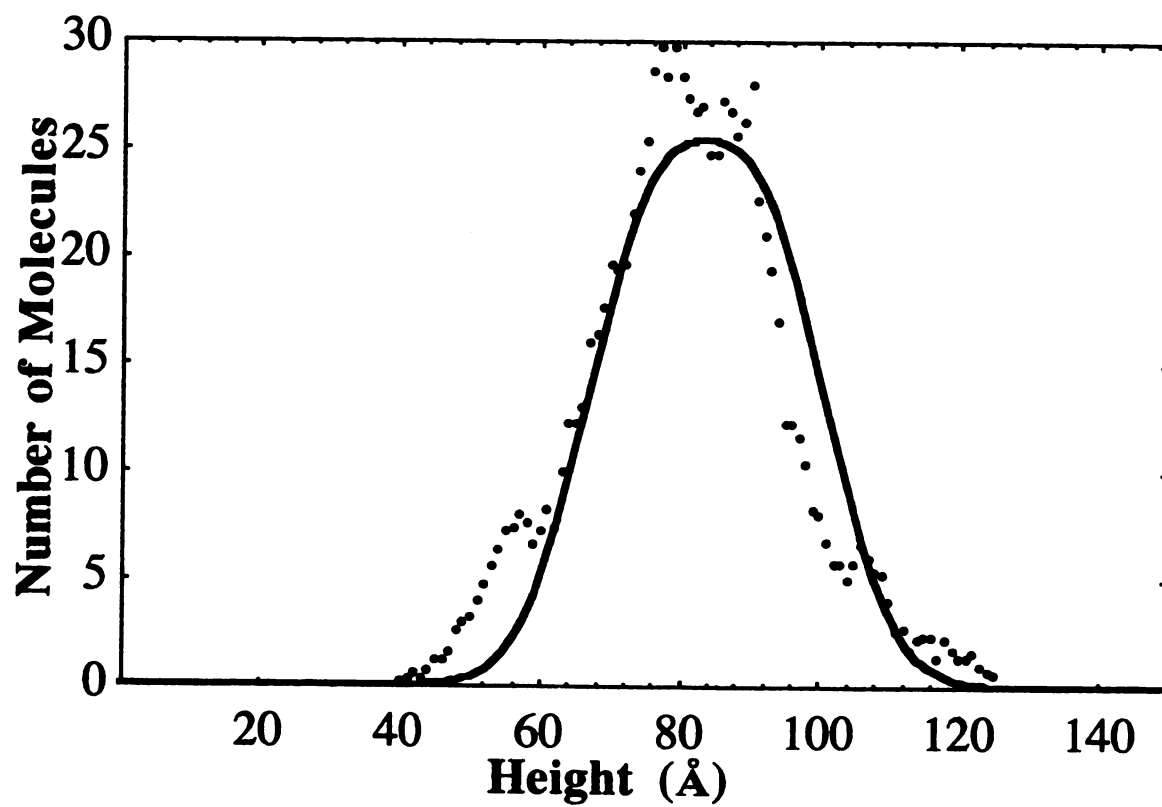


Figure 19: RAMS data on ADPgpp G336D mutant.

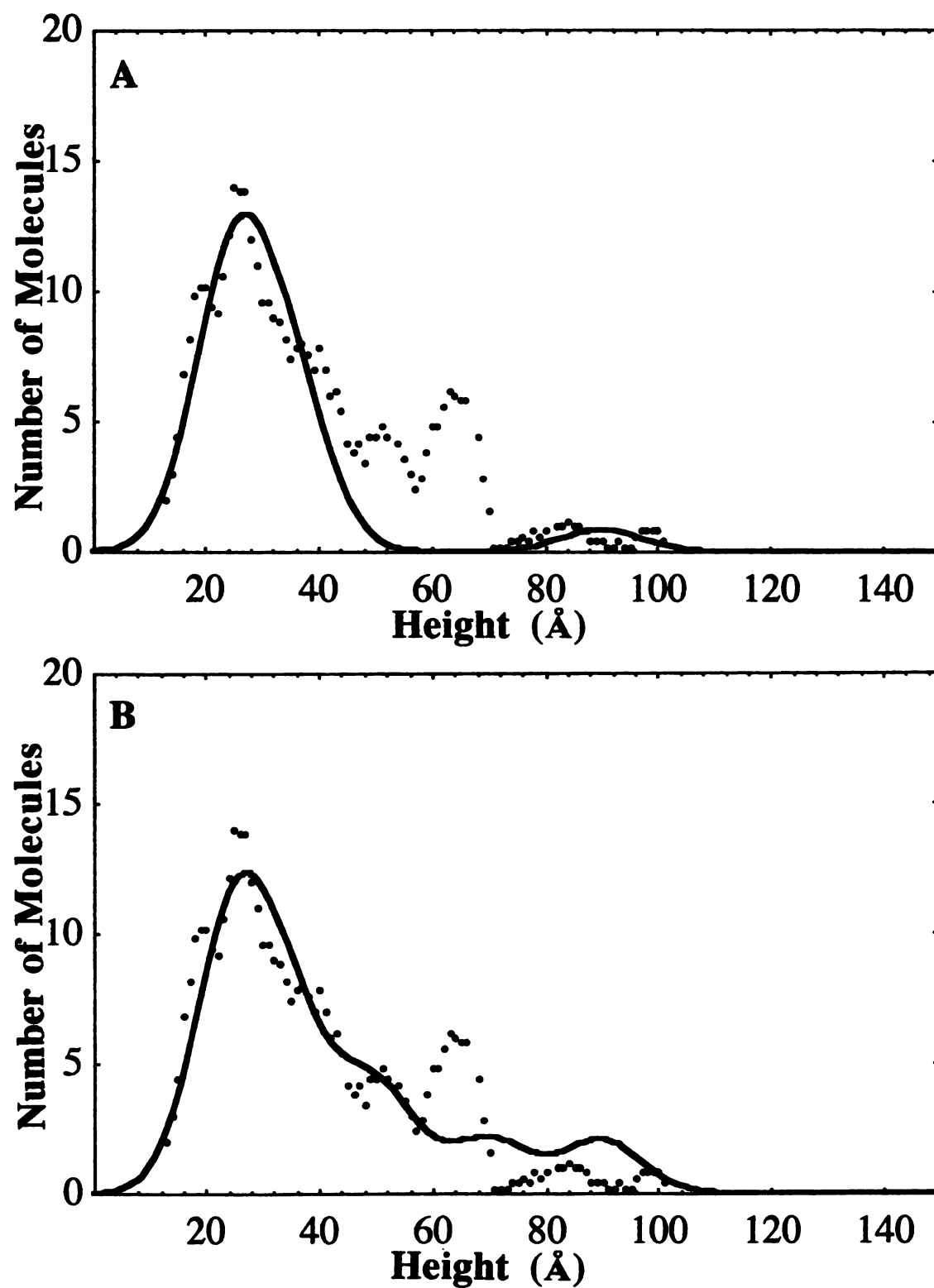


Figure 20: RAMS data on ADPgpp from *Anabaena* fitted to (A) monomer (B) 68% monomer 31% tetramer.

Table 2: Summary of the protein dimensions as derived from RAMS microscopy and x-ray crystallography.

	Molecule	Dimensions (Å) ^a	Predicted SOM ^b percentage	Measured dimensions (Å)	Observed SPM percentage ^c	Oligomerization ratio ^d
Con A monomer		33	49	28	57	0.13
Con A dimer		39 39 57	35 16	36 59	34 9	
Con A tetramer		28 41 79	63 29 8	28 46 69	52 33 14	0.67
		41 56 79	55 30 15	46 59 79	69 16 15	0.20
WGA monomer		22	65	20	69	1
WGA dimer		38 48	21 14	35 45	21 10	0
		41 46 61	45 35 20			
Streptavidin monomer		18	59	16	71	1
Streptavidin dimer		24 47	33 8	27 49	27 2	0
Streptavidin tetramer		28 44 45	56 22 21			0
		44 45 56	39 37 24			
Protein A monomer		13 24 120	77 22 1	12 20 120	67 24 9	
Ferritin		120	33	73 112 130	20 41 39	

^a These are the dimensions measured from the crystal structure deposited with the Brookhaven Protein DataBank, as referenced in the text. The molecule was visualized with Insight II, Biosym.

^b Predicted percentage of the molecules oriented such that the corresponding value is the height of the molecule. This is calculated as described in the text.

^c Area under the peak presented as a percentage of the total area considered for each protein and each oligomerization state, for easier comparison to the SOM ratios. For Concanavalin A, the area under the peak at 46 Å had contributions of both dimer and tetramer, and the area under the peak at 28 Å had contributions from both monomer and dimer, and the area under the peak at 59 Å had contributions from both monomer and tetramer.

^d Ratio of monomer, dimer and tetramer as deduced from the gaussian fits to the RAMS data.

CHAPTER 3 : LIGAND BINDING STUDIES

INTRODUCTION

Recent measurements with a large number of diverse DNA binding proteins including Rec A, Par A, and CAP, have suggested that enhanced binding and activity occur through substrate-induced oligomerization of DNA bound proteins.^{82,83,84} A mutant ADP-glucose pyrophosphorylase has demonstrated activator-induced oligomerization.⁷² Thus it is critical to the understanding of protein structure to be cognizant of substrate-induced conformational changes. With the tools and theory presented in the previous chapter, studies can be done to determine ligand-induced oligomerization of proteins. RAMS³⁵ has proven to be a technique readily applicable to this type of study as will be demonstrated in experiments on Wheat Germ Agglutinin and Streptavidin.

Wheat Germ Agglutinin (WGA) binds N-acetylglucosamine (GLcNac) and derivative sugars (i.e., the β -(1-4)-linked oligomers such as di-N-acetylchitobiose or tri-N-acetylchitobiose)^{46,85,86} known to be present on membrane surfaces of cells.^{87,88,89} The binding of WGA to oligosaccharides of GLcNac displays even higher affinities, with a dissociation constant of 4.0 and 1.2×10^{-5} M for the di- and tri-saccharide, respectively, compared with a dissociation constant of 7.6×10^{-4} M for the mono-saccharide.⁴⁷ Tryptophan residues at the binding site seem to be essential for binding activity, as shown by examining binding following treatment with a tryptophan modifying agent.⁹⁰ RAMS analysis indicated that WGA was a monomer under the high dilution and low ionic strength conditions used in experiments described in chapter 2. It was suspected that the subunit association may be strengthened by substrate binding, so a study was undertaken to determine if WGA would be stabilized in the dimer form upon binding GLcNac, di-GLcNac, and tri-GLcNac. Since the high dilution conditions were suspected to cause dissociation of the subunits, the protein was treated with the sugar before dilution. Also,

experiments were performed treating the protein with the sugar after dilution to measure the ability of the sugar to promote re-association of the subunits.

Streptavidin, a protein produced by the actinobacterium *Streptomyces avidinii*, is remarkable for its extraordinarily strong affinity for d-biotin. The dissociation constant for the streptavidin-biotin link is 10^{-15} M.⁹¹ This protein has a tetrameric structure, and each identical subunit has a molecular weight of 13.5 kDa⁹¹ and dimensions from the crystal structure 18 Å by 24 Å by 47 Å. The tetramer has dimensions 42 Å by 42 Å by 56 Å.⁶⁸ Despite the lack of sulfur containing residues and disulfide bonds,⁹² the tetramer is relatively stable with regard to dissociation, and this stability is increased upon binding biotin.⁹³ One notable structural feature is that contacts made by tryptophan (Trp)- 120 of one subunit with biotin bind by an adjacent subunit through the dimer-dimer interface, where two stable symmetric dimers are associated to form a tetramer having dihedral D₂ symmetry.⁹² In an experiment by Cantor's group in Boston University the wild type protein bound to biotin did not dissociate from tetramer to monomers in an SDS gel unless the protein was heated prior to running the gel. A mutant was constructed which had a biotin-Streptavidin dissociation constant of about 10^{-8} M compared to the biotin-Streptavidin dissociation constant for wild type of 10^{-15} M.⁶⁶ This mutant Streptavidin, which bound biotin less strongly because of conversion of tryptophan 120 to phenylalanine, dissociated into monomers in an SDS gel whether or not the protein was heated prior to running the gel,⁹⁴ indicating that biotin binding plays a pivotal role in increasing the stability of the quaternary structure.

MATERIALS AND METHODS

Materials. Materials used were described in the previous chapter, with a few additions. Biotin and N-Acetylglucosamine (GLcNAc) were obtained from Sigma. Di-

GLcNac and tri-GLcNac were a gift from Professor Melvin Schindler, Department of Biochemistry, MSU. RAMS data was collected and analyzed as described in the previous chapter, and the Mathematica routine in Appendix E was used to evaluate the proportion of the different oligomers in the protein size distribution.

