5-2001

Hydrodynamic Properties of DNA Polymerases

Angela Sherell Byrd

Hydrodynamic Properties of DNA Polymerases

An Undergraduate Honors Thesis

Louisiana State University
Department of Biological Sciences

Angela S. Byrd
May 2001

Abstract

Escherichia coli DNA Polymerase I (Pol I) and *Thermus aquaticus* DNA Polymerase I (Taq) are homologous polymerases that function in different environments. *E. coli*'s optimum temperature is 37°C, while *T. aquaticus*' is 75°C. In the present study, hydrodynamic properties of the two sets of polymerases were studied using analytical ultracentrifugation to determine if the proteins undergo a change in conformation with temperature and to resolve the current ambiguity in the orientation of the 5' nuclease domain of Taq. The sedimentation coefficient, a property related to the size and shape of a macromolecule, of Taq, Pol I, and their shortened versions Klentaq and Klenow were measured at 4°C, 20°C, and 40°C using the sedimentation velocity method. S values were normalized to $S_{20,w}$ values (using density and viscosity measurements of solution conditions and partial specific volume measurements from equilibrium runs) and compared. All polymerases experienced a small decrease in $S_{20,w}$ with temperature, indicating a slight conformational expansion in conformation with increasing temperature. Taq and Pol I were found to have smaller S values than expected based on their molecular weights when modeled as spheres, suggesting that the proteins are more elongated and less spherical. HYDROPRO, a computer program that calculates hydrodynamic properties based on protein crystal structure, was used to predict S values, and the predicted S values were compared to measured ones. Taq's predicted S values support the compact orientation of the 5' nuclease domain, however, this contradicts the conclusions made from the $S_{20,w}$ data. Future studies using dynamic light scattering measurements of the diffusion coefficient and the radius of gyration, other hydrodynamic properties, as well as alternate measurements of the proteins' partial specific volumes, will be used to make final conclusions concerning the solution structures of the polymerases.

Table of Contents

Literature Review

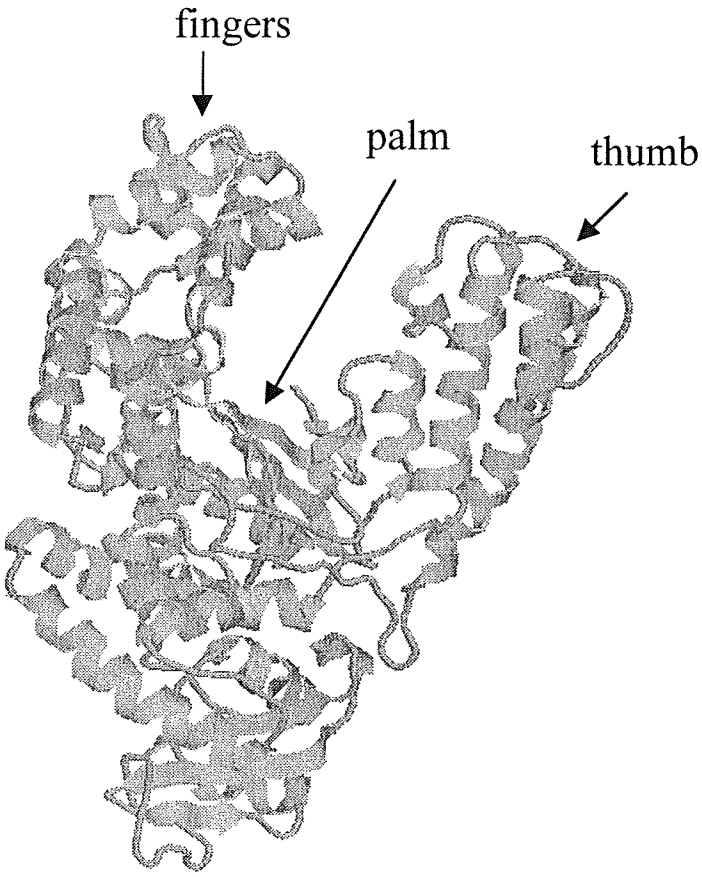
DNA Polymerases

DNA polymerases are enzymes with the essential job of replicating the genome and passing information from one generation to the next. More specifically, DNA polymerases catalyze the addition of a deoxynucleoside triphosphate (dNTP) to the free 3' hydroxyl of a primer chain, under the direction of a template strand (Li, et al., 1998). The mechanism of this phosphoryl transfer reaction is conserved among both prokaryotic and eukaryotic polymerases and involves two metal ions. The two metal ions facilitate the reaction by assisting in the 3' hydroxyl's attack on the incoming dNTP, stabilizing the transition state, and assisting in the leaving of the pyrophosphate (Beese and Steitz, 1991; Steitz, 1999).

DNA polymerases sometimes contain not only a polymerase domain, but also a 5' nuclease domain and a 3'-5' exonuclease domain. The 5' nuclease domain removes RNA primers and damaged DNA nucleotides. The 3'-5' exonuclease domain has an editing function; it removes misincorporated nucleotides during DNA synthesis (Voet, et al., 1999).

DNA polymerases are divided into several families based on analyses of their amino acid sequences and, when available, their crystal structures. All polymerases (prokaryotic and eukaryotic) have a similar overall structure, resembling the right human hand with thumb, palm, and fingers regions (Figure 1). The active site of the polymerase reaction is located in the palm region. The fingers region interacts with the incoming dNTP, and the thumb region helps position the DNA. Although the fingers and thumb regions vary among families, the palm regions are all similar. One family of polymerases

Figure 1. The polymerase domain of *Thermus aquaticus* DNA polymerase I with fingers, thumb, and palm regions.



is the DNA polymerase I family, which includes *Escherichia coli* DNA Polymerase I and *Thermus aquaticus* DNA Polymerase I (Steitz, 1999).

Escherichia coli DNA Polymerase I and the Klenow fragment

E. coli DNA Polymerase I (Pol I) has polymerase activity, 3'-5' exonuclease activity, and 5' nuclease activity. Pol I is made of two fragments, the large fragment and the small fragment, that can be separated by proteases (Klenow and Henningsen, 1970). The large fragment contains the polymerase domain and the 3'-5' exonuclease domain. The small fragment contains the 5' nuclease activity. Together these fragments make up Pol I with a molecular weight of 103,118. The primary roles of Pol I are removal of RNA primers in lagging strand DNA synthesis and repair of damaged DNA (Voet, et al., 1999).

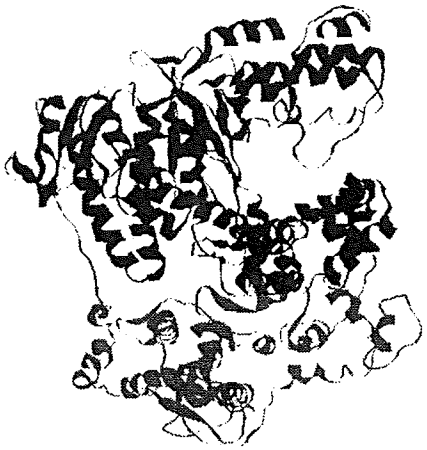
The large fragment of Pol I is also called the "Klenow fragment" or just "Klenow" (MW=68,094). Although there is currently no crystal structure of full-length Pol I, Klenow was the first polymerase to have its crystal structure solved. Klenow is folded into two domains, a small domain (which contains the 3'-5'exonuclease activity) and a large domain (which contains the polymerase active site). The two active sites are separated by about 35Å in the crystal structure. The deep cleft of the large domain is the active site of the polymerase domain and has a size and shape that accommodates B-DNA (Ollis, et al., 1985; Beese, et al., 1993).

Thermus aquaticus DNA Polymerase I and the Klentaq fragment

Thermus aquaticus DNA Polymerase I (Taq) was isolated and purified in 1976 (Chien, et al., 1976). It was determined to have a pH optimum of 7-8 and an optimum functional temperature of 75°C. Taq is even stable at temperatures up to 93-95°C, making it extremely useful in PCR experiments (Lawyer, et al., 1989). Taq contains the

carboxy terminus polymerase domain and the amino terminus 5' nuclease domain, but has no 3'-5' exonuclease activity (Lawyer, et al., 1989; Chien, et al., 1976). Although the 3'-5' exonuclease activity is missing, remnants of this domain are present, as shown in crystal structure comparisons of Taq and Klenow. However, the vestigal domain is missing the residues present in Klenow that bind the two metal ions (which are different than the two metal ions which facilitate the polymerase reaction) and which catalyze the 3'-5' exonuclease reaction (Kim, et al., 1995). Taq is homologous to Pol I. The two enzymes have an overall 38% amino acid sequence similarity (Lawyer, et al., 1989).

Four crystal structures of full-length Taq have been reported. Two structures are of the enzyme alone (Kim, et al., 1995; Urs, 1999). One structure is of Taq with a piece of double-stranded DNA bound to the cleft of the polymerase active site (Eom, et al., 1996). The fourth is of Taq with an inhibitory antibody bound to the polymerase active site (Murali, et al., 1998). Comparisons of the overall shape of Taq in these four structures raise questions about the position of the 5' nuclease domain (Figure 2). In two of the structures the Taq molecule is elongated with the active sites of the polymerase and 5' nuclease domains about 70Å apart (Kim, et al., 1995; Eom, et al., 1996). However, in the other two structures, the 5' nuclease domain is closer in, reducing the distance between the active sites to about 38Å (Murali, et al., 1998; Urs, 1999). Also, preliminary, unpublished radius of gyration data measured using X-ray scattering report that the molecule is more compact and not elongated (Kim, et al., 1995). These conflicting reports are a basis for this project's attempt to clarify the position of the 5' nuclease domain of Taq using analytical ultracentrifugation.



1CMW

1TAQ

Figure 2. Two orientations of the 5' nuclease domain of Taq. In the 1CMW structure, the 5' nuclease domain is folded in, resulting in a more compact, spherical structure. In the 1TAQ structure, the domain is extended, producing a more elongated particle.

Klentaq is the shortened version of Taq. The amino terminus is deleted, leaving only the polymerase domain and the vestigal 3'-5' exonuclease domain. The large domains of Klentaq and Klenow have 49.6% sequence homology and extremely similar folds (Korolev, et al., 1995). Comparisons with Klenow have pointed out amino acid substitutions that may contribute to Klentaq's enhanced stability. For example, Klentaq exhibits an increased number of hydrophobic residues in its buried surface areas as compared to Klenow. Klentaq also has several substitutions of proline for alanine, as well as a large number of substitutions to amino acids of the opposite charge, leading to an overall change in the charge distribution of the protein. These substitutions have been shown to enhance stability in other protein systems and are sometimes used to explain thermostability (Korolev, et al., 1995).

In the present study, Klentaq and Klenow are used as comparative controls. Since they do not have a 5' nuclease domain, they should act like more compact, spherical molecules.

Although Taq and Klentaq show many structural and functional similarities with Pol I and Klenow, the enzymes function in very different environments. While *E. coli*'s optimum temperature is 37°C, *T. aquaticus*' is 75°C. This makes study of the similarities and differences among these proteins intriguing. A better understanding of the differences in structure and organization of these enzymes will, therefore, lead to an explanation for their temperature differences and an overall better characterization of the proteins.

Analytical Ultracentrifugation

One way to study the solution structures of the two sets of polymerases is by using analytical ultracentrifugation. Equations relating the behavior of macromolecules in the ultracentrifuge to their physical properties are used to obtain information from analytical ultracentrifugation.

The primary piece of information obtained from analytical ultracentrifugation is the sedimentation coefficient. The sedimentation coefficient (s) is defined as the velocity divided by the strength of the centrifugal field and is related to other molecular properties as shown in Equation 1 (Cantor and Schimmel, 1980; van Holde, et al., 1998):

$$s = \frac{v}{\omega^2 r} = \frac{M(1-\bar{v}\rho)}{N_o f} \qquad \textit{Equation 1}$$

M=molecular weight
ρ=density of the solution
f=frictional coefficient of the protein
ω=angular velocity=rpm(2π/60)
$\bar{v}$=partial specific volume of the protein
N_o=Avogadro's number
v=velocity of the protein in the centrifuge
r=distance from center of rotation at any time

Using the Einstein-Sutherland equation and making the appropriate substitutions will rearrange this equation into what is known as the Svedberg equation:

$$D_o = (k_B T)/f \qquad \textit{Equation 2. Einstein-Sutherland equation}$$

D_o=diffusion coefficient of the protein
f=frictional coefficient of the protein
k_B=Boltzman constant
T=temperature

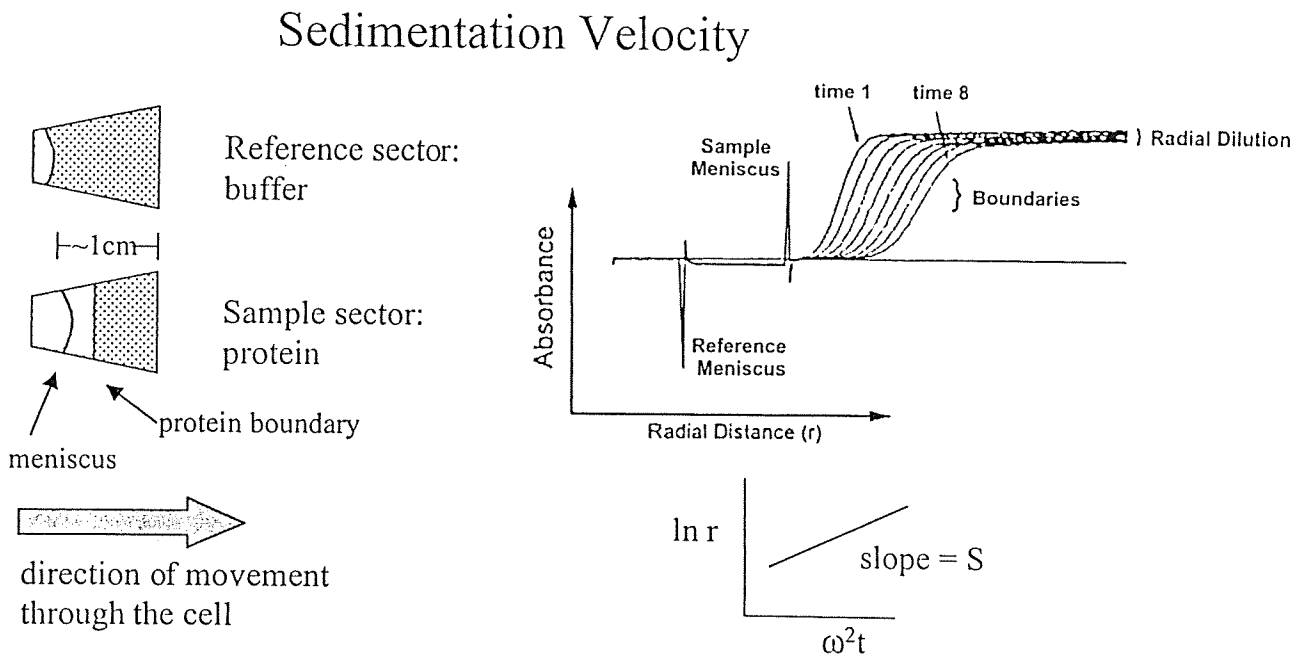
$$M = \frac{sRT}{D_o(1-\bar{v}\rho)} \qquad \textit{Equation 3. Svedberg equation}$$

M=molecular weight
D_o=diffusion coefficient of the protein
T=temperature
ρ=density of the solution
s=sedimentation coefficient
R=gas constant
$\bar{v}$=partial specific volume of the protein

Values of the sedimentation coefficient (also known as S values) are measured in Svedbergs, where 1 Svedberg=10^{-13} seconds. Since the equation relates S values to molecular weight, the buoyancy factor ($1-\bar{v}\rho$), and the diffusion coefficient, information on the size and shape of macromolecules can be obtained from analytical ultracentrifugation studies. The buoyancy factor relates the density of the protein to the density of the solvent. Proteins are more dense than aqueous solvents. The partial specific volume parameter in the buoyancy factor contains contributions from the packing density of the protein and from the hydration of the protein. The diffusion coefficient of the protein contains contributions from the size and shape of the hydrated protein (Cantor and Schimmel, 1980; van Holde, et al., 1998).

There are two major types of experiments performed in the analytical ultracentrifuge: sedimentation velocity and sedimentation equilibrium. In **sedimentation velocity** experiments, a homogeneous solution is placed into the ultracentrifuge and is spun at a particular angular velocity. When the centrifugal force is applied, the molecules begin to move towards the bottom of the cell, resulting in a region at the top of the cell in which there are no particles and only pure solvent remains. The moving boundary between the protein solution and the solvent is called the protein boundary. The movement of the protein boundary down the centrifuge cell with time can be tracked by the ultracentrifuge's spectrophotometer as it scans the length of the cell. Scans of absorbance versus radius yield the position of the boundary versus time (Figure 3), which is then used to calculate the protein's sedimentation coefficient. The velocity (v) of sedimentation is equal to the change in the boundary position with time. A plot of

Figure 3. Overview of the general steps in a sedimentation velocity experiment.



Equation: $s\ \omega^2 = \frac{dr/r}{dt} = \frac{d \ln r}{dt}$

s = sedimentation coefficient

ω = angular velocity of the rotor in radians per second

dr = change in radial position

r = radial distance from the center of rotation (sometimes normalized as "r/rm")

dt = change in time

the logarithm of the boundary position as a function of time and the angular velocity results in a line with slope s (Cantor and Schimmel, 1980):

$$s = \frac{v}{\omega^2 r} = \frac{dr/dt}{\omega^2 r} \qquad s = \omega^2 [d(\ln r)/dt] \qquad \textit{Equations 4 and 5}$$

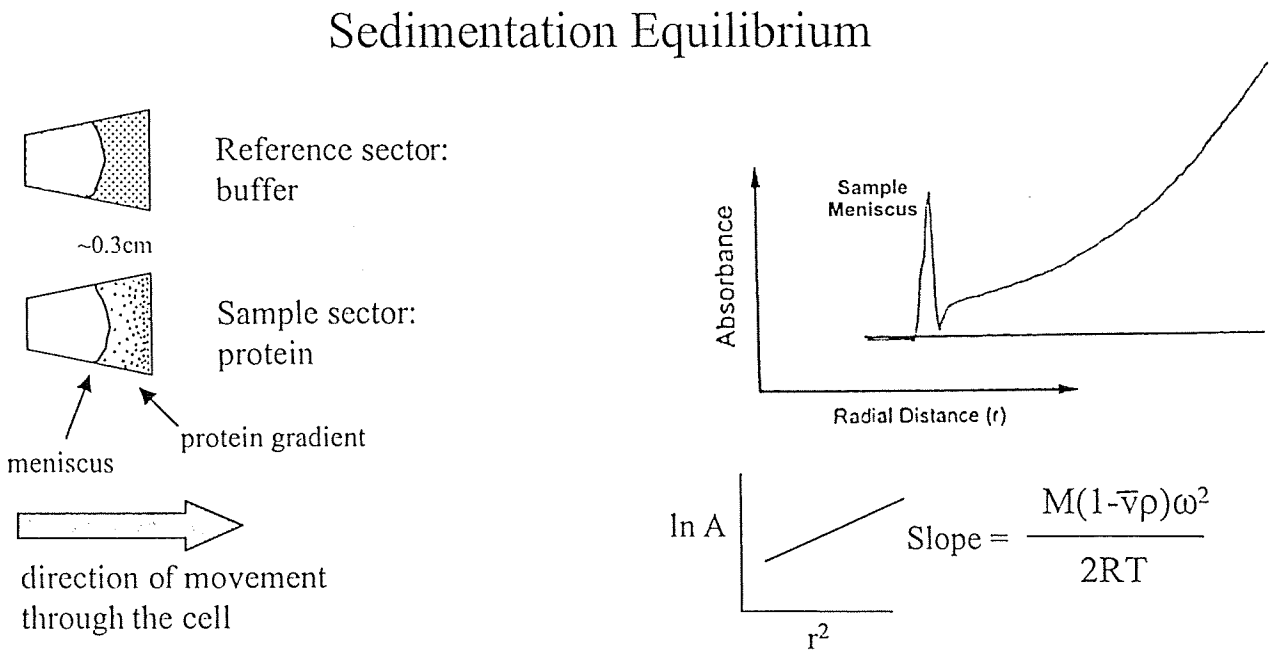
Changes in the S value with changes in temperature or other experimental conditions can, therefore, be used to monitor changes in protein conformation.