WGA-GLcNac sample preparation. WGA was treated with the sugar while still in the concentrated form, in an effort to avoid dilution induced dissociation of the subunits of the dimer WGA. The monosaccharide treatment was performed in the following manner: A solution was made 4.8 mg/ml in wheat germ agglutinin (WGA) and 0.16 M N-acetylglucosamine (GLcNac) in 10 mM HEPES pH 7.5. After about 30 minutes of incubation at room temperature, the solution was diluted with 10 mM HEPES pH 7.5, making a solution 2.4 μ g/ml in WGA and 0.00008 M in GLcNac. The dilute solution was deposited on MgCl_2 treated mica and incubated at room temperature for 20 minutes, then rinsed twice with 200 μ l aliquots of MilliQ water. The disaccharide treatment was performed as follows: A solution was made 1 M di-GLcNac and 6 mg/ml WGA. After incubating at room temperature for several hours, the above solution was diluted to 0.01 M GLcNac and 0.06 mg/ml WGA, and after refrigeration for 7 days diluted again to 0.0003 M GLcNac and 1.8 μ g/ml WGA. The dilute solution was deposited on MgCl_2 treated mica and incubated at room temperature for 20 minutes, then rinsed twice with 200 μ l aliquots of MilliQ water. The tri-GLcNac treatment was as follows: A solution was made 0.2 M tri-GLcNac and 6 mg/ml WGA. After incubating at room temperature for several hours, the above solution was diluted to 0.02 M GLcNac and 0.06 mg/ml WGA, and after refrigeration for 7 days diluted again to 0.0006 M GLcNac and 1.8 μ g/ml WGA. The dilute solution was deposited on MgCl_2 treated mica and incubated at room temperature for 20 minutes, then rinsed twice with 200 μ l aliquots of MilliQ water.

To assess the ability of GLcNac to promote reassembling of the WGA monomers into the native dimer form, the protein was diluted before treatment with the three sugars. The mono-GLcNac preparation was done as follows: Twenty μl of 0.06 mg/ml WGA in 10 mM HEPES pH 7.5 was added to 1.9 mg solid mono-GLcNac, making a solution 0.43 M in GLcNac. After 30 minutes the solution was diluted 10x with 10 mM HEPES pH 7.5, and deposited on freshly cleaved mica along with 20 μl 5 mM MgCl_2 . The di-GLcNac preparation was done as follows: Forty μl of 0.06 mg/ml WGA in 10 mM HEPES pH 7.5 was added to 4.9 mg of solid di-GLcNac, making a solution 0.28 M in di-GLcNac. After 30 minutes the solution was diluted 10x with 10 mM HEPES pH 7.5, then deposited on MgCl_2 treated mica. The tri-GLcNac preparation was done as follows: Ten μl of 0.06 mg/ml WGA was added to 0.9 mg of solid tri-GLcNac, making a solution 0.14 M in tri-GLcNac. After 30 minutes the solution was diluted 10x with 10 mM HEPES pH 7.5 and deposited on MgCl_2 treated mica. All samples were incubated at room temperature for 10 minutes, then rinsed twice with 200 μl aliquots of MilliQ water and allowed to dry overnight in petri dishes under ambient temperature and humidity.

Streptavidin-biotin sample preparation. The biotin treatment was performed as follows: a solution was made 0.03 mg/ml in streptavidin and 0.008 M biotin in 10 mM HEPES pH 7.5. That was diluted to 0.003 mg/ml streptavidin and 0.0008 M biotin several hours later. The dilute solution was deposited on MgCl_2 treated mica and incubated at room temperature for 20 minutes, then rinsed twice with 200 μl aliquots of MilliQ water.

The ability of biotin to promote the reassembling of the Streptavidin monomers was tested by first diluting the Streptavidin to 2 $\mu\text{g/ml}$ in 10 mM HEPES pH 7.5, then after incubation at room temperature for 10 minutes making this solution 0.0008 M in biotin. The solution was deposited on MgCl_2 treated mica and incubated at room temperature for 20 minutes, then rinsed twice with 200 μl aliquots of MilliQ water.

RESULTS

WGA. In all cases, the pre-dilution GLcNac treatment produced both monomers and dimers distributed on the mica, while the post-dilution treatment produced height distributions indicative of monomer only. This further confirms that a dissociation of WGA from dimer to monomer³⁵ is caused by dilution of the protein. The separation of subunits does not seem to be reversible by the addition of GLcNac after dilution.

The pre-dilution mono-GLcNac treated WGA RAMS data is presented in Figure 21. The data has been fitted to monomer dimensions 23 x 32 x 48 Å (Figure 21A) and dimer dimensions 36 x 48 x 60 Å (Figure 21B). Neither monomer nor dimer fits the data well, but as shown in Figure 22, a combination of 50% monomer 50% dimer improves the fit considerably. The post-dilution mono-GLcNac treated RAMS data is presented in Figure 23. Figure 23A shows the data fitted to monomer dimensions 21 x 32 x 47 Å, and figure 23B shows the data fitted to dimer dimensions 39.5 x 47 x 59 Å. The combination of the monomer and dimer curves indicate that these molecules are composed of 99% monomer and 1% dimer.

The pre-dilution di-GLcNac WGA RAMS data was fitted to dimensions corresponding to monomer, 25 x 33 x 46 Å, and dimensions corresponding to dimer, 40 x 46 x 60 Å. Figure 24 shows that a mixture of 68% monomer and 32% dimer fit the data much better than either monomer or dimer alone. The RAMS data for the post-dilution di-GLcNac WGA treatment is presented in Figure 25. The fit to (A) the dimensions corresponding to monomer, 19 x 29 x 43 Å, and (B) the dimensions corresponding to

dimer, 42 x 42 x 59 Å, clearly show that all molecules are accounted for by monomer only.

The RAMS data for the pre-dilution tri-GLcNac WGA treatment is presented in Figure 26. The data was fitted to dimensions corresponding to monomer, 25 x 38 x 46 Å, and dimensions corresponding to dimer, 38 x 46 x 63 Å. The best fit to the data is produced by a linear combination of 81% of the monomer curve and 19% of the dimer curve, shown in Figure 26. The post-dilution treatment with tri-GLcNac produced the data in Figure 27. Fits to dimensions corresponding to (A) monomer, 23 x 34 x 46 Å, and (B) dimer, 40 x 46 x 56 Å, reveal that this data is produced by monomers only of WGA.

Streptavidin. When a concentrated solution of Streptavidin is treated with biotin then diluted, the molecules retain their native tetramer conformation. However, when the solution is first diluted then biotin is added, the molecules dissociate into a mixture of dimers and monomers. This indicates that the dissociation of Streptavidin into subunits, caused by the dilution, is not completely reversible, but leaves most of the molecules in the dissociated state. This can be compared to the RAMS measurements made without the addition of biotin at all³⁵ (chapter 2), which show the Streptavidin molecules completely dissociated into monomers. In the dilute protein solution biotin seems to play a role in partially stabilizing the dimer, even making possible the rejoining of monomers to dimers, but is not able to bring about complete return to the tetramer state. If biotin is already bound to Streptavidin the dimer-dimer attraction is increased to the point that only tetramers are detectable. Further dilution at this point does not dissociate the tetramer.

The pre-dilution biotin treatment produces a distribution of molecules with the main peak in the height histogram at 60 Å as shown in Figure 28. The curve is the SOM prediction for Streptavidin in the tetramer oligomerization state. The data does not closely fit the SOM predictions in this case, probably because there is preferential orientation of the

molecules in their upright orientation as opposed to the majority of the molecules oriented having the largest surface area in contact with the mica surface, as is assumed for SOM. When Streptavidin forms two-dimensional crystals the molecules also orient standing upright, as is shown in Appendix A. It is likely that the molecule has chemical affinity for the mica that preferentially turns it upright, both as individual molecules and when two-dimensional crystals are formed.

The Streptavidin solution treated with biotin after dilution produced the RAMS data presented in Figures 29 and 30. The data is fitted to dimensions corresponding to (A) monomer $18 \times 24 \times 47 \text{ \AA}$, (B) dimer, $28 \times 44 \times 45 \text{ \AA}$, in Figure 29 and (A) tetramer, $44 \times 45 \times 56$, and (B) a linear combination of 24% monomer and 76% dimer, in Figure 30.

DISCUSSION

Here we have demonstrated that the RAMS technique can be readily applied to study ligand-induced conformational changes. For both wheat germ agglutinin and Streptavidin ligand binding causes a shift in the distribution of oligomers from lower to higher oligomerization state if the protein is treated with the ligand before dilution. This ligand effect is highly reduced or eliminated if ligand treatment is performed after the protein is diluted.

The shift in oligomerization is readily observable for the mono-, di- and tri-GLcNac treated WGA. The statistical model developed in chapter 2³⁵ holds well for these molecules as judged by the reasonably good fit between the observed and predicted distribution of oligomers. One may therefore conclude that there is no preferential orientation beyond that predicted by SOM. It is not surprising that GLcNac treatment does not cause the rejoining of subunits in the dilute solution, since the low protein concentration makes the average intermolecular distances in the solution large.

For Streptavidin, biotin binding appears to definitely promote the retention of tetramer conformation in the concentrated solution, but preferential orientation of the molecules on the mica make SOM fail in this case. It can reasonably be concluded, however, that the Streptavidin in the pre-dilution biotin treated solution does not dissociate into an oligomer smaller than the tetramer, since the height observed is larger than any dimension of monomer or dimer for this molecule. It is furthermore reasonable that the oligomer observed in this case is tetramer, since the height of two-dimensional crystals of Streptavidin is 60 Å, indicating that preferential orientation of individual Streptavidin molecules in the same manner is also highly possible. Again, it is not surprising that the post-dilution biotin treatment does little to promote reassembly of the tetramer, considering the high dilution of the protein at the time the biotin is added. It is interesting that a significant proportion of dimers are found in the post-dilution biotin treatment of Streptavidin, but if biotin treatment is omitted, the protein forms monomers.

The difference in oligomerization effects between post-dilution and pre-dilution ligand treatment for both WGA and Streptavidin show that the dissociation of subunits frequently seen in chapter 2 for Con A, Streptavidin, and WGA is indeed caused by the low dilution of the protein used in those studies. Since pre-dilution treatment causes an increase in the measured size for both Streptavidin and WGA but post-dilution treatment causes little or no increase in the measured size, this experiment confirms previous experiments which conclude that high dilution causes dissociation of oligomeric proteins.^{50,58}

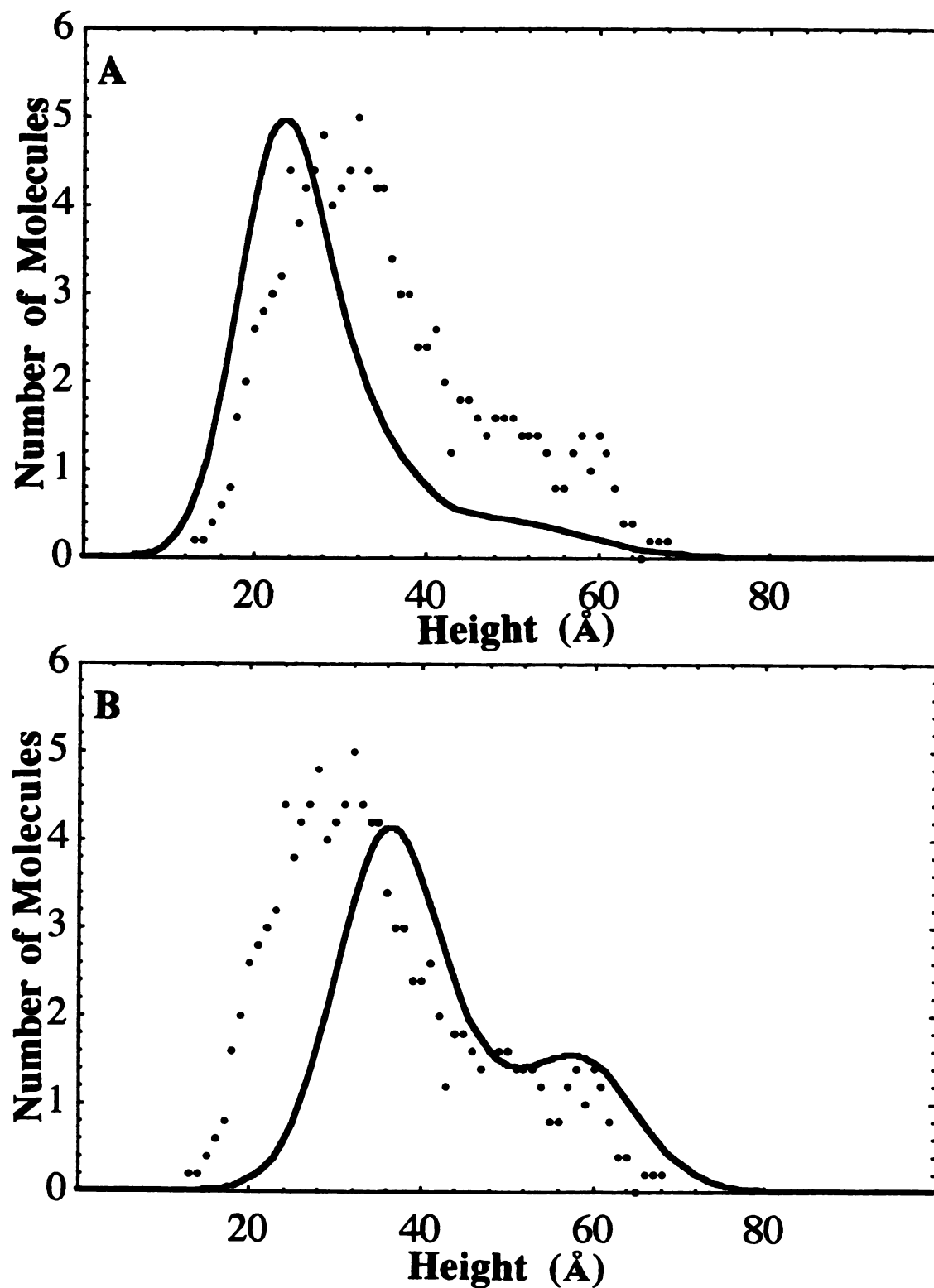


Figure 21: Pre-dilution mono-GLcNac treated WGA fitted to (A) monomer (B) dimer.

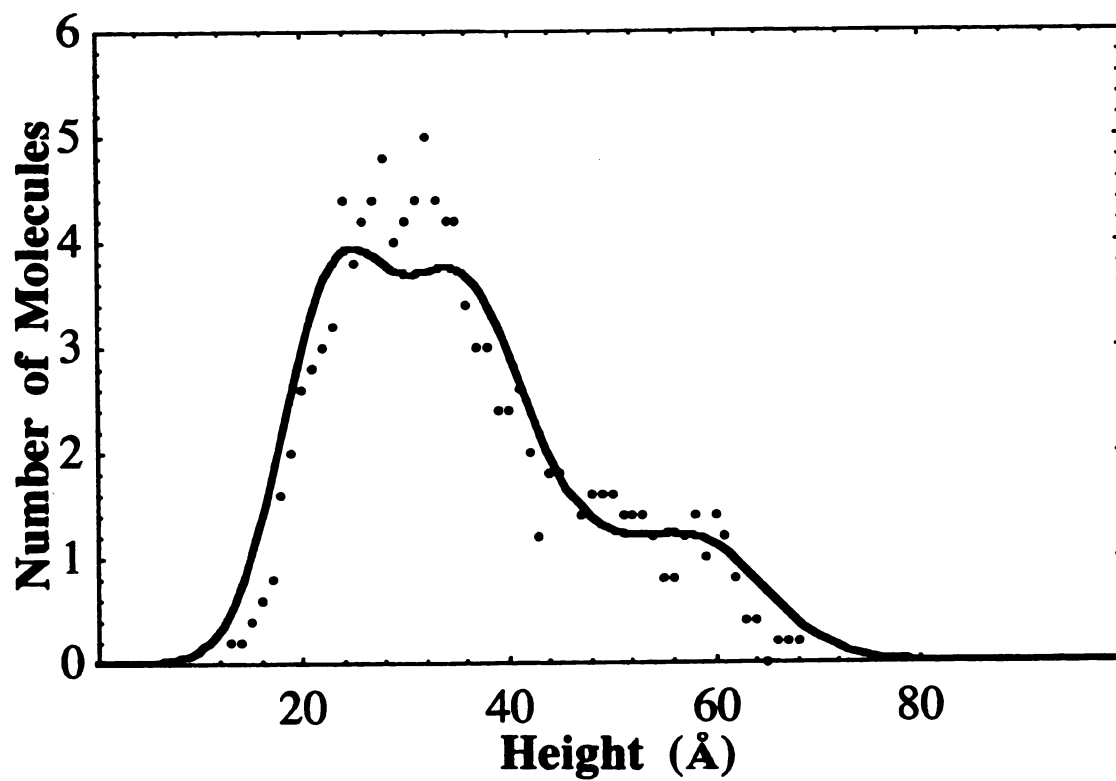


Figure 22: Pre-dilution mono-GLcNac treated WGA fitted to 50% monomer, 50% dimer.

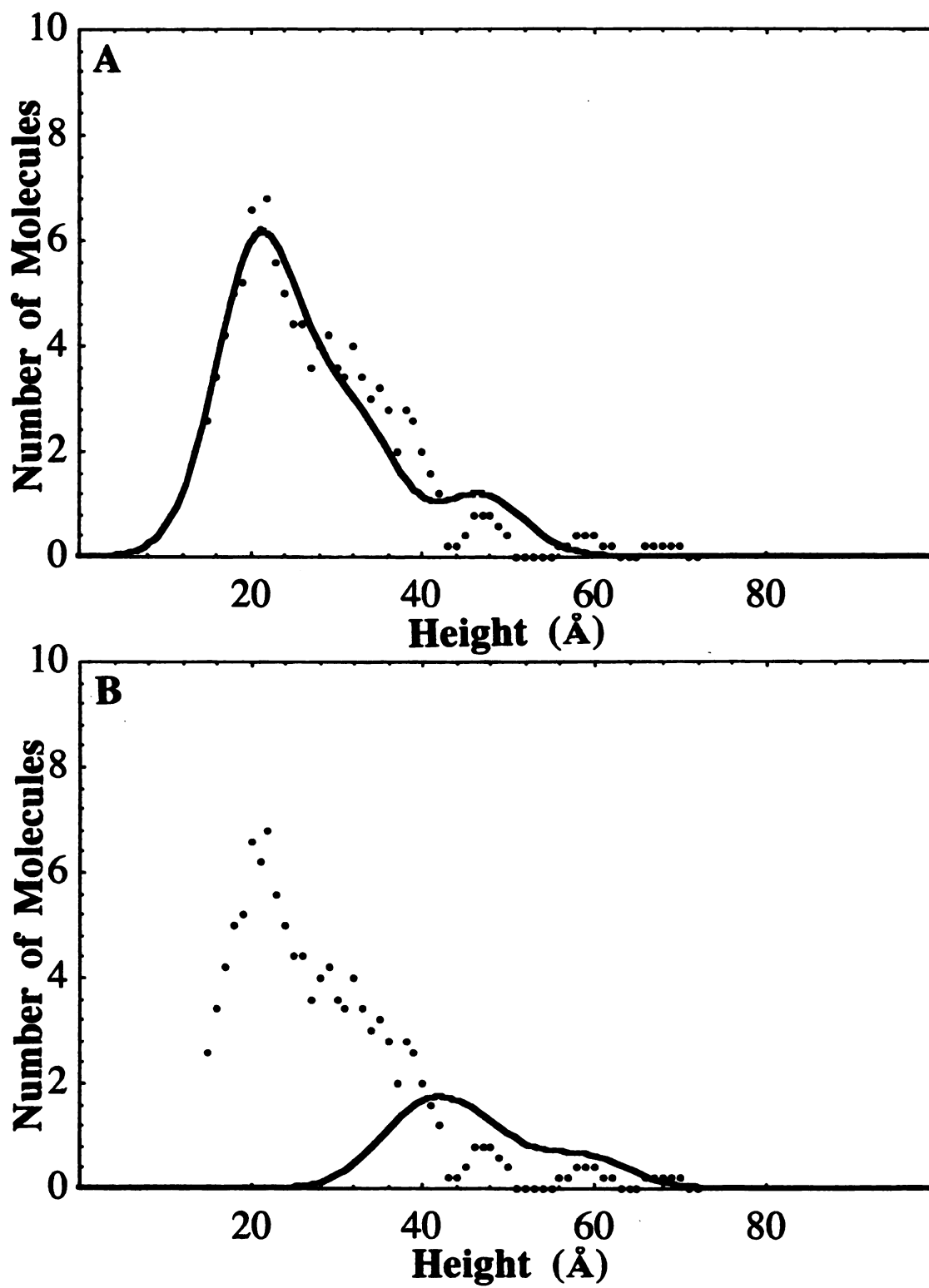


Figure 23: Post-dilution mono-GLcNac treated WGA fitted to (A) monomer (B) dimer.

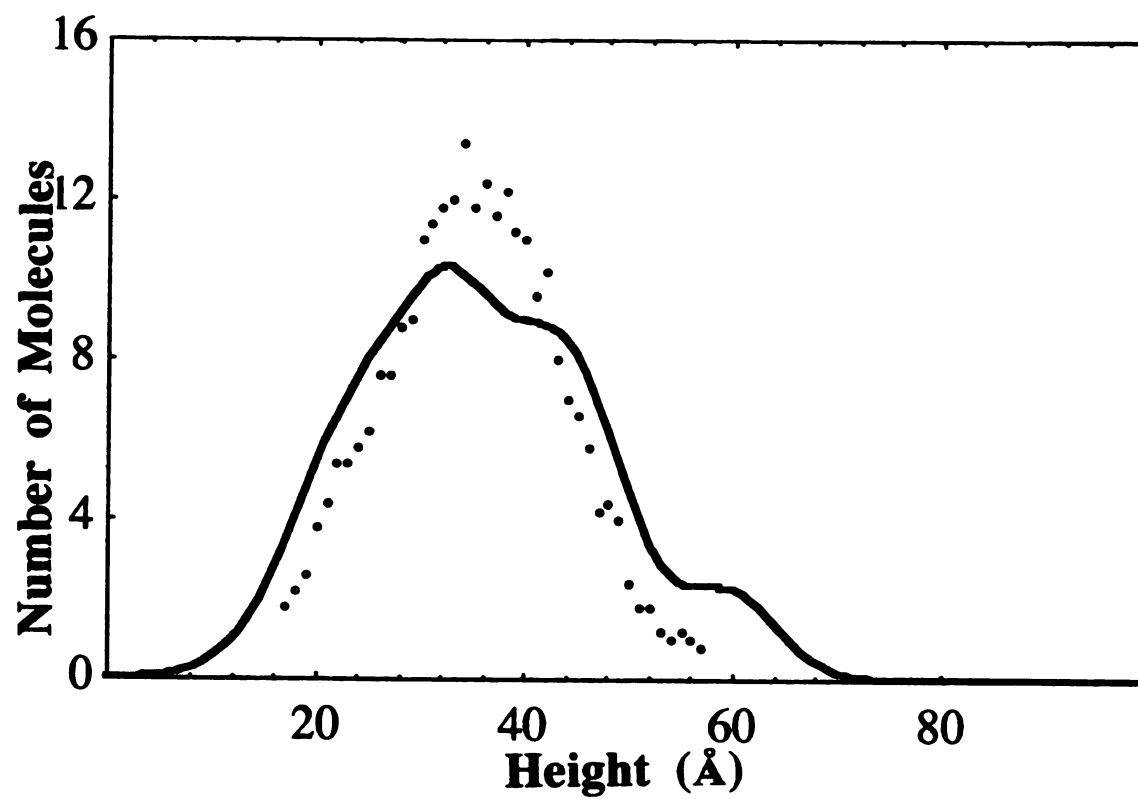


Figure 24: Pre-dilution di-GLcNac treated WGA fitted to 68% monomer, 32% dimer.