In **sedimentation equilibrium** experiments, the angular velocity of the ultracentrifuge is much lower, resulting in the formation of a stable distribution of particles from the top to bottom of the cell instead of a moving protein boundary (Figure 4). Again, the spectrophotometer scans the length of the cell and takes absorbance readings. At some time after the start of centrifugation (generally 14-50 hours), the solution will reach equilibrium and the particle distribution gradient will no longer change with time. The resulting curve (see Results, Figures 7-10) can be used to determine the molecular weight of a protein, aggregation states, and binding stoichiometries. In a solution with only one species, the absorbance increases exponentially with the square of the distance from the center of rotation (r). A graph of the logarithm of absorbance versus r^2 yields a straight line with slope $[M(1-\bar{v}\rho)\omega^2]/2RT$ (Cantor and Schimmel, 1980):

$$\frac{d \ln A}{d r^2} = \frac{M(1-\bar{v}\rho)\omega^2}{2RT} \qquad \textit{Equation 6}$$

Since the equations used to relate the protein's S value to other hydrodynamic factors include properties (such as density and partial specific volume) that depend on the temperature and characteristics of the solution in which the particle is dissolved, corrections for solution properties must be made. S values are conventionally converted

Figure 4. Overview of the general steps in a sedimentation equilibrium experiment.



Fitting Equation: $A_r = A_o[HM(x^2-x_o^2)] + E$

A_r = absorbance at radius x

A_o = absorbance at reference radius x_o

$H = (1-\bar{v}\rho)\omega^2 / (2RT)$

E = baseline offset

$\bar{v}$ = partial specific volume

ρ = density

ω = angular velocity of rotor

R = gas constant

T = temperature

M= molecular weight

to $S_{20,w}$ values, which is the value of the sedimentation coefficient if the run had taken place at 20°C and in water. Measurements of the solution's density and viscosity are used to correct to these standardized values:

$$S_{20,w} = S_T \frac{\eta_T}{\eta_{20,w}} \frac{(1-\bar{v}_{20,w}\rho_{20,w})}{(1-\bar{v}_T\rho_T)} \qquad \textit{Equation 7}$$

S_T=sedimentation coefficient at temperature T
$\bar{v}_{20,w}$=partial specific volume of particle in water at 20°C
$\rho_{20,w}$=density of water at 20°C
η_T=viscosity of solution at T
$\eta_{20,w}$=viscosity of water at 20°C
$\bar{v}_T$=partial specific volume of particle in buffer at T
ρ_T=density of solution at T

Standardized $S_{20,w}$ values can then be compared without having to take into account the solution conditions in which the experiment was run (Svedberg, et al., 1959; Cantor and Schimmel, 1980; van Holde, et al., 1998).

HYDROPRO—Predicting hydrodynamic properties

HYDROPRO is a computer program that calculates hydrodynamic properties (for example, the sedimentation coefficient, radius of gyration, and the diffusion coefficient) from the atomic coordinates of a macromolecule using bead modeling (de la Torre, et al., 1994).

The protein's atomic coordinates, usually imported from the Protein Data Bank, and conditions of the solution are entered. The program first represents each nonhydrogen atom by spheres of the size of their covalent radii. The spheres are then expanded to a value of radius *a*, denoted the "atomic element radius" (AER). The AER is an adjustable parameter entered by the user. It is assumed to take into account hydration effects (the hydration layer) in order to give a more accurate picture of the size and shape of a protein in solution. The resulting primary hydrodynamic particle (PHP) is then filled with closely-packed beads, yielding the "filling model." The filling model is used to

calculate the volume and the radius of gyration of the molecule. Next the internal beads are removed, leaving the "rough shell model." The rough shell model represents the surface of the protein with beads of radius σ and is then used to calculate the remaining hydrodynamic properties. Extrapolation of calculated values of hydrodynamic properties resulting from the use of different σ values are reported by the program (de la Torre, et al., 2000).

Materials and Methods

Isolation and Purification of Polymerases

Taq, Klentaq, Pol I, and Klenow were isolated and purified by members of Dr. LiCata's lab (Kausiki Datta, Carmen Ruiz, and Rena Karantzeni) using modifications of published procedures (Barnes, 1995; Engelke, et al., 1990; Joyce, et al., 1983; Minkley, et al., 1984).

All sedimentation velocity and sedimentation equilibrium runs were performed in a buffer containing 10mM Tris, 125mM KCl, and 5mM $MgCl_2$, pH 7.9. Small centrifuged G-50 columns were used to exchange the buffer of all protein solutions, according to the method of Penefsky (Penefsky, 1977). All proteins were concentrated with Centricon YM-10 centrifugal filter devices (Amicon) to an absorbance of around 0.35 at 280nm.

Sedimentation Velocity Runs

Sedimentation velocity runs were performed in the Beckman XL-A Optima analytical ultracentrifuge using Epon charcoal-filled double-sector cells and the An-60 Ti rotor (Figure 3). 425µl of buffer was placed in the reference sector, and 400µl of protein was placed into the sample sector. All runs were performed at 38,000rpm for 3.5 hours, and absorbance scans were taken at 280nm.

A sedimentation velocity run was performed for each protein at 4°C, 20°C, and 40°C. 40°C was the highest temperature used since it is the maximum temperature at which the XL-A can run. The 20°C runs for each polymerase were performed first. Since Pol I and Klenow are not heat stable, their 40°C runs were performed last. Since

Taq and Klentaq are heat stable, their 40°C runs were performed second, and their 4°C runs were performed last. 20°C runs were then repeated following the extreme temperature runs in order to check the proteins' stability and the reproducibility of the data.

Runs were analyzed using Origin Velocity software provided by Beckman. Each sedimentation coefficient determined from the sedimentation velocity runs was converted to a $S_{20,w}$ value using Equation 7.

Sedimentation Equilibrium Runs

Sedimentation equilibrium runs were performed in the Beckman XL-A Optima analytical ultracentrifuge using Epon charcoal-filled double-sector cells and the An-60 Ti rotor (Figure 4). 125µl of buffer was placed in the reference sector, and 110µl of protein was placed into the sample sector. All runs were performed at 20°C for approximately 22 hours (or until equilibrium was reached), and absorbance scans were taken at 280nm. A sedimentation equilibrium run was performed for each protein. Pol I and Taq equilibrium runs were performed at 9500rpm, and Klenow and Klentaq equilibrium runs were performed at 11,000rpm.

Runs were analyzed using Origin Equilibrium software provided by Beckman.

HYDROPRO

The HYDROPRO computer program was downloaded from the University of Murcia Macromolecules Group website (http://leonardo.fcu.um.es/macromol). PDB files were imported into the program, and other required information (such as molecular weight) was entered and calculations performed as described in de la Torre, et al., 2000.

Sedimentation coefficients obtained from HYDROPRO were compared to actual values obtained in the ultracentrifuge.

Density Measurements

Density measurements of the buffer at 4°C, 20°C, and 40°C were taken in the Anton-Paar DMA 58 Digital Densitometer. The densitometer contains an oscillating U-shaped glass capillary tube and measures the period of oscillation in order to determine the density.

Viscosity Measurements

Viscosity measurements were made using a calibrated Cannon-manning Semi-micro kinematic viscometer. The efflux time (the time it takes the solution to fall a set distance through the glass tube) was converted to viscosity in Poise by using the viscometer constant and the density of the solution. Viscosity measurements of the buffer at each temperature were taken.

efflux time (s) * viscometer constant (mm^2s^2) *

density (g/cm^3) = viscosity in centiPoise *Equation 8*

Determination of Partial Specific Volumes

In this study, partial specific volumes were obtained in two ways: 1) by changing the value of the partial specific volume during equilibrium run analysis until the correct molecular weight of the proteins were obtained; and 2) using a computer program (SEDNTERP, http://biochem.uthscsa.edu/auc/software) which calculates the partial specific volume based on amino acid sequence.

Results

Uncorrected Sedimentation Coefficients

Sedimentation coefficients were obtained for Klenow, Pol I, Klentaq, and Taq at 4°C, 20°C, and 40°C (Table 1). Figure 5 shows an example of sedimentation velocity data. The changes in the S values with temperature were similar for all four proteins (Figure 6).

Table 1. Measured, uncorrected sedimentation coefficients at 4°C, 20°C, and 40°C.