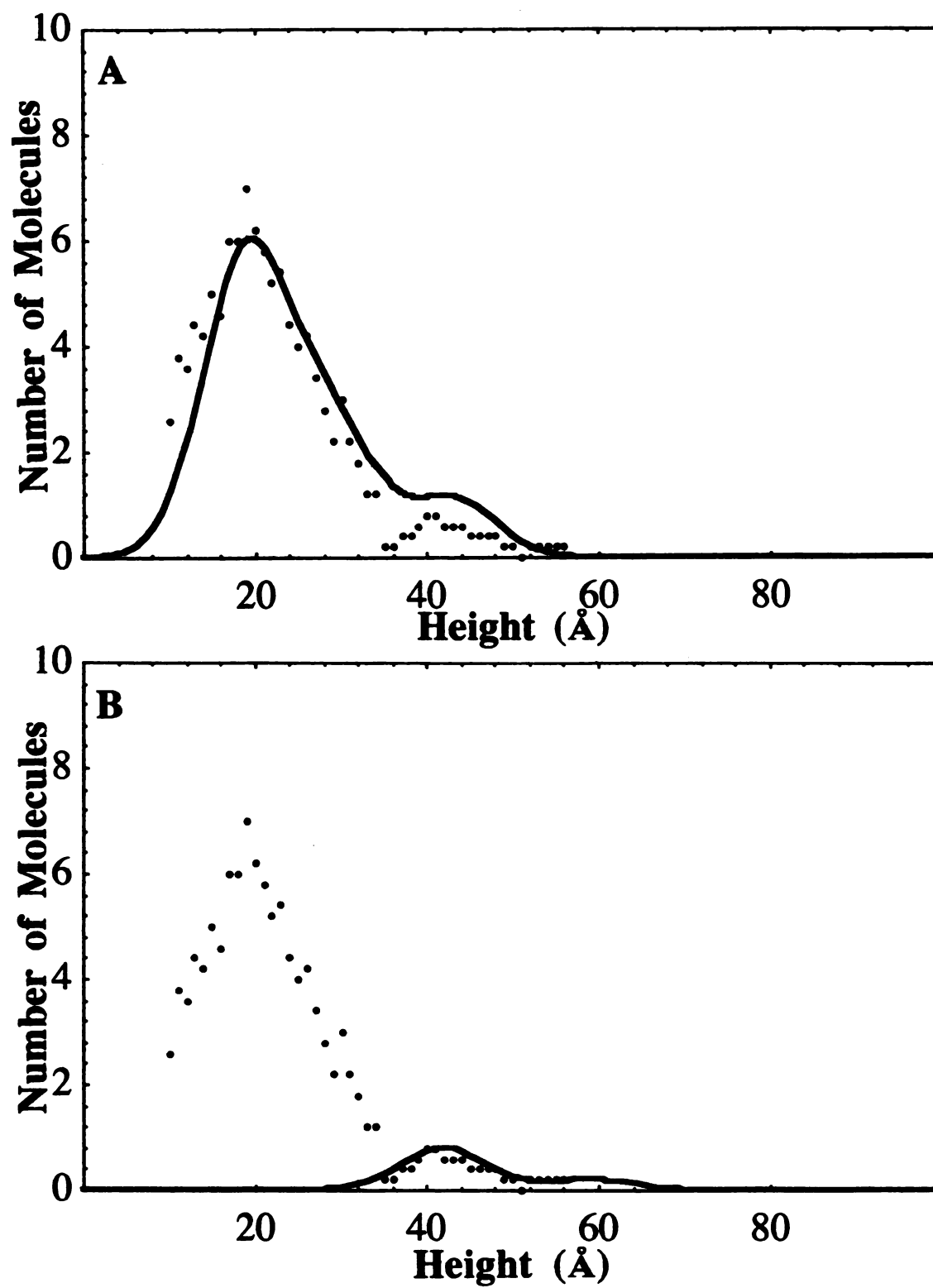


Figure 25: Post-dilution di-GlcNac treated WGA fitted to (A) monomer (B) dimer.

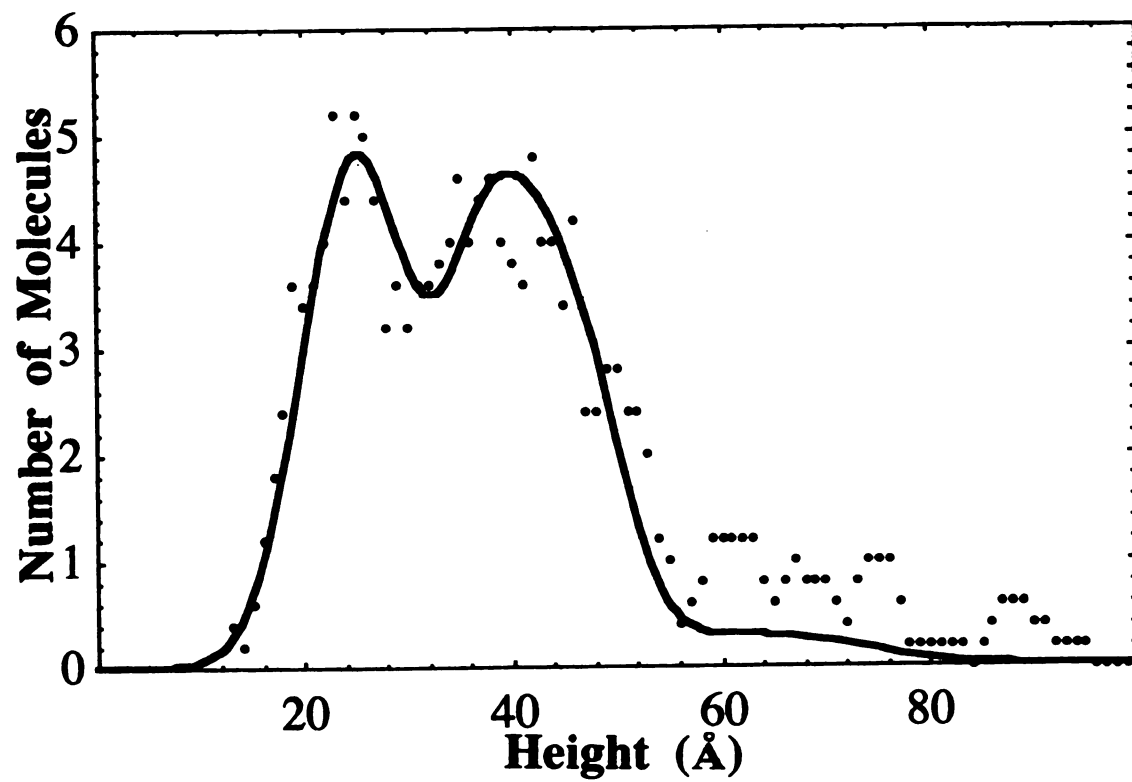


Figure 26: Pre-dilution tri-GLcNac treated WGA fitted to 81% monomer 19% dimer.

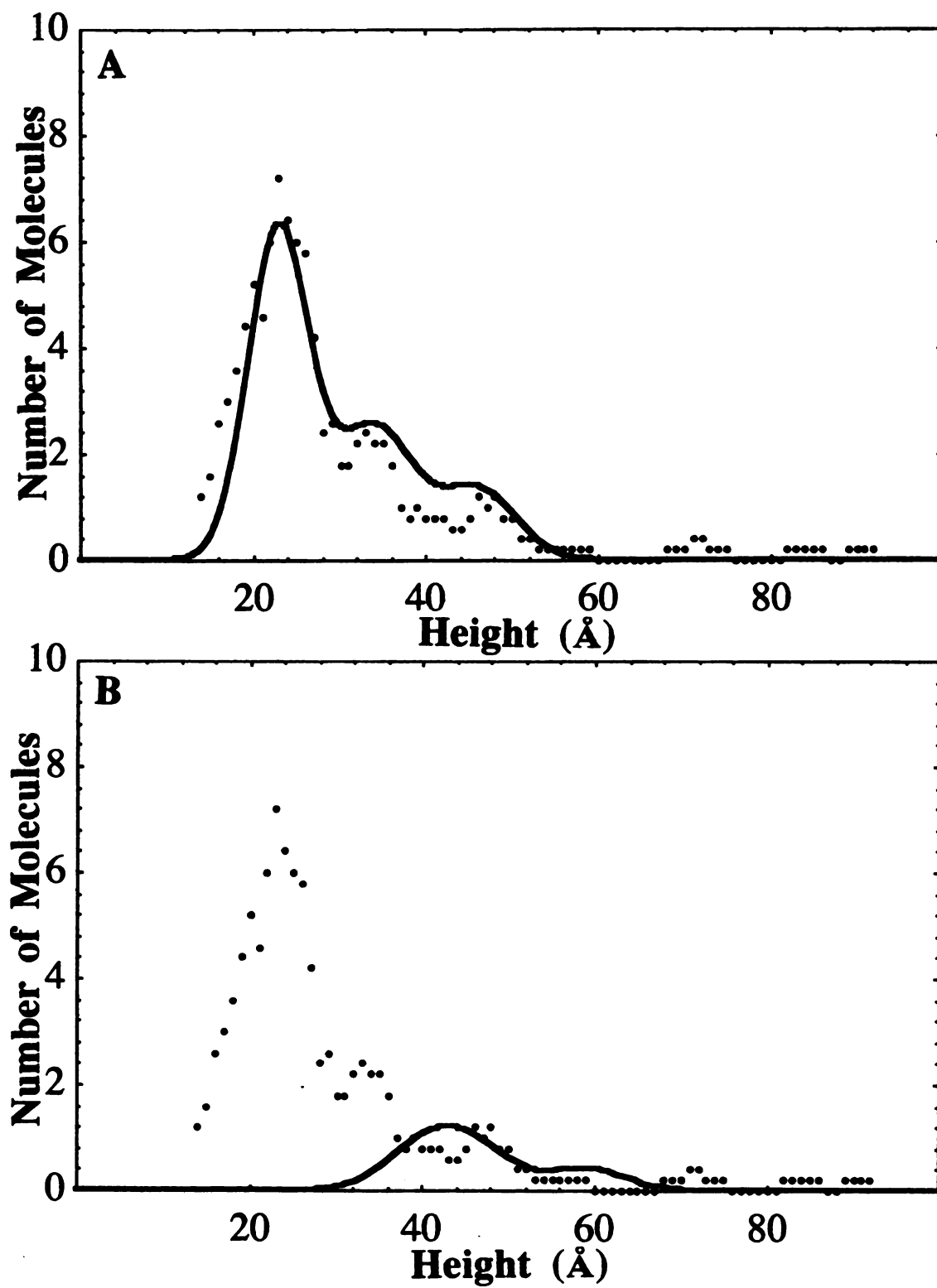


Figure 27: Post-dilution tri-GlcNac treated WGA fitted to (A) monomer (B) dimer.

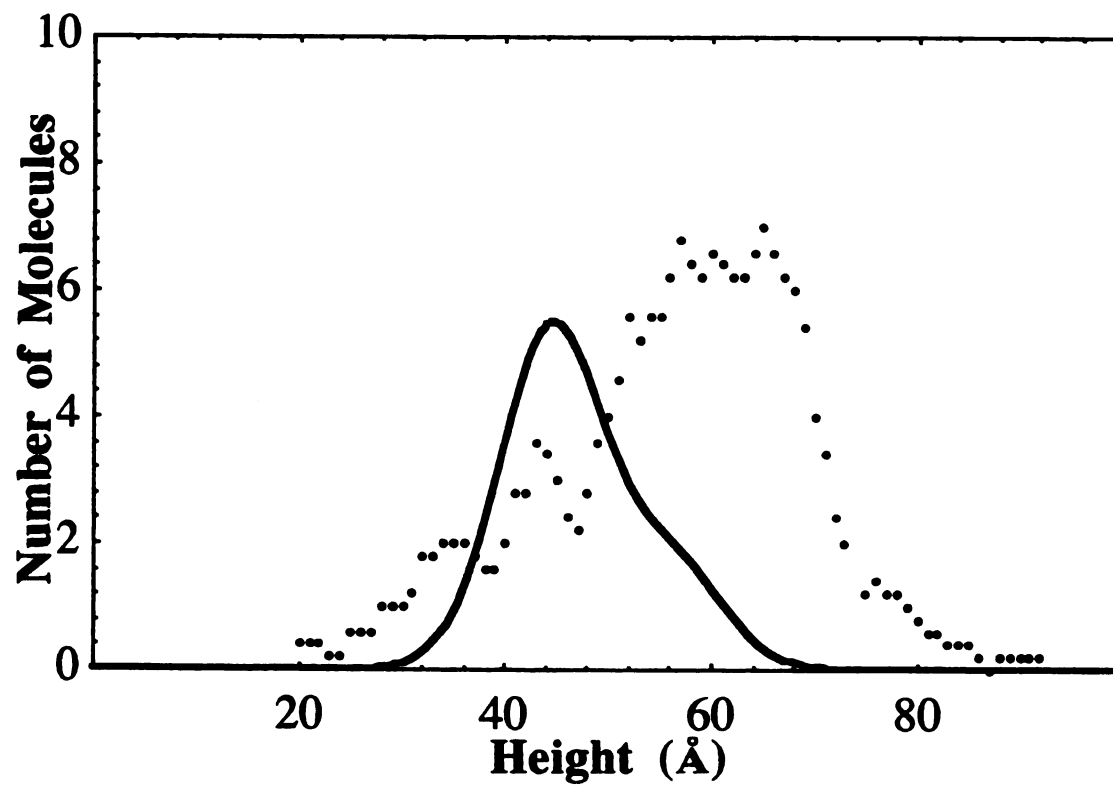


Figure 28: Pre-dilution biotin treated Streptavidin fitted to tetramer.

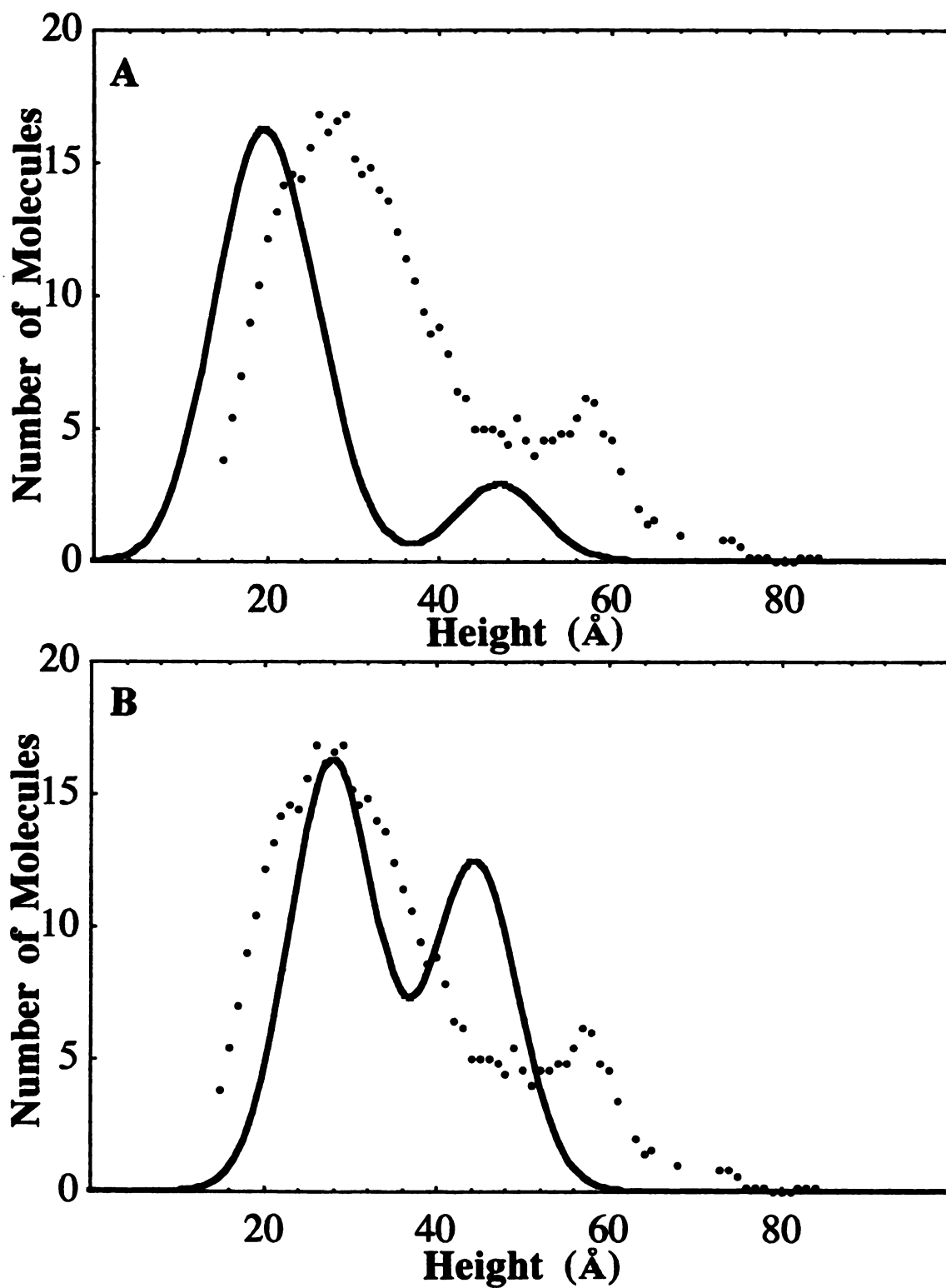


Figure 29: Post-dilution biotin treated Streptavidin fitted to (A) monomer (B) dimer.

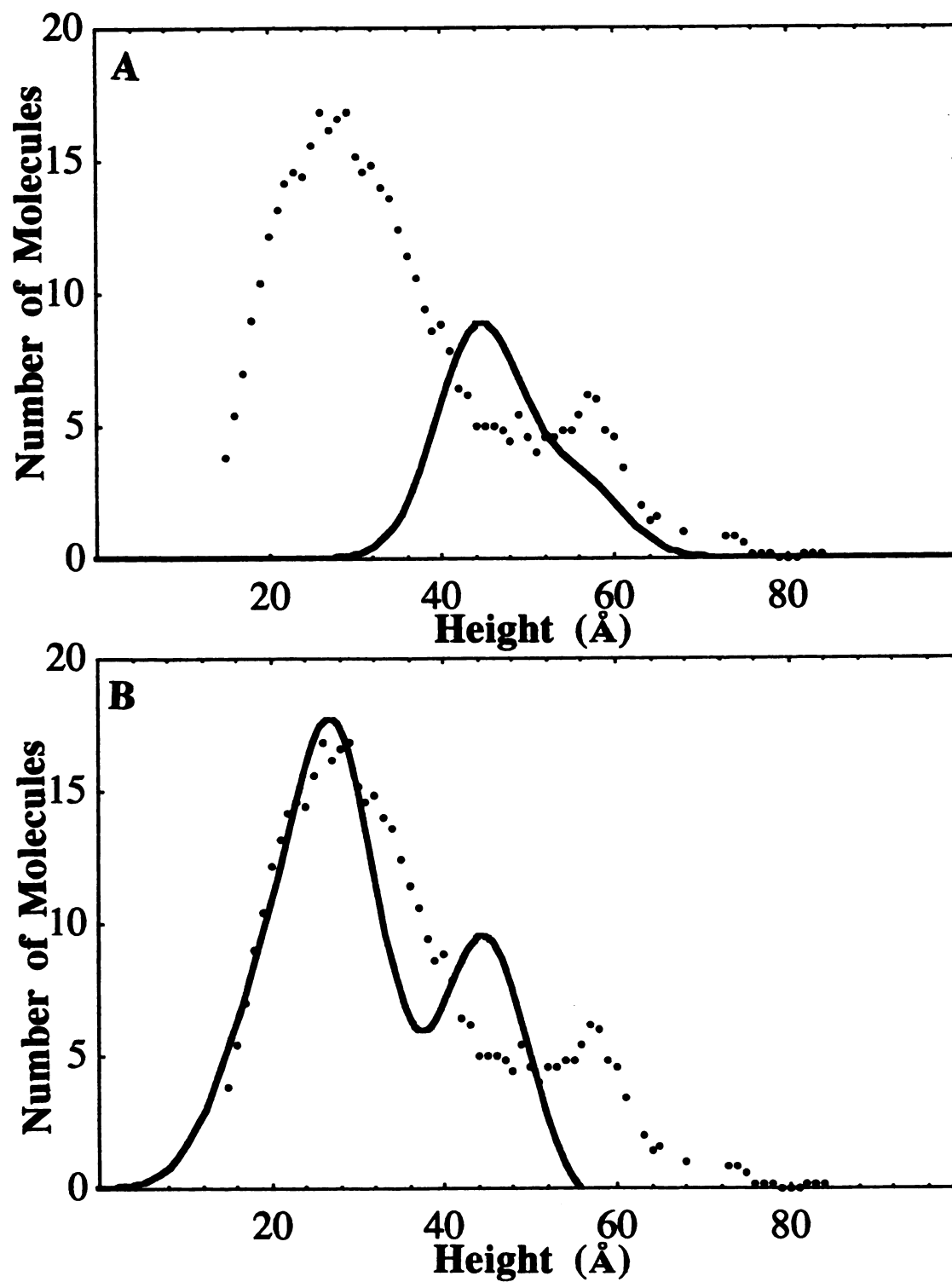


Figure 30: Post-dilution biotin treated Streptavidin fitted to (A) tetramer (B) 24% monomer 76% dimer.

CHAPTER 4: CONCLUSIONS

In this thesis we have presented numerous examples of how RAMS microscopy can be used to determine molecular size and ligand-induced conformational changes. Experiments with colloidal gold particles have demonstrated that the height measurements used in RAMS are not affected by the artifacts caused by finite tip size. Analysis of Ferritin by RAMS microscopy has demonstrated that protein molecules can be accurately sized using SFM. RAMS on WGA and Con A have shown that the technique can accurately measure the three dimensions of an asymmetrical molecule. The study of Protein A has demonstrated that even a molecule with a very irregular shape consisting of an elongated string of 4 small domains and one larger domain can give results consistent with the actual shape of the molecule, identifying the smaller sized part and the larger sized part of the molecule. The RAMS analysis of Streptavidin shows how the quaternary structure of a protein can be deduced using this technique. The application of RAMS to determine ligand-induced changes in oligomerization of protein molecules has been shown for two molecules, wheat germ agglutinin and Streptavidin. This demonstrates that RAMS³⁵ is the technique of choice for the study of ligand-induced conformational changes for most other proteins.^{1,2,3,4,5,6,7}

Random Adsorption Molecular Sizing is the only technique which can furnish direct measurements of molecular size without requiring high protein concentration or harsh fixation methods. Techniques exist which can be used to give an estimate of molecular size, but each has disadvantages compared to RAMS. Gel filtration is commonly used to determine the size of oligomeric proteins,⁹⁵ but the column must be calibrated by passing a set of proteins of known dimensions through it, and the reliability of the measurement is dependent on how closely the protein of interest resembles the proteins used to calibrate the column.⁹⁶ Since the protein of interest is of unknown shape, this makes the reliability of this technique uncertain. Sedimentation equilibrium measurements are dependent on the

amino acid composition, overall shape, hydration, and surface roughness of the molecule.⁹⁷ SDS-PAGE will give an accurate molecular weight for a monomeric protein, but generally does not give information about quaternary structure because the technique involves dissociation of subunits with detergents.⁹⁸ Non-denaturing gel electrophoresis is highly affected by molecular charge, which may be changed upon binding a highly charged ligand, so this makes ligand binding studies using this method impractical.⁹⁹ The growth of crystals with the quality required for X-ray crystallography can be a difficult and time consuming task. All of these techniques require high concentrations of protein, a possible problem when one wants to study concentration-dependent oligomeric changes. Microscopy techniques such as SEM and TEM may be used for large molecules such as Ferritin, but usually require the sample to be dehydrated and metal coated, eliminating the possibility that the molecule is in an active conformation.^{79,80} RAMS can reveal the three-dimensional size and oligomerization state of molecules that still hold their solvation water,⁴¹ and give direct measurements of size without relying on uncertain calibration with other protein molecules, and can produce these size measurements with only a few minutes of sample preparation time. This technique should find many applications in the study of the interrelationship of protein structure and function.

The technique to determine molecular size presented here, termed Random Adsorption Molecular Sizing, has demonstrated its usefulness in measurements on a number of different molecules. The use of the projected height of adsorbed protein molecules as the defining parameter for molecular dimensions makes the RAMS technique minimally dependent on probe tip geometry and extends the range of the SFM from nanometer to angstrom resolution. The relative ease and speed with which these high resolution measurements of molecular dimension may now be performed should provide investigators with new opportunities to examine changes in protein conformation and

oligomerization that depend on protein concentration and ligand or allosteric activator binding.

REFERENCES

References

- ¹ Kosk-Kosicka, D., Bzedga, T., Wawrzynow, A. (1989) *J. Biol Chem.* 264, 19495-9.
- ² Kosk-Kosicka, D., Bzedga, T., (1988) *J. Biol. Chem.* 263, 18184-18189.
- ³ Carlson, C. A., Parsons, T. F., & Preiss, J. (1976) *The Journal of Biological Chemistry* 251, 7886.
- ⁴ Stenger, J. E., Tegtmeyer, P, Mayr, G. A., Reed, M., Wang, Y. Wang, P., Hough, P. V. C., Mastrangelo, I. A. (1994) *EMBO J.* 13 6011-6020.
- ⁵ Davey, M. J., Funnell, B. E. (1994) *J. Biol Chem.* 269 29908-29913.
- ⁶ Koblan, K., S., Ackers, G. K. (1991) *Biochemistry* 30, 7817-7821.
- ⁷ Takahashi, M. (1989) *J. Biol. Chem.* 264(1), 288-95.
- ⁸ Bennig, G., Rohrer, H. (1982) *Helv. Phys. Acta.*, 55, 726.
- ⁹ Bennig, G., Quate, C. F., Gerber, C. (1986) *Phys. Rev Lett.*, 56, 930.
- ¹⁰ Pohl, D. W., Denk, W., Lanz, M. (1984) *Appl. Phys. Lett.* 44, 651.
- ¹¹ Lal, R., John, S. A. (1994) *Am. J. Physiol.*, 266, C1-C21.
- ¹² Marti, O., Amrein, M. STM and SFM in Biology, Academic Press, New York (1993).
- ¹³ Putman, C. A. J; Van der Werf, K. O.; Grooth, B. G.; Van Hulst, N. F.; Greve, J. (1994) *Appl. Phys. Lett.* 64 (18) 2454-2456.
- ¹⁴ Zhong, Q.; Inniss, D.; Kjoller, K.; Elings, V. B. (1993) *Surf. Sci. Lett.* 290 L688.
- ¹⁵ Keller, D. J., & Franke, F. S. (1993) *Surf. Sci.* 294, 409.
- ¹⁶ Eppel, S. J., Zypman, F. R., & Marchant, R. E. (1993) *Langmuir* 9, 2281.
- ¹⁷ Allen, M. J., Hud, N. V., Balooch, M., Tench, R. J., Seikhais, W. J., & Balhorn, R. (1992) *Ultramicroscopy* 42-44, 1095.
- ¹⁸ Markiewicz, P., & Goh, M. C. (1994) *Langmuir* 10, 5-7.
- ¹⁹ Radmacher, M., Cleveland, J. P., Fritz, M., Hansma, H. G., Hansma, P. K. (1994) *Biophysical Journal* 66, 2159-2165.
- ²⁰ Ducker, W. A.; Xu, Z.; Israelachvili, J. N. (1994) *Langmuir* 10, 3279-3289.
- ²¹ Rabinovich, Y. I.; Yoon, R.-H. (1994) *Langmuir* 10, 1903-1909.
- ²² Weisenhorn, A. L.; Maivald, P.; Butt, H.-J.; Hansma, P. K. (1992) *Physical Review B* 45, 11226-11232.