Polymerase (Molecular weight)	S, 4°C	S, 20°C	S, 40°C
Klenow (68,094)	2.82	4.40	6.69
Pol I (103,118)	3.00	4.77	7.14
Klentaq (61,155)	2.50	3.92	5.95
Taq (93,909)	3.08	4.70	7.13

Obtaining $\bar{v}$ data for correction of raw sedimentation coefficients

1) Calculations

Partial specific volumes ($\bar{v}$ or vbar) for each polymerase were predicted using their amino acid sequences. The partial specific volume is the volume per unit mass of the particle (for example, ml/g). A computer program (SEDNTERP, http://biochem.uthscsa.edu/auc/software) calculated the partial specific volumes using the PDB files (Table 2). The program sums the partial specific volumes of the individual residues to theoretically calculate the vbar:

$$\bar{v} = \frac{\sum n_i M_i \bar{v}_i}{\sum n_i M_i} \qquad \textit{Equation 9}$$

$\bar{v}_i$ = partial specific volume of the ith component
n_i = number of moles of the ith component
M_i = molecular weight of the ith component

2) Sedimentation equilibrium runs

Equilibrium runs of each protein (Figures 7-10) at 20°C were also used to obtain partial specific volumes since the actual molecular weights are known. The partial

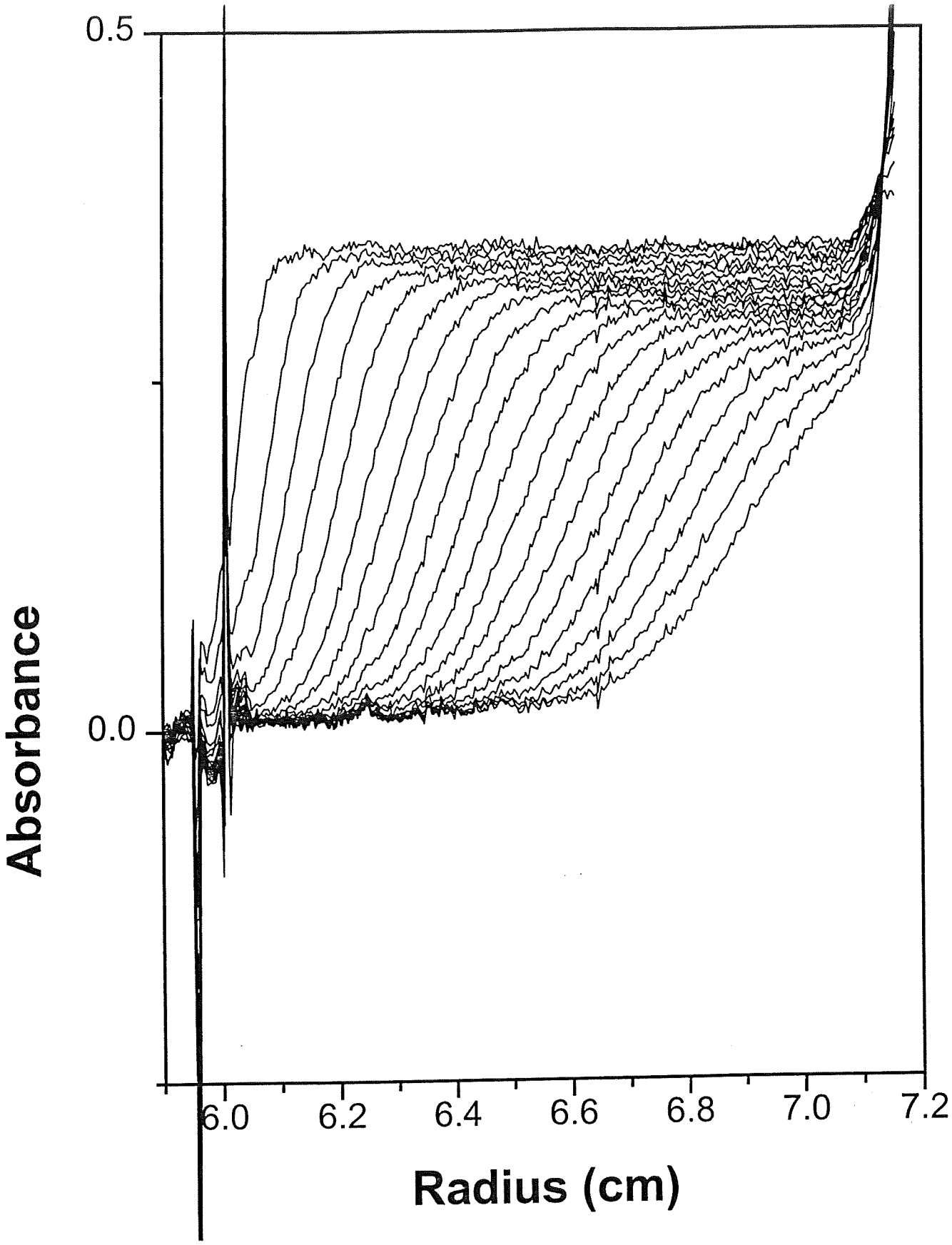


Figure 5. Representative sedimentation velocity data. The twenty absorbance scans taken during a sedimentation velocity run of Taq at 40°C are shown. The sedimentation coefficient obtained from this run is 7.13 S.

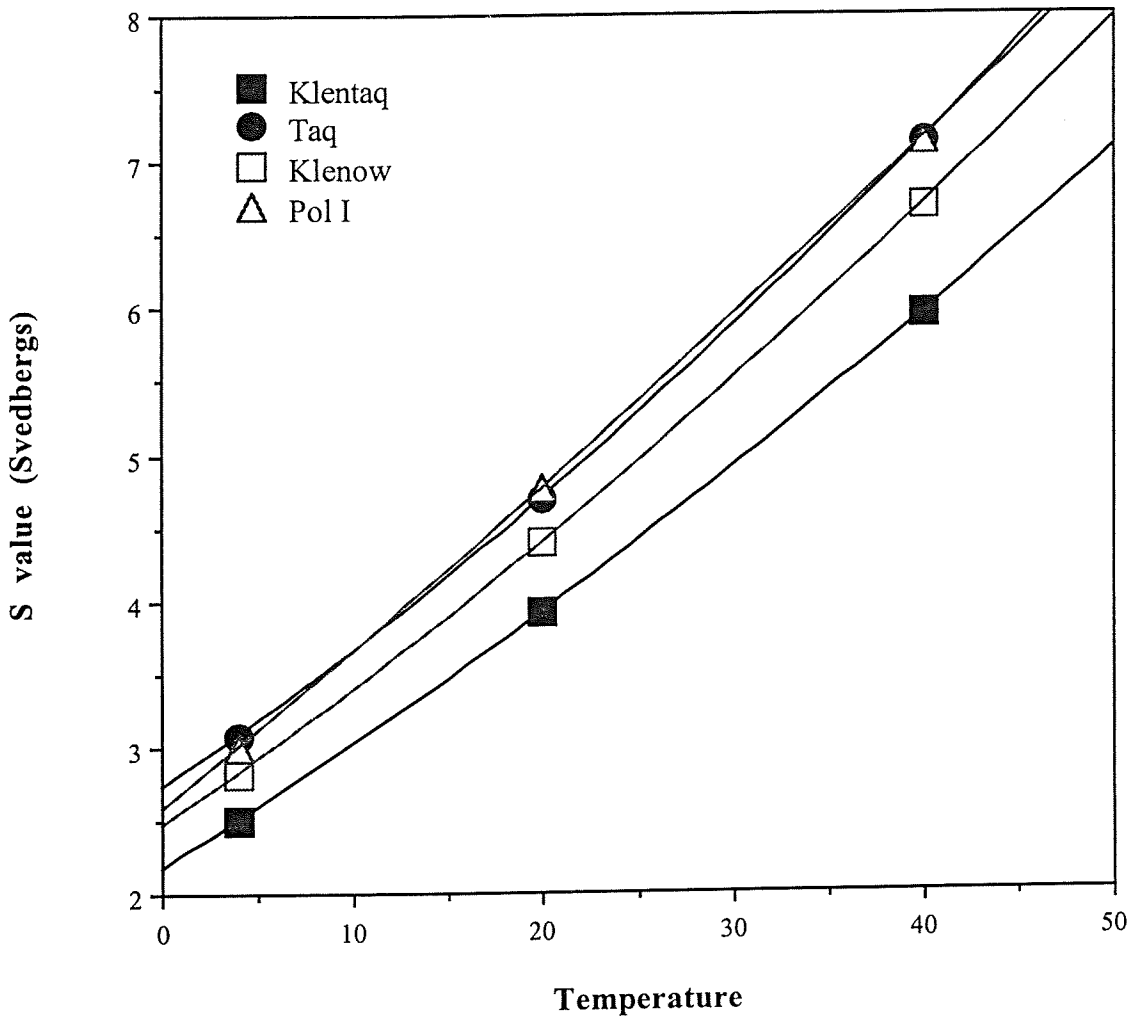


Figure 6. Raw sedimentation coefficients vs. temperature. Directly obtained, uncorrected sedimentation coefficients are shown vs. temperature for the four proteins examined. Lines are polynomial fits for visual clarity only.

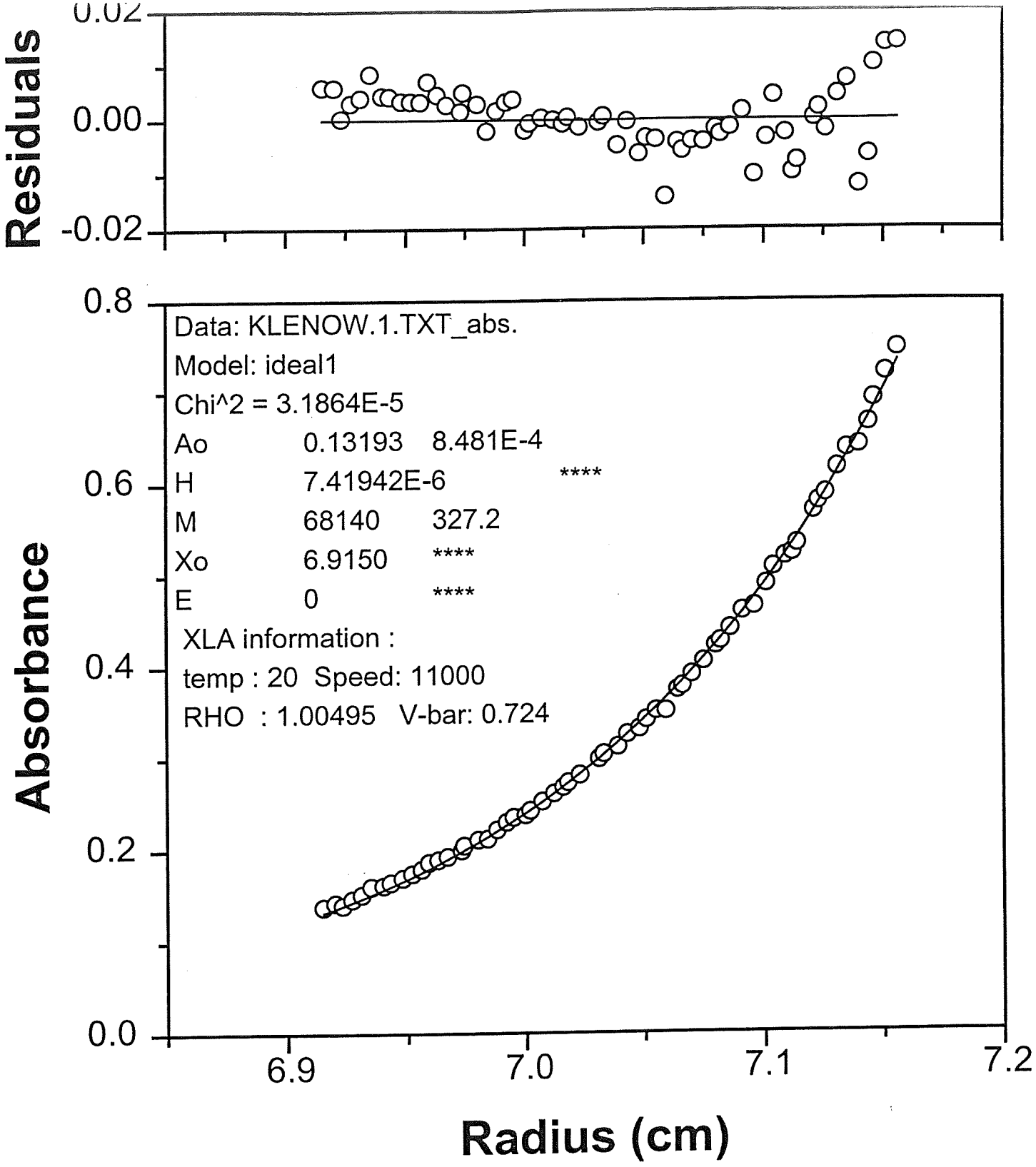


Figure 7. Klenow sedimentation equilibrium run. The bottom panel shows the equilibrium curve fit to a one species fitting equation. The vbar was altered until the correct molecular weight was obtained. The top panel is a plot of the residuals vs. radius. Random scatter in the residuals plot indicates a good fit. Patterns or systematic distributions of the residuals may indicate the presence of more than one species.

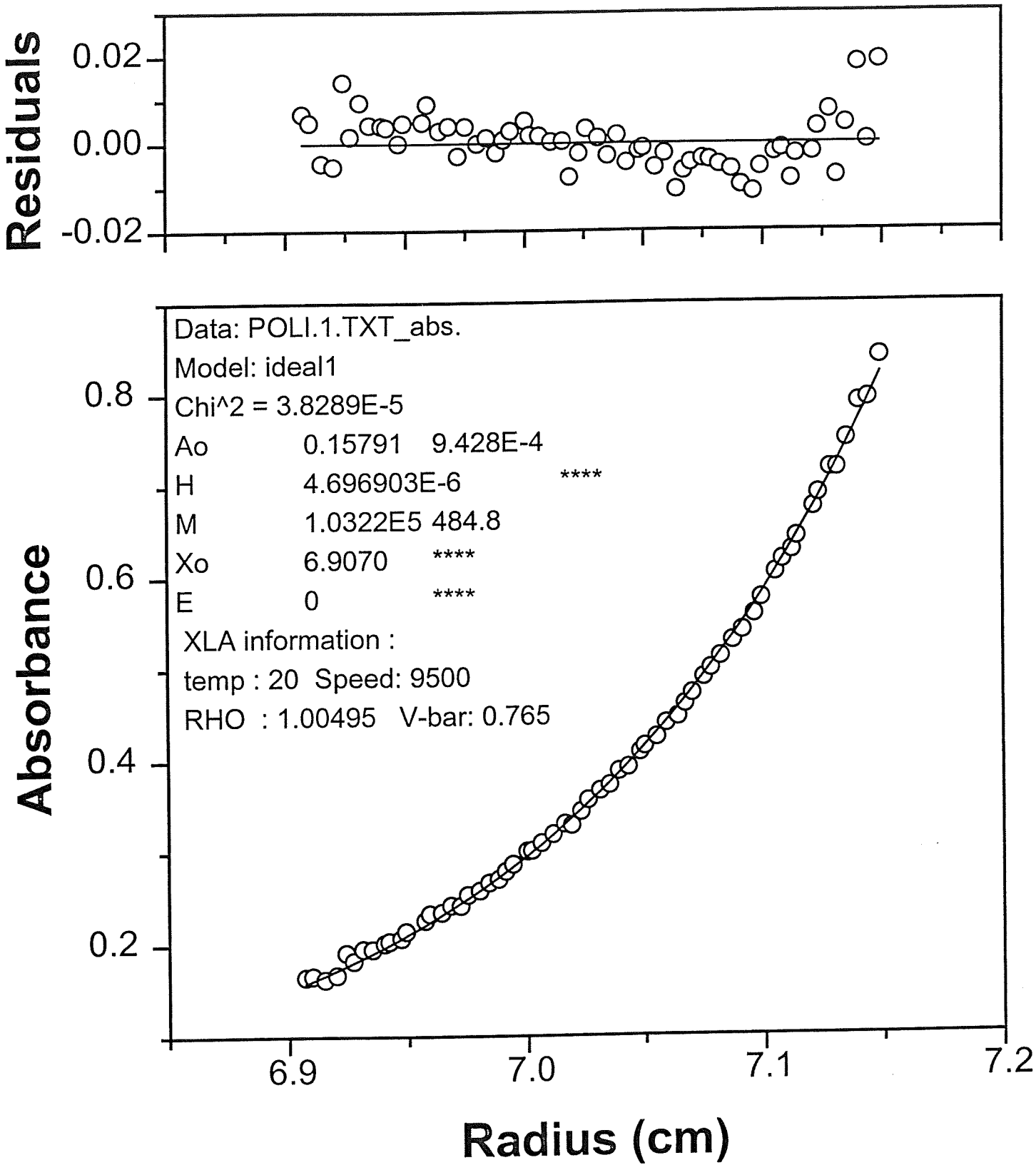


Figure 8. Pol I sedimentation equilibrium run. The bottom panel shows the equilibrium curve fit to a one species fitting equation. The vbar was altered until the correct molecular weight was obtained. The top panel is a plot of the residuals vs. radius.

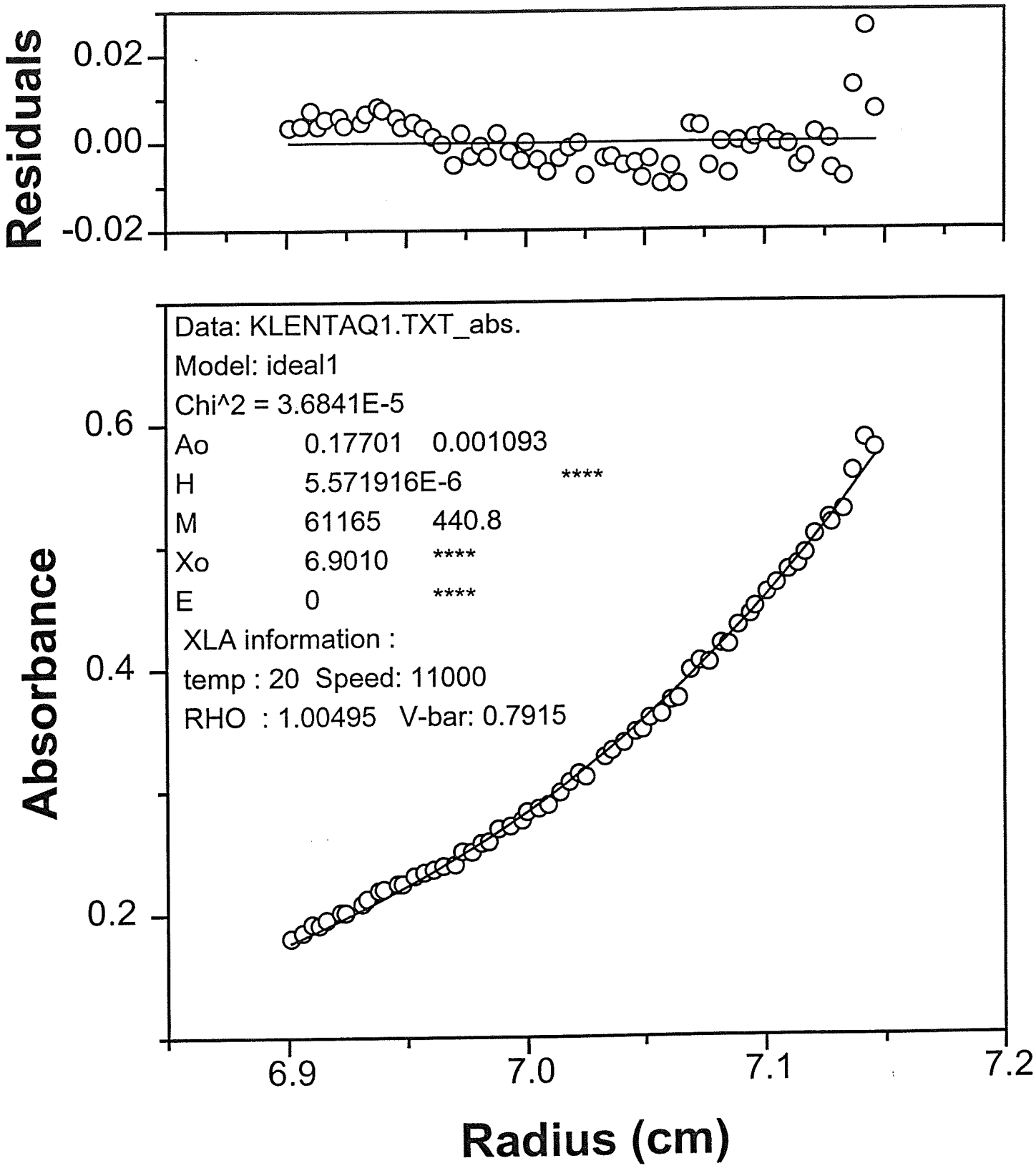


Figure 9. Klentaq sedimentation equilibrium run. The bottom panel shows the equilibrium curve fit to a one species fitting equation. The vbar was altered until the correct molecular weight was obtained. The top panel is a plot of the residuals vs. radius.