-
- ²³ Tsao, Y.H.; Evans, D. F.; Wennerstrøm, H. (1993) *Science* 262, 547-549.
- ²⁴ F. Schabert; A. Hefti; K. Goldie; A. Stemmer; A. Engel; E. Meyer; R. Overney; H.-J. Güntherodt, *Ultramicroscopy* 42-44, 1118-1124 (1992).
- ²⁵ Ohnishi, S., Hara, M., Furuno, T., Sasabe, H. (1992) *Biophys. J.*, 63, 1425-1431.
- ²⁶ Radmacher, M., Fritz, M., Cleveland, J. P., Walters, D. A., Hansma, P. K. *Langmuir*, 10, 3809-3814 (1994).
- ²⁷ Häberle, W., Hørber, J. H. K., Ohnesorge, F., Smith, D. P. E., Binnig, G. (1992) *Ultramicroscopy*, 42-44, 1161-1167.
- ²⁸ Bustamante, C., Vasenka, J., Tang, C. L., Reese, W., Guthold, M., & Keller, R. (1992) *Biochemistry* 31, 22.
- ²⁹ Hansma, H. G., Sinsheimer, R. L., Li, M. Q., & Hansma, P. K. (1992a) *Nucleic Acids Res.* 20, 3585. and Hansma, H. G., Vasenka, J., Seigerist, C., Kelderman, G., Morret H., Sinsheimer, R. L., Elings, V., Bustamante, C., & Hansma, P. K. (1992b) *Science*, 242, 209.
- ³⁰ Henderson, E. (1992) *Nucleic Acids Res.* 20, 445.
- ³¹ Shaiu, W.-L., Vesenka, J., Jondle, D., Henderson, E., & Larson, D. D. (1993) *J. Vac. Sci. Technol. A* 11(4), 820-823.
- ³² Roberts, C. J., Williams, P. M., Davies, J., Dawkes, A. C., Sefton, J., Edward, J. C., Haymes, A. G., Bestwick, C., Davies, M. C., & Tendler, S. J. B. (1995) *Langmuir* 11, 1822-1826.
- ³³ Ill, C. R., Keivens, V. M., Hale, J. E., Nakamura, K. K., Jue, R. A., Cheng, S., Melcher, E. D., & Drake, B. (1993) *Biophys. J.* 64, 919.
- ³⁴ Marchant, R. E., Lea, A. S., Andrade, J. D., & Bockenstedt, P. (1992) *Journal of Colloid and Interface Science* 148, 261.
- ³⁵ Gilchrist, M., Preiss, J., Schindler, M., Dantus, M. Imaging the Molecular Dimensions and Oligomerization of Single Protein Molecules by Random Adsorption Molecular Sizing (RAMS) Microscopy. *Biochemistry* (in preparation).
- ³⁶ Cantor, C. R. Biophysical Chemistry W. H. Freeman, San Francisco 1980.
- ³⁷ Sommerville, J., Scheer, U. Electron Microscopy in Molecular Biology: a practical approach, IRL Press, Washington, DC 1987.
- ³⁸ Dykstra, M. J. A Manual of Applied Techniques for Biological Electron Microscopy Plenum, New York 1993.
- ³⁹ Ornstein, L., & Davis, B. J. (1962) Disk Electrophoresis. Reprinted by Distillation Products Industries, Rochester.
- ⁴⁰ Bezanilla, M., Manne, S., Laney, D. E., Lyubchenko, Y. L., & Hansma, H. G. (1995) *Langmuir* 11, 655.

-
- ⁴¹ Hu, J., Xiao, X.-D., Ogletree, D. F., & Salmeron, M. (1995) *Science* 268, 267.
- ⁴² Thundat, T., Zheng, X.-Y., Chen, G. Y., & Warmack, R. J. (1993) *Surface Science Letters* 294, L939-L943.
- ⁴³ Harrison, P. M. (1963) *J. Mol. Biol.* 6, 404.
- ⁴⁴ Massover, W. H. (1993) *Micron* 24, 389-437.
- ⁴⁵ Yau, S.-T., Zhou, Y. (1995) *Modern Physics Letters B* 9, 187-193.
- ⁴⁶ Allen, A. K., Neuberger, A., Sharon, N. (1973) *Biochem. J.* 131, 155-162.
- ⁴⁷ Nagata, Y., Burger, M. M. (1974) *J. Biol. Chem.* 249, 3116-3122.
- ⁴⁸ Privat, J. P., Delmotte, F., Monsigny, M. (1974) *FEBS Letts.* 46, 224-228.
- ⁴⁹ Protein DataBank, 9WGA, C. S. Wright.
- ⁵⁰ Steindler, D. A., Bradley, R. H. (1983) *Neuroscience* 10(1), 219-241.
- ⁵¹ Denman, R. B., Wedler, F. C. (1984) *Arch. Biochem. Biophys.* 232(2), 427-40.
- ⁵² Bowles, D.J., Pappin, D.J. (1988) *TIBS* 13, 60-64.
- ⁵³ Sharon, N., Lis, H. (1989) *Science* 246, 227-235.
- ⁵⁴ Kéry, V. (1991) *Int. J. Biochem.* 23, 631-640.
- ⁵⁵ Lin, S. S., Levitan, I. B. (1991) *Trends Neurosci.* 14, 273-278.
- ⁵⁶ Reeke, G.N., Becker, J. W., Cunningham, B. A., Wang, J. L., Yahara, I., Edelman, G. M. Structure and Function of Concanavalin A. From Concanavalin A. T. K. Chodhury and A. K. Weiss Ed. Plenum Publishing Corp., New York.
- ⁵⁷ Protein DataBank, 1CN1, Shoham et. al.
- ⁵⁸ Gordon, J. A., Young, R. K. (1978) *The Journal of Biological Chemistry* 254, 1932-1937.
- ⁵⁹ Gunther, G. R., Wang, J. L., Yahara, I., Cunningham, B. A., Edelman, G. M. (1973) *Proc. Nat. Acad. Sci. USA* 70, 1012
- ⁶⁰ Dumas, M., Schwab, M. E., Thoenen, H. (1979) *J. Neurobiol.* 10, 179-197.
- ⁶¹ Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y., Shimada, I. (1992) *Biochemistry* 31, 9665-9672.
- ⁶² Sjodahl, J. (1977) *Eur. J. Biochem.* 73, 343-351. and Sjodahl, J. (1977) *Eur. J. Biochem.* 78, 471-490.
- ⁶³ Ghetie, V., Nilsson, K., Sjoquist, J. (1974) *Eur. J. Immunol.* 4, 500-505.

-
- ⁶⁴ Protein DataBank, 1FC2, Deisenhofer 1981.
- ⁶⁵ Bjørk, I, Petersson, B.-Å., Sjöquist, J. (1972) *Eur. J. Biochem.* 29, 579-584.
- ⁶⁶ Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J., Salemme, F. R. (1989) *Science* 243, 85-88.
- ⁶⁷ Green, N. M. (1975) *Adv. Protein Chem.* 29, 85-133.
- ⁶⁸ Protein DataBank, 1PTS, Dupont Protein Crystallography Group
- ⁶⁹ Hendrickson, W. A.; Pähler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A.; Phizackerley, R. P. (1989) *Proc. Natl. Acad. Sci.* 86 2190-2194.
- ⁷⁰ Ghosh, P., Meyer, C., Remey, E., Peterson, D., & Preiss, J. (1992) *Archives of Biochemistry and Biophysics* 296, 122.
- ⁷¹ Stark, D. M., Timmerman, K. T., Barry, G. F., Preiss, J., & Kishore, G. M. (1992) *Science* 258, 287-292.
- ⁷² Carlson, C. A., Parsons, T. F., & Preiss, J. (1976) *The Journal of Biological Chemistry* 251, 7886.
- ⁷³ Iglesias, A. A., Kakefuda, G., Preiss, J. (1991) *Plant Physiol.* 97, 1187-1195.
- ⁷⁴ Chang, Y.-y., Kakefuda, G., Iglesias, A. A., Buikema, W. J., & Preiss, J. (1992) *Plant Molecular Biology* 20, 37.
- ⁷⁵ Fanger, B. O., Stephens, J. E., Staros, J. V. (1989) *FASEB J.* 3(1), 71-5.
- ⁷⁶ Kraus, S., Quaiser, A., Oren, M., Montenarh, M. (1988) *J. Virol.* 62(12), 4737-44.
- ⁷⁷ Ghosaini, L. R., Brown, A. M., Sturtevant, J. M. (1988) *Biochemistry* 27(14), 5257-61.
- ⁷⁸ Clore, G. M., Gronenborn, A. M. NMR of Proteins, CRC Press, Boca Raton 1993.
- ⁷⁹ Revet, B., Brahms, S., Brahms, G. (1995) *Proc. Natl. Acad. Sci. USA* 92(16), 7535-9.
- ⁸⁰ Smith, J. M. A., Ford, G. C., Harrison, P. M., Yariv, J., Kalb, A. J. (1989) *J. Mol. Biol.* 205, 465-467.
- ⁸¹ Radmacher, M., Fritz, M., Hansma, H. G., & Hansma, P. K. (1994b) *Science* 265, 1577.
- ⁸² Takahashi, M. (1989) *J. Biol. Chem.* 264(1), 288-95.
- ⁸³ Davey, M. J., Funnell, B. E. (1994) *J. Biol. Chem.* 269 29908-29913.
- ⁸⁴ Ghosaini, L. R., Brown, A. M., Sturtevant, J. M. (1988) *Biochemistry* 27(14), 5257-61.
- ⁸⁵ Burger, M. M., Goldberg, A. R. (1967) *Proc. Natl. Acad. Sci. USA* 57, 359-366.

-
- ⁶⁴ Protein DataBank, 1FC2, Deisenhofer 1981.
- ⁶⁵ Björk, I, Petersson, B.-Å., Sjöquist, J. (1972) *Eur. J. Biochem.* 29, 579-584.
- ⁶⁶ Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J., Salemme, F. R. (1989) *Science* 243, 85-88.
- ⁶⁷ Green, N. M. (1975) *Adv. Protein Chem.* 29, 85-133.
- ⁶⁸ Protein DataBank, 1PTS, Dupont Protein Crystallography Group
- ⁶⁹ Hendrickson, W. A.; Pähler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A.; Phizackerley, R. P. (1989) *Proc. Natl. Acad. Sci.* 86 2190-2194.
- ⁷⁰ Ghosh, P., Meyer, C., Remy, E., Peterson, D., & Preiss, J. (1992) *Archives of Biochemistry and Biophysics* 296, 122.
- ⁷¹ Stark, D. M., Timmerman, K. T., Barry, G. F., Preiss, J., & Kishore, G. M. (1992) *Science* 258, 287-292.
- ⁷² Carlson, C. A., Parsons, T. F., & Preiss, J. (1976) *The Journal of Biological Chemistry* 251, 7886.
- ⁷³ Iglesias, A. A., Kakefuda, G., Preiss, J. (1991) *Plant Physiol.* 97, 1187-1195.
- ⁷⁴ Charng, Y.-y., Kakefuda, G., Iglesias, A. A., Buikema, W. J., & Preiss, J. (1992) *Plant Molecular Biology* 20, 37.
- ⁷⁵ Fanger, B. O., Stephens, J. E., Staros, J. V. (1989) *FASEB J.* 3(1), 71-5.
- ⁷⁶ Kraus, S., Quaiser, A., Oren, M., Montenarh, M. (1988) *J. Virol.* 62(12), 4737-44.
- ⁷⁷ Ghosaini, L. R., Brown, A. M., Sturtevant, J. M. (1988) *Biochemistry* 27(14), 5257-61.
- ⁷⁸ Clore, G. M., Gronenborn, A. M. NMR of Proteins, CRC Press, Boca Raton 1993.
- ⁷⁹ Revet, B., Brahms, S., Brahms, G. (1995) *Proc. Natl. Acad. Sci. USA* 92(16), 7535-9.
- ⁸⁰ Smith, J. M. A., Ford, G. C., Harrison, P. M., Yariv, J., Kalb, A. J. (1989) *J. Mol. Biol.* 205, 465-467.
- ⁸¹ Radmacher, M., Fritz, M., Hansma, H. G., & Hansma, P. K. (1994b) *Science* 265, 1577.
- ⁸² Takahashi, M. (1989) *J. Biol. Chem.* 264(1), 288-95.
- ⁸³ Davey, M. J., Funnell, B. E. (1994) *J. Biol Chem.* 269 29908-29913.
- ⁸⁴ Ghosaini, L. R., Brown, A. M., Sturtevant, J. M. (1988) *Biochemistry* 27(14), 5257-61.
- ⁸⁵ Burger, M. M., Goldberg, A. R. (1967) *Proc. Natl. Acad. Sci. USA* 57, 359-366.