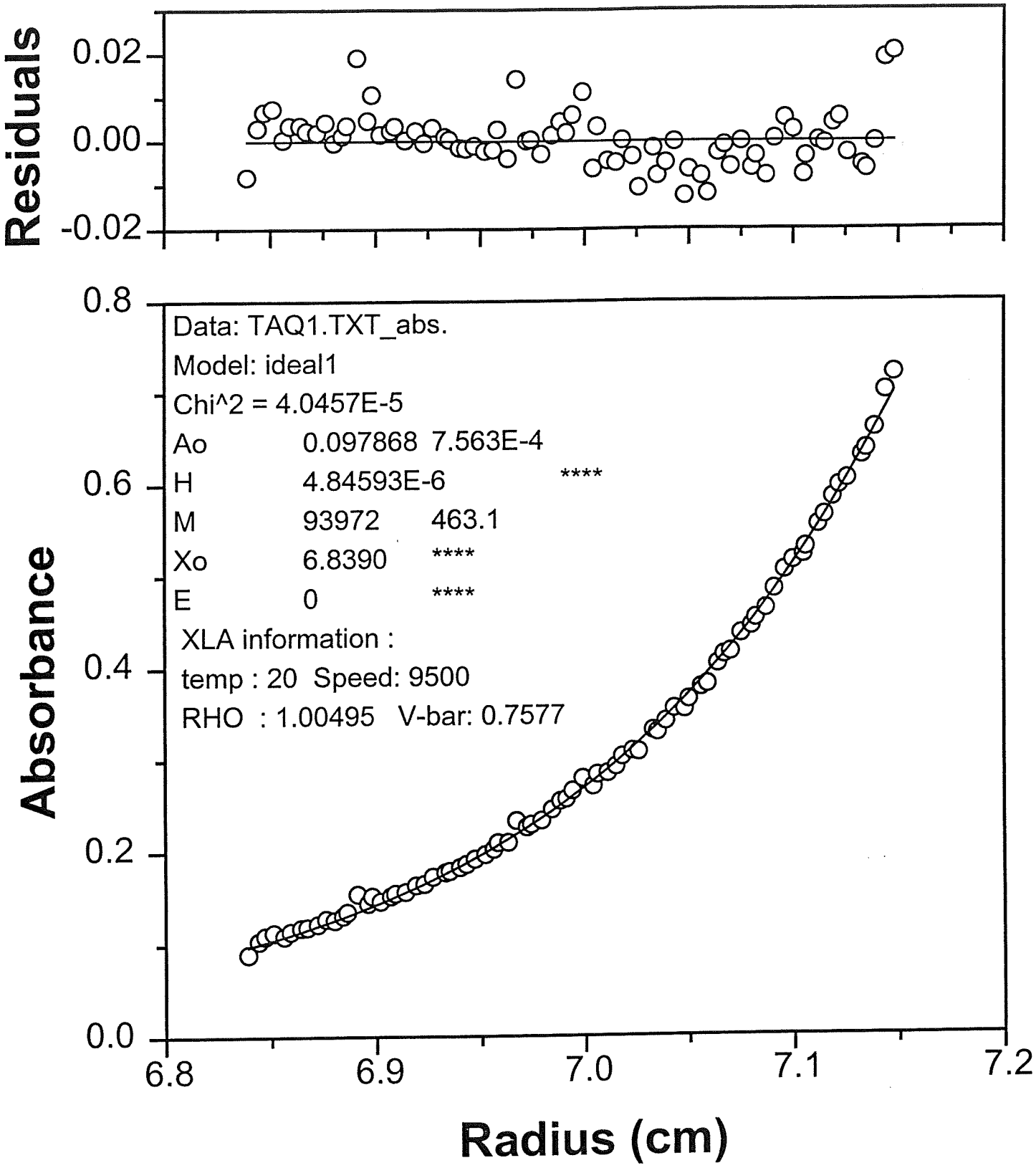


Figure 10. Taq sedimentation equilibrium run. The bottom panel shows the equilibrium curve fit to a one species fitting equation. The vbar was altered until the correct molecular weight was obtained. The top panel is a plot of the residuals vs. radius.

specific volume was altered until the equilibrium run data resulted in the correct molecular weight (Table 2).

Table 2. Partial specific volumes (ml/g).

Polymerase	Predicted partial specific volume from amino acid sequence	Partial specific volume using known molecular weight and equilibrium run
Klenow	0.7430	0.7240
Pol I	0.7417	0.7650
Klentaq	0.7442	0.7915
Taq	0.7438	0.7577

Equilibrium runs also demonstrated that higher order oligomers of the proteins are not forming. Correct molecular weights were obtained from fitting the curves to a one species fitting equation using reasonable values of vbar. If oligomers were forming, the molecular weights obtained would reflect the molecular weight of the oligomer, or the data would have fit to the equations for a solution of several species and not to the one species fit.

Equilibrium data also demonstrated that Pol I begins to degrade after about one week of storage. Several equilibrium runs performed on Pol I samples greater than one week old resulted in data that would not fit to the one species fit. Instead, fits of this data identified two species of molecular weights of approximately 95,000 and 20,000.

Correction to $S_{20,w}$ values

Sedimentation coefficients were converted to $S_{20,w}$ values (Table 4, Figures 11-12). The measured density and viscosity at each run condition (Table 3) were used, as well as the partial specific volumes determined from the equilibrium runs and the partial specific volumes from the protein sequences (Table 2). $S_{20,w}$ values calculated using the predicted vbars were similar to $S_{20,w}$ values calculated using vbars obtained from equilibrium runs.

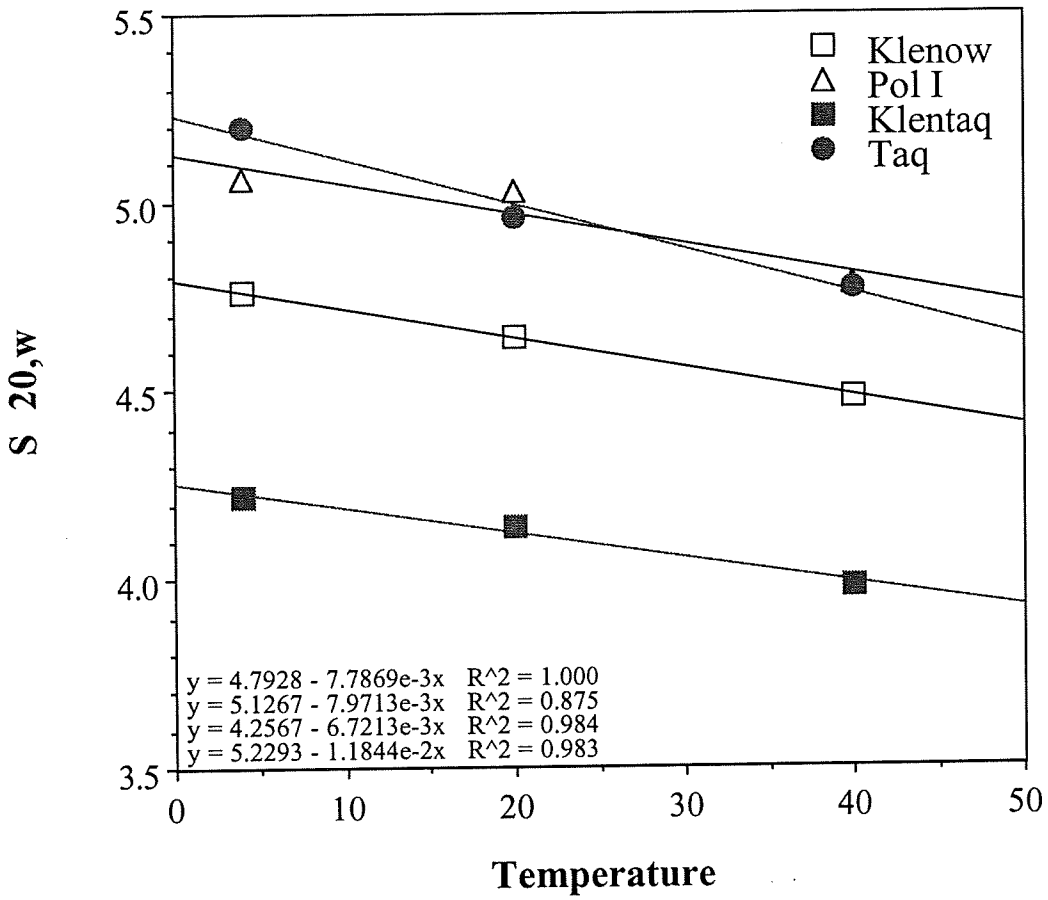


Figure 11. $S_{20,w}$ values vs. temperature. S values corrected using the vbars predicted from the protein sequences are shown for all proteins. Fits to the equation for a line are also shown in the same order (top to bottom) as the proteins are listed in the legend.

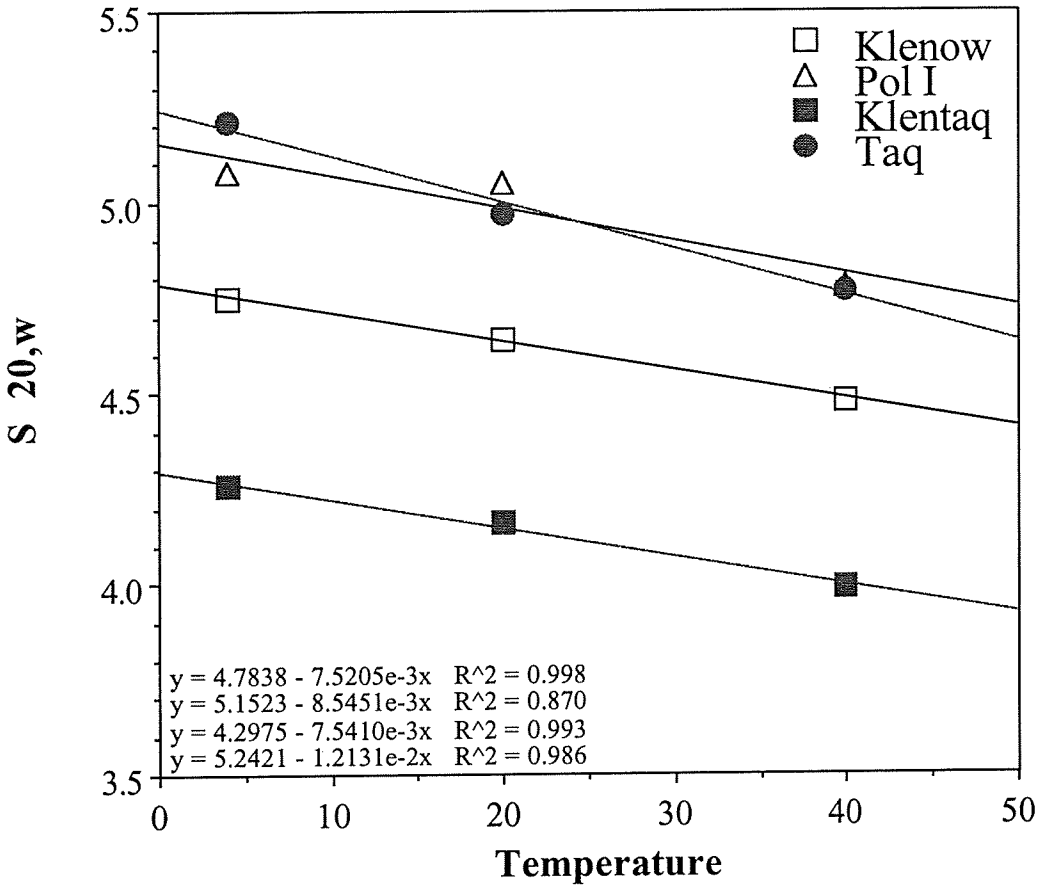


Figure 12. $S_{20,w}$ values vs. temperature. S values corrected using the vbars obtained from the equilibrium runs are shown for all proteins. Fits to the equation for a line are also shown in the same order (top to bottom) as the proteins are listed in the legend.

Table 3. Densities and viscosities of solution conditions.

	4°C	20°C	40°C
Density (g/cm^3)	1.00699	1.00495	0.99876
Viscosity (Poise)	0.01649	0.01037	0.00670

Table 4. Sedimentation coefficients converted to $S_{20,w}$ values using vbars predicted from protein sequences and vbars from the equilibrium runs. (Data plotted in Figures 11-12)

	Using predicted vbar			Using vbar from equilibrium runs		
Polymerase	$S_{20,w}$ 4°C	$S_{20,w}$ 20°C	$S_{20,w}$ 40°C	$S_{20,w}$ 4°C	$S_{20,w}$ 20°C	$S_{20,w}$ 40°C
Klenow	4.76	4.64	4.48	4.75	4.64	4.48
Pol I	5.06	5.03	4.78	5.08	5.05	4.78
Klentaq	4.22	4.14	3.98	4.26	4.16	3.99
Taq	5.20	4.96	4.77	5.21	4.97	4.77

$S_{20,w}$ values from the 20°C sedimentation velocity runs of the polymerases were also compared to values of sedimentation coefficients of other proteins of known shape and molecular weight (Figure 13). A plot of $(s\bar{v}^{1/3})/(1-\bar{v}\rho)$ vs. molecular weight of unhydrated, spherical molecules yields a straight line with a slope of 2/3 (van Holde, et al., 1998). Globular proteins fall on a line parallel to this line since they are close to being spherical; however, asymmetric proteins fall below the line (van Holde, et al., 1998). Another way to visualize the shapes of the polymerases is to plot the measured S values vs. molecular weight and compare them to the S values of spherical molecules of the same molecular weight (Figure 14). This plot shows that the S values of Taq and Pol I are lower than expected for spherical molecules.

Calculation of S values from structural data

HYDROPRO was used to predict the sedimentation coefficients of Klentaq, Klenow, and Taq from their PDB files. Information for Pol I could not be predicted since it does not have a solved crystal structure. The molecular weight, partial specific volume, temperature, solvent density, and solvent viscosity were entered parameters. For each protein, the S value was calculated using the predicted partial specific volume from the

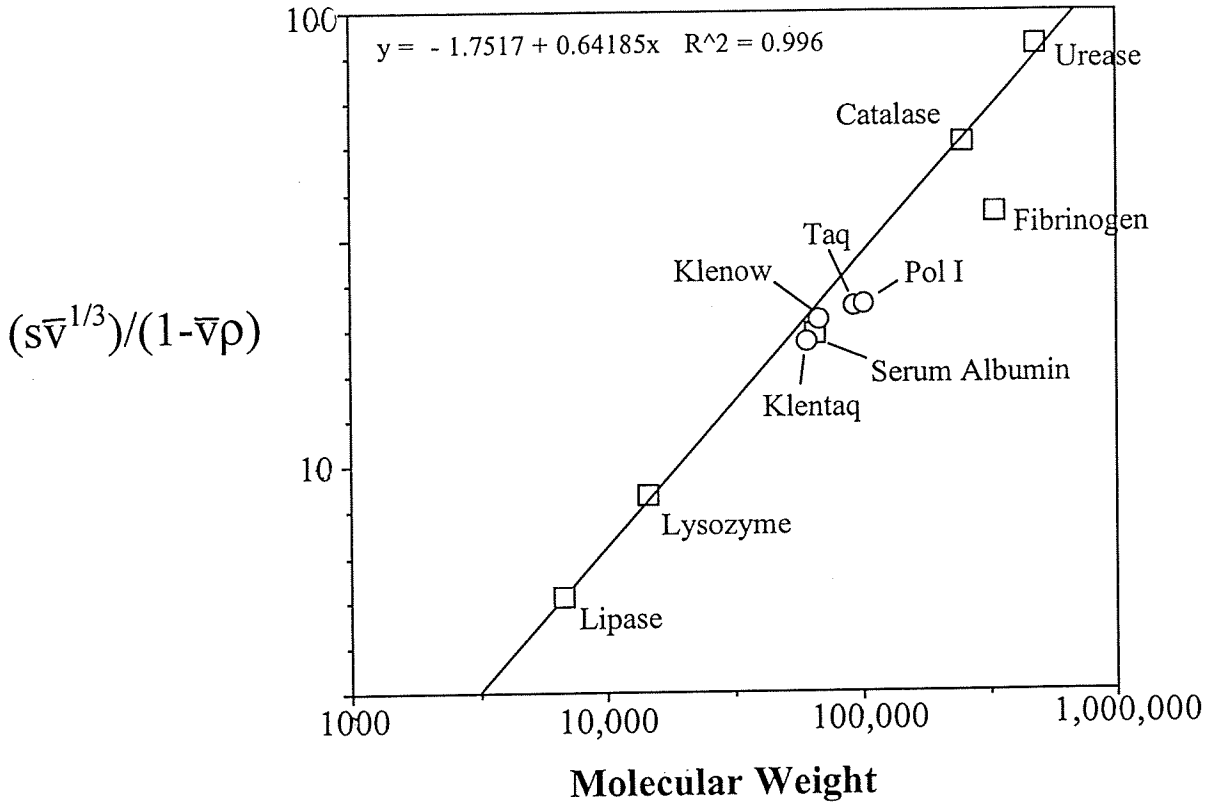


Figure 13. Polymerase $S_{20,w}$ data compared to that of other proteins. Globular proteins fall on or near the line, which is parallel to a line for unhydrated, perfectly spherical molecules (not shown). Asymmetric proteins (for example, fibrinogen) fall far below the line. Klentaq and Klenow are near the globuar protein line. Taq and Pol I are farther from the line; however, they are not as far away as highly asymmetric proteins like fibrinogen.

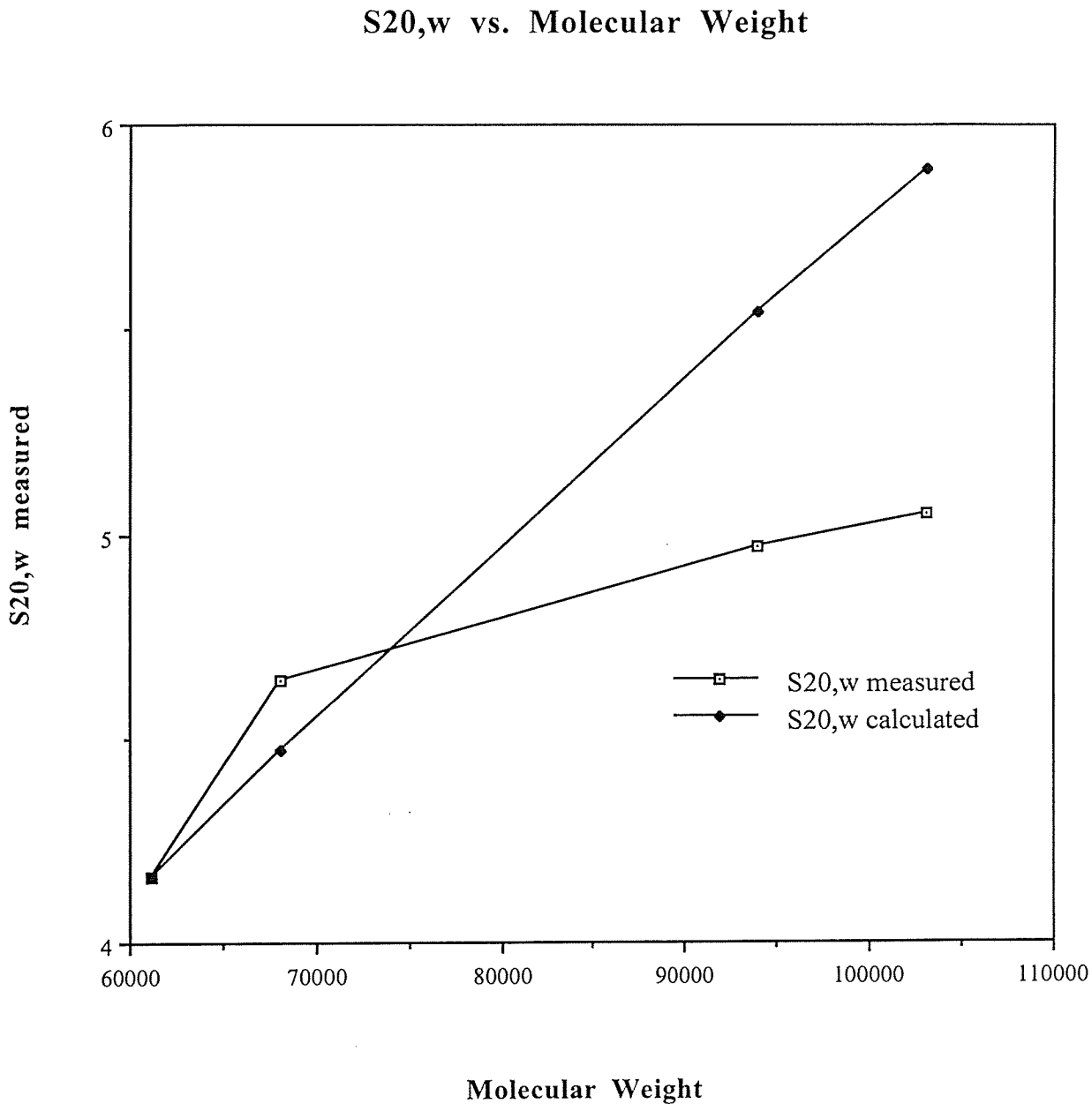


Figure 14. 20°C $S_{20,w}$ values vs. molecular weight. Measured $S_{20,w}$ values vs. molecular weight are shown. $S_{20,w}$ values of spherical, unhydrated molecules of the same molecular weights as the four polymerases are also plotted. Taq and Pol I's S values are lower than expected, indicating that they are more asymmetrical than Klentaq and Klenow.

amino acid sequence as well as the partial specific volume calculated from the equilibrium runs. Three different values of AER were used: 3.1, 4.0, and 5.0. Also, four sigmas (radii of beads in the shell) were used for the extrapolation. The sigma range used was below the value of the AER in each case. This approach and these parameter choices are based on the published approach used by de la Torre and associates for a large set of globular proteins (de la Torre et al., 2000). Results of predicted S values for Klentaq, Klenow, and the two orientations of Taq are reported in Tables 5-8. Calculations were also performed for Klentaq coordinates extracted from the two full-length Taq PDB files to determine if the specifics of the crystal structure (such as space group and packing) effected the calculated values (Tables 9-10). In each table the calculated S that most agrees with the $S_{20,w}$ value (calculated using the vbar from the equilibrium run) is in bold, and the percent difference between the two is shown.

In general the predicted, calculated S values show poor correlation with measured $S_{20,w}$ values. The best agreement (0.9%) is between the compact Taq structure and the measured $S_{20,w}$ of Taq. This agreement contradicts the solution structure for full length Taq predicted from analysis of the measured $S_{20,w}$'s alone. These calculations do show, however, show that the compact Taq structure should have a higher S value than the elongated structure.

Table 5. Klenow HYDROPRO predictions
Measured, uncorrected S value, 20°C = 4.40; $S_{20,w}$ = **4.64**

AER	Sigma range	S value using vbar predicted from amino acid sequence	S value using vbar calculated from equilibrium run
3.1	2.4-3.0	4.188	**4.454**
4.0	3.3-3.9	3.935	4.185
5.0	4.3-4.9	3.950	4.202

Difference between $S_{20,w}$ and closest calculated S: 4%

Table 6. Taq HYDROPRO predictions—Elongated structure (1TAQ)
Measured, uncorrected S value, 20°C = 4.70; $S_{20,w}$ = **4.97**

AER	Sigma range	S value using vbar predicted from amino acid sequence	S value using vbar calculated from equilibrium run
3.1	2.4-3.0	**4.680**	4.426
4.0	3.3-3.9	4.363	4.126
5.0	4.3-4.9	4.024	3.806

Difference between $S_{20,w}$ and closest calculated S: 5.8%

Table 7. Taq HYDROPRO predictions—Compact structure (1CMW)
Measured, uncorrected S value, 20°C = 4.70; $S_{20,w}$ = **4.97**

AER	Sigma range	S value using vbar predicted from amino acid sequence	S value using vbar calculated from equilibrium run
3.1	2.4-3.0	**5.014**	4.743
4.0	3.3-3.9	5.076	4.802
5.0	4.3-4.9	4.590	4.343

Difference between $S_{20,w}$ and closest calculated S: 0.9%

Table 8. Klentaq HYDROPRO predictions—Klentaq crystal structure (1KTQ)
Measured, uncorrected S value, 20°C = 3.92; $S_{20,w}$ = **4.16**

AER	Sigma range	S value using vbar predicted from amino acid sequence	S value using vbar calculated from equilibrium run
3.1	2.4-3.0	**3.767**	3.032
4.0	3.3-3.9	3.589	2.889
5.0	4.3-4.9	3.562	2.867

Difference between $S_{20,w}$ and closest calculated S: 9.4%

Table 9. Klentaq HYDROPRO predictions—Extracted from elongated Taq structure
Measured, uncorrected S value, 20°C = 3.92; $S_{20,w}$ = **4.16**

AER	Sigma range	S value using vbar predicted from amino acid sequence	S value using vbar calculated from equilibrium run
3.1	2.4-3.0	**3.817**	3.070
4.0	3.3-3.9	3.316	2.667
5.0	4.3-4.9	3.733	3.002

Difference between $S_{20,w}$ and closest calculated S: 8.2%

Table 10. Klentaq HYDROPRO predictions—Extracted from compact Taq structure
Measured, uncorrected S value, 20°C = 3.92; $S_{20,w}$ = **4.16**

AER	Sigma range	S value using vbar predicted from amino acid sequence	S value using vbar calculated from equilibrium run
3.1	2.4-3.0	3.817	3.071
4.0	3.3-3.9	3.606	2.901
5.0	4.3-4.9	**4.092**	3.292

Difference between $S_{20,w}$ and closest calculated S: 1.6%

Discussion and Conclusions

Temperature dependence of $S_{20,w}$

The sedimentation coefficients determined from sedimentation velocity experiments indicate that, overall, all of the polymerases experience similar changes in conformation with temperature. The $S_{20,w}$ values decrease slightly with an increase in temperature (Figures 11-12), indicating that the molecules are moving slower and are therefore becoming larger or less spherical. The slopes of the temperature dependence of sedimentation coefficients are all very similar except for Taq, which has a slightly steeper slope indicating its postulated expansion may be larger. However, with the small data set and the relatively large corrections necessary to obtain $S_{20,w}$ values, this slope difference is likely to be within error of the other proteins.

If the macromolecules were perfectly spherical, their sedimentation coefficients would be proportional their molecular weights raised to the 2/3 power (van Holde, 1998). The $S_{2o,w}$ data for all four proteins are below the line representing this relationship for unhydrated, spherical molecules, but the data for full length Taq and Pol I deviate further from the line than Klentaq and Klenow (Figure 13). Figure 13 is a double logarithmic plot, so small changes in x, y coordinates are reflective of larger changes in shape. This suggests that Taq and Pol I are more elongated in solution than Klentaq and Klenow. This data supports the orientation of Taq in the 1TAQ crystal structure in which the 5' nuclease is extended. Also, if the 5' nuclease domains of Taq and Pol I are extended, causing their sedimentation coefficients to be considerably lower, it would make sense that Klentaq and Klenow's S values would be closer to spherical since they do not contain the 5' nuclease domain. Figure 14 is a plot of S values of spheres the size of the

polymerases, as well as of the measured S values. This plot gives a better visualization of the deviation of Taq and Pol I from spherical.

HYDROPRO calculations of sedimentation coefficients

The HYDROPRO data produced conflicting results. For Taq, the predicted S values for the 1CMW structure are closer to the measured S and to the $S_{20,w}$ using both the calculated and the predicted vbars. This suggests that the compact orientation of the 5' nuclease domain is the correct one. However, the HYDROPRO data overall is not completely consistent. The predicted S values at different AER's are sometimes extremely different even using the same crystal structure and vbar. In order to better asses the results from HYDROPRO, the partial specific volumes should be measured another way in order to verify that they are correct