-
- ⁸⁶ Privat, J. P., Charleir, M. (1978) *Eur. J. Biochem.* 84, 79-85.
- ⁸⁷ Gurd, J. W. (1977) *Biochemistry* 16, 369-374.
- ⁸⁸ Nicholson, G. L. (1974) *Int. Rev. Cytol.* 39, 89-190.
- ⁸⁹ Pinto de Silva, P., Torrisi, M. R., Kachar, B. (1981) *J. Cell. Biol.* 91, 361-372.
- ⁹⁰ Jordan, F., Bassett, E., Redwood, W. R. (1977) *Biochem. Biophys. Res. Comm.* 75 1015-1021.
- ⁹¹ Weber, P; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. (1989) *Science* 243, 85-88.
- ⁹² Pähler, A.; Hendrickson, W. A.; Kolks, M. A. G.; Argaraña, C. E.; Cantor, C. R. (1987) *The Journal Of Biological Chemistry* 262, 13933-13937.
- ⁹³ Sano, T., Pansdori, M. W., Smith, C. L., Cantor, C. R. (1994) *Advances in Biomagnetic Separation*, eds. Uhlén, M., Hornes, E., Olsvik, Ø. (Eaon, Natick, MA)pp. 21-29.
- ⁹⁴ Sano, T.; Cantor, C. R. (1995) *Proc. Natl. Acad. Sci.* 92 3180-3184.
- ⁹⁵ Chinami, M., Sasaki, S., Hachiya, N., Yuge, K., Ohsugi, T., Maeda, H., Shingu, M. (1994) *J. Gen. Virol.* 75, 227-281.
- ⁹⁶ LeMaire et. al. (1989) *Anal. Biochem.* 177, 50-56.
- ⁹⁷ Minton, A. P. (1989) *Anal. Biochem.* 176, 209-216.
- ⁹⁸ Shapiro, A. L. (1967) *Biochem. Biophys. Res. Comm.* 28, 815-820.
- ⁹⁹ Hedrick, J. L., Smith, A. J. (1968) *Arch. Biochem. Biophys.* 126, 155-164.

APPENDICES

APPENDIX A

SOLID FILMS OF PROTEIN

Introduction If the concentration of the protein solution is too high to allow the molecules to isolate themselves on the mica surface (generally greater than 0.05 mg/ml), a solid film of protein is formed. For streptavidin and ADPgpp, this film has characteristics of a two dimensional crystal, with uniform vertical dimensions, while in other cases the film is rather amorphous. In both situations, some structural information can still be gleaned from analyzing these images. In the case of an amorphous film that completely covers the mica, it is necessary to scrape away a portion of the film to reveal the flat background of the mica in order to be sure of measuring the correct thickness of the film.

Streptavidin. Streptavidin has previously been known to form 2 dimensional crystals by spreading on lipid layers,¹ and by spreading on a film of poly(1-benzyl-L-histadine).² The crystals formed by spreading on the poly(1-benzyl-L-histadine) revealed ordered 2-D arrays of Streptavidin molecules with 44 Å spacing. Since the molecular dimensions are approximately 42 x 42 x 56 Å³ this lattice spacing would indicate that the molecules are oriented with their long sides together. Therefore, the thickness of the layer would be the larger dimension of the molecule, 56Å. In six different SFM experiments using protein concentrations of 0.5 to 5 mg/ml, Streptavidin forms islands or a lattice type formation on the mica after deposition of the protein and drying the sample. This type of formation was seen in all experiments when the protein concentration was greater than 0.05

¹ Ku, A. C.; Darst, S. A.; Robertson, C. R.; Gast, A. P. Kornberg, R. D. (1993) *J. Phys. Chem.* 97, 3013-3016.

² Furano, T.; Sasabe, H. (1993) *Biophysical Journal* 65, 1714-1717.

³ Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. (1989) *Science* 243, 85.

mg/ml. These formations always have a very uniform thickness of about 60 Å throughout the entire formation. This indicates a type of 2-D molecular ordering similar to that found in the 2-D crystals of Streptavidin. See figure 31 for a typical two-dimensional array of Streptavidin.

ADPgpp. When ADPgpp wild type formed a layer on the mica, it was in the form of a uniform lattice or islands 100 Å tall. The concentration necessary to form a layer instead of isolated particles is about 0.08 mg/ml. In four of five experiments in which a solid coating was formed this uniform layer was observed. The remaining sample produced a layer solidly covering the mica, for which the height could not be determined because the flat mica surface was not visible by either gaps in the layer or scraping a portion of the layer with the tip. Since the crystal structure for this protein is unknown, the interpretation of this data is somewhat uncertain, however, the molecular weight of the intact tetrameric enzyme, 200kDa, would be amenable to the interpretation that the molecule was not terribly elongated and one dimension of the molecule was 90 Å. The single molecule SFM data then can be used to predict the other two dimensions of the molecule to be 50 Å and 70 Å. The uniformity of the formations within and between different experiments suggests that these are a formation with at least short-range ordering of the molecules. The one sample of the ADPgpp enzyme from *Anabaena* in which a height could be determined also showed uniform flat islands 100 Å tall, indicating a similar structure for this analogous enzyme. See figure 32 for a typical image of a two-dimensional array of ADPgpp wild type molecules.

Protein A. Protein A, as mentioned in chapter 2, is an elongated molecule about 18 Å in diameter and 120 Å long. When the concentration of protein is too high to form individual particles on the mica, Protein A will form a layer, and in the six different experiments in which a layer was seen, it was always less than 30 Å tall. A concentration greater than 0.05 mg/ml will form this layer instead of isolated particles on the surface. The layer was not smooth or uniform in height, indicating the absence of long-range order to the molecular packing, but the molecules apparently lay flat on the mica and form one or two layers of protein under the conditions of these experiments.

Con A. When Con A forms a solid layer on the mica, the layer is uneven, and it is 40 to 75 Å tall. This would indicate that these formations are composed of Con A dimers with some tetramer form also possibly present. The thicker layer could also be the result of two layers of dimers.



Figure 31: Two-dimensional array of Streptavidin.

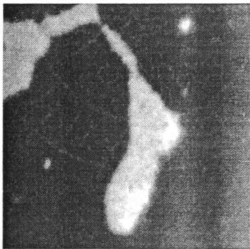


Figure 32: Two-dimensional array of ADPgpp wild type.

APPENDIX B

VISUAL BASIC PROGRAM TO AUTOMATICALLY MEASURE AND TABULATE THE HEIGHT OF MOLECULES IN AN AFM IMAGE

```
Sub Different2()  
Application.ScreenUpdating = False  
Dim B  
Dim Co  
Dim CV  
Dim ACV  
Dim BCV  
Dim CCV  
Dim DCV  
Dim ECV  
Dim FCV  
Dim GCV  
Dim HCV  
Dim JCV  
Dim KCV  
Dim LCV  
Dim MCV  
Dim NCV  
Dim OCV  
Dim PCV  
Dim QCV  
Dim RCV  
Dim SCV  
Dim TCV  
Dim UCV  
Dim VCV  
Dim WCV  
Dim XCV  
Dim YCV  
Dim ZCV  
  
B = 0  
Co = 0  
SHEETS("Sheet1").Select  
Range("F5").Select  
SHEETS("Sheet2").Select  
Range("B1").Select  
ActiveCell.FormulaR1C1 = "value"  
Range("C1").Select  
ActiveCell.FormulaR1C1 = "background"  
Range("D1").Select  
ActiveCell.FormulaR1C1 = "height"  
Range("D2").Select  
ActiveCell.FormulaR1C1 = "=RC[-2]-RC[-1]"
```

```

Range("D2").Select
Selection.Copy
Range("D3:D500").Select
ActiveSheet.Paste
Application.CutCopyMode = False
Range("B2").Select
SHEETS("Sheet1").Select
For Lines = 1 To 244 Step 1
    Do Until ActiveCell = ""
        CV = Selection.Value
        If ActiveCell < 50 Then
            B = CV + B
            Co = Co + 1
        End If
        If ActiveCell < 50 Then GoTo Line2

        ActiveCell.Offset(-1, -1).Range("A1").Select
        ACV = Selection.Value

        ActiveCell.Offset(0, 1).Range("A1").Select
        BCV = Selection.Value

        ActiveCell.Offset(0, 1).Range("A1").Select
        CCV = Selection.Value

        ActiveCell.Offset(1, -2).Range("A1").Select
        DCV = Selection.Value

        ActiveCell.Offset(0, 2).Range("A1").Select
        ECV = Selection.Value

        ActiveCell.Offset(1, -2).Range("A1").Select
        FCV = Selection.Value

        ActiveCell.Offset(0, 1).Range("A1").Select
        GCV = Selection.Value

        ActiveCell.Offset(0, 1).Range("A1").Select
        HCV = Selection.Value

        ActiveCell.Offset(-3, -3).Range("A1").Select
        ICV = Selection.Value

        ActiveCell.Offset(0, 1).Range("A1").Select
        JCV = Selection.Value

        ActiveCell.Offset(0, 1).Range("A1").Select
        KCV = Selection.Value

        ActiveCell.Offset(0, 1).Range("A1").Select

```

LCV = Selection.Value

**ActiveCell.Offset(0, 1).Range("A1").Select
MCV = Selection.Value**

**ActiveCell.Offset(1, -4).Range("A1").Select
NCV = Selection.Value**

**ActiveCell.Offset(0, 4).Range("A1").Select
OCV = Selection.Value**

**ActiveCell.Offset(1, -4).Range("A1").Select
PCV = Selection.Value**

**ActiveCell.Offset(0, 4).Range("A1").Select
QCV = Selection.Value**

**ActiveCell.Offset(1, -4).Range("A1").Select
RCV = Selection.Value**

**ActiveCell.Offset(0, 4).Range("A1").Select
SCV = Selection.Value**

**ActiveCell.Offset(1, -4).Range("A1").Select
TCV = Selection.Value**

**ActiveCell.Offset(0, 1).Range("A1").Select
UCV = Selection.Value**

**ActiveCell.Offset(0, 1).Range("A1").Select
VCV = Selection.Value**

**ActiveCell.Offset(0, 1).Range("A1").Select
WCV = Selection.Value**

**ActiveCell.Offset(0, 1).Range("A1").Select
XCV = Selection.Value**

ActiveCell.Offset(-2, -2).Range("A1").Select

If CV < ACV Then GoTo Line2

If CV < BCV Then GoTo Line2

If CV < CCV Then GoTo Line2

If CV < DCV Then GoTo Line2

If CV < ECV Then GoTo Line2

If CV < FCV Then GoTo Line2

If CV < GCV Then GoTo Line2

If CV < HCV Then GoTo Line2

If CV < ICV Then GoTo Line2

If CV < JCV Then GoTo Line2

```

If CV < KCV Then GoTo Line2
If CV < LCV Then GoTo Line2
If CV < MCV Then GoTo Line2
If CV < NCV Then GoTo Line2
If CV < OCV Then GoTo Line2
If CV < PCV Then GoTo Line2
If CV < QCV Then GoTo Line2
If CV < RCV Then GoTo Line2
If CV < SCV Then GoTo Line2
If CV < TCV Then GoTo Line2
If CV < UCV Then GoTo Line2
If CV < VCV Then GoTo Line2
If CV < WCV Then GoTo Line2
If CV < XCV Then GoTo Line2

```

Line1:

```

Selection.Font.ColorIndex = 3
If ActiveCell > 300 Then Selection.Font.ColorIndex = 4
ActiveCell.Copy
'Calculates the average background

```

```

Dim B1
Dim B2
Dim B3
Dim B4
Dim B5
Dim B6
Dim B7
Dim B8
Dim B9
Dim B10
Dim B11
Dim B12
Dim B13
Dim B14
Dim B15
Dim B16
Dim B17
Dim B18
Dim B19
Dim B20
Dim B21
Dim B22
Dim B23
Dim B24
Dim B25
Dim B26
Dim B27
Dim B28
Dim B29
Dim B30
Dim B31

```

Dim B32
 Dim B33
 Dim B34
 Dim B35
 Dim B36
 Dim B37
 Dim B38
 Dim B39
 Dim B40
 Dim B41
 Dim BK1
 Dim BK2
 Dim BK3
 Dim BK4

Dim BK

ActiveCell.Offset(-5, -5).Range("A1").Select
 B1 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B2 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B3 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B4 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B5 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B6 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B7 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B8 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B9 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B10 = Selection.Value
 ActiveCell.Offset(0, 1).Range("A1").Select
 B11 = Selection.Value
 ActiveCell.Offset(1, 0).Range("A1").Select
 B12 = Selection.Value
 ActiveCell.Offset(1, 0).Range("A1").Select
 B13 = Selection.Value
 ActiveCell.Offset(1, 0).Range("A1").Select
 B14 = Selection.Value
 ActiveCell.Offset(1, 0).Range("A1").Select
 B15 = Selection.Value
 ActiveCell.Offset(1, 0).Range("A1").Select
 B16 = Selection.Value
 ActiveCell.Offset(1, 0).Range("A1").Select
 B17 = Selection.Value

ActiveCell.Offset(1, 0).Range("A1").Select
B18 = Selection.Value
ActiveCell.Offset(1, 0).Range("A1").Select
B19 = Selection.Value
ActiveCell.Offset(1, 0).Range("A1").Select
B20 = Selection.Value
ActiveCell.Offset(1, 0).Range("A1").Select
B21 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B22 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B23 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B24 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B25 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B26 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B27 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B28 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B29 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B30 = Selection.Value
ActiveCell.Offset(-1, 0).Range("A1").Select
B31 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B32 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B33 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B34 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B35 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B36 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B37 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B38 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B39 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B40 = Selection.Value
ActiveCell.Offset(0, -1).Range("A1").Select
B41 = Selection.Value
ActiveCell.Offset(6, 5).Range("A1").Select

BK1 = (B1 + B2 + B3 + B4 + B5 + B6 + B7 + B8 + B9 + B10)
 BK2 = (B11 + B12 + B13 + B14 + B15 + B16 + B17 + B18 + B19 + B20 +
 B21)
 BK3 = (B21 + B22 + B23 + B24 + B25 + B26 + B27 + B28 + B29 + B30 +
 B31)
 BK4 = (B32 + B33 + B34 + B35 + B36 + B37 + B38 + B39 + B40 + B41)
 BK = (BK1 + BK2 + BK3 + BK4) / 41

'Excludes background points if < 80% or > 120% of average

If B1 > 1.2 * BK Then B1 = BK
 If B1 < 0.8 * BK Then B1 = BK
 If B2 > 1.2 * BK Then B2 = BK
 If B2 < 0.8 * BK Then B2 = BK
 If B3 > 1.2 * BK Then B3 = BK
 If B3 < 0.8 * BK Then B3 = BK
 If B4 > 1.2 * BK Then B4 = BK
 If B4 < 0.8 * BK Then B4 = BK
 If B5 > 1.2 * BK Then B5 = BK
 If B5 < 0.8 * BK Then B5 = BK
 If B6 > 1.2 * BK Then B6 = BK
 If B6 < 0.8 * BK Then B6 = BK
 If B7 > 1.2 * BK Then B7 = BK
 If B7 < 0.8 * BK Then B7 = BK
 If B8 > 1.2 * BK Then B8 = BK
 If B8 < 0.8 * BK Then B8 = BK
 If B9 > 1.2 * BK Then B11 = BK
 If B11 < 0.8 * BK Then B11 = BK
 If B12 > 1.2 * BK Then B12 = BK
 If B12 < 0.8 * BK Then B12 = BK
 If B13 > 1.2 * BK Then B13 = BK
 If B13 < 0.8 * BK Then B13 = BK
 If B14 > 1.2 * BK Then B14 = BK
 If B14 < 0.8 * BK Then B14 = BK
 If B15 > 1.2 * BK Then B15 = BK
 If B15 < 0.8 * BK Then B15 = BK
 If B16 > 1.2 * BK Then B16 = BK
 If B16 < 0.8 * BK Then B16 = BK
 If B17 > 1.2 * BK Then B17 = BK
 If B17 < 0.8 * BK Then B17 = BK
 If B18 > 1.2 * BK Then B18 = BK
 If B18 < 0.8 * BK Then B18 = BK
 If B19 > 1.2 * BK Then B19 = BK
 If B19 < 0.8 * BK Then B19 = BK
 If B20 > 1.2 * BK Then B20 = BK
 If B21 > 1.2 * BK Then B21 = BK
 If B21 < 0.8 * BK Then B21 = BK
 If B22 > 1.2 * BK Then B22 = BK
 If B22 < 0.8 * BK Then B22 = BK
 If B23 > 1.2 * BK Then B23 = BK
 If B23 < 0.8 * BK Then B23 = BK

If B24 > 1.2 * BK Then B24 = BK
 If B24 < 0.8 * BK Then B24 = BK
 If B25 > 1.2 * BK Then B25 = BK
 If B25 < 0.8 * BK Then B25 = BK
 If B26 > 1.2 * BK Then B26 = BK
 If B26 < 0.8 * BK Then B26 = BK
 If B27 > 1.2 * BK Then B27 = BK
 If B27 < 0.8 * BK Then B27 = BK
 If B28 > 1.2 * BK Then B28 = BK
 If B28 < 0.8 * BK Then B28 = BK
 If B29 > 1.2 * BK Then B29 = BK
 If B29 < 0.8 * BK Then B29 = BK
 If B30 > 1.2 * BK Then B30 = BK
 If B30 < 0.8 * BK Then B30 = BK
 If B31 > 1.2 * BK Then B31 = BK
 If B31 < 0.8 * BK Then B31 = BK
 If B32 > 1.2 * BK Then B32 = BK
 If B32 < 0.8 * BK Then B32 = BK
 If B33 > 1.2 * BK Then B33 = BK
 If B33 < 0.8 * BK Then B33 = BK
 If B34 > 1.2 * BK Then B34 = BK
 If B34 < 0.8 * BK Then B34 = BK
 If B35 > 1.2 * BK Then B35 = BK
 If B35 < 0.8 * BK Then B35 = BK
 If B36 > 1.2 * BK Then B36 = BK
 If B36 < 0.8 * BK Then B36 = BK
 If B37 > 1.2 * BK Then B37 = BK
 If B37 < 0.8 * BK Then B37 = BK
 If B38 > 1.2 * BK Then B38 = BK
 If B38 < 0.8 * BK Then B38 = BK
 If B39 > 1.2 * BK Then B39 = BK
 If B39 < 0.8 * BK Then B39 = BK
 If B40 > 1.2 * BK Then B40 = BK
 If B40 < 0.8 * BK Then B40 = BK
 If B41 > 1.2 * BK Then B41 = BK
 If B41 < 0.8 * BK Then B41 = BK

BK1 = (B1 + B2 + B3 + B4 + B5 + B6 + B7 + B8 + B9 + B10)

BK2 = (B11 + B12 + B13 + B14 + B15 + B16 + B17 + B18 + B19 + B20)

BK3 = (B21 + B22 + B23 + B24 + B25 + B26 + B27 + B28 + B29 + B30 +

B31)

BK4 = (B32 + B33 + B34 + B35 + B36 + B37 + B38 + B39 + B40 + B41)

BK = (BK1 + BK2 + BK3 + BK4) / 41

SHEETS("Sheet2").Select

ActiveSheet.Paste

ActiveCell.Offset(0, 1).Range("A1").Select

Selection.Value = BK

ActiveCell.Offset(1, -1).Range("A1").Select

SHEETS("Sheet1").Select

Line2:

```

    ActiveCell.Offset(1, 0).Range("A1").Select
    Loop
    ActiveCell.Offset(-250, 1).Range("A1").Select
Next Lines
Application.ScreenUpdating = True
SHEETS("Sheet3").Select
ActiveCell.Offset(2, 0).Range("A1").Select
Selection.Value = Co
ActiveCell.Offset(2, 0).Range("A1").Select
Selection.Value = B
End Sub

```

Sub Color()

```

    Do Until ActiveCell = ""
        Do Until ActiveCell = ""
            Application.ScreenUpdating = False
            If Selection.Font.ColorIndex = 2 Then Selection.Font.ColorIndex = 3
            ActiveCell.Offset(1, 0).Range("A1").Select
            Loop
            ActiveCell.Offset(-250, 1).Range("A1").Select
        Loop
        Application.ScreenUpdating = True
    End Sub

```

APPENDIX C**VISUAL BASIC ROUTINE TO AUTOMATICALLY OPEN SEVERAL
TEXT FILES AND PROCESS THE DATA THEN CLOSE THE FILES.**

Sub series()

```
Workbooks.Open Filename:="D:\EXCEL\MARTHA\cona2.XLS"  
SHEETs.Add  
SHEETs.Add  
Application.Run Macro:="PICKS.XLS!different2"  
ActiveWorkbook.Save  
ActiveWorkbook.Close
```

```
Workbooks.Open Filename:="D:\EXCEL\MARTHA\cona3.XLS"  
SHEETs.Add  
SHEETs.Add  
Application.Run Macro:="PICKS.XLS!different2"  
ActiveWorkbook.Save  
ActiveWorkbook.Close
```

```
Workbooks.Open Filename:="D:\EXCEL\MARTHA\cona4.XLS"  
SHEETs.Add  
SHEETs.Add  
Application.Run Macro:="PICKS.XLS!different2"  
ActiveWorkbook.Save  
ActiveWorkbook.Close
```

```
Workbooks.Open Filename:="D:\EXCEL\MARTHA\cona5.XLS"  
SHEETs.Add  
SHEETs.Add  
Application.Run Macro:="PICKS.XLS!different2"  
ActiveWorkbook.Save  
ActiveWorkbook.Close  
End Sub
```

APPENDIX D

**C++ PROGRAM TO CONVERT PARK SCIENTIFIC INSTRUMENTS
BINARY IMAGE TO X, Y TEXT ARRAY SUITABLE FOR READING
INTO EXCEL.**

```
//primitive binary to text file converter for Park scientific
#include <stdio.h>
#include <stdlib.h>

#define FNAME "1220003a"

int main(void)
{
    FILE *fp,*fp1;
    int word;

    /* place the word in a file */
    fp = fopen(FNAME, "rb");
    fp1 = fopen("test.txt","wb");
    if (fp == NULL)
    {
        printf("Error opening file %s\n", FNAME);
        exit(1);
    }
    if (fp1 == NULL)
    {
        printf("Error opening file %s\n", "test.txt");
        exit(1);
    }
    for (int i=1;i<=256;i++) {
        for (int j=1;j<=256;j++) {
            word=getw(fp);
            if (ferror(fp)){
                printf("Error writing to file\n");
                exit(1);
            }
            else
                fprintf(fp1,"%d",word);
            if(j==256) fprintf(fp1,"\n");
            else fprintf(fp1,"t");
        }
        // fprintf(fp1,"\n");
    }
    fclose(fp);
}
```

```
    fclose(fp1);  
  
    return 0;  
}
```

APPENDIX E

**MATHEMATICA ROUTINE FOR READING A TEXT FILE OF
TABULATED AFM DATA AND FITTING THE DATA TO APPROPRIATE
GAUSSIAN CURVES.**

(*Reads a text file with the data and fits the data to a specified gaussian curve. The coefficient to each exponential is the percentage of molecules at a particular height according to SOM theory. If the gaussians have different widths, it is necessary to change the coefficient to make the area under the curve correspond to the expected distribution of protein sizes according to SOM theory.*)

```
M=54 Exp[-((x-28)^2)/50] + 33 Exp[-((x-36)^2)/50] + 13
Exp[-((x-57)^2)/50]
data=ReadList["ConAtext2",Number, RecordLists->True];
Fit[data, {M}, {x}]
g2=Plot[%, {x, 0, 100}, PlotRange->{0, 20}];
g1=ListPlot[data, PlotRange->{{0, 100}, {0, 20}}];
Show[g1, g2]
```



MICHIGAN STATE UNIV. LIBRARIES



31293015722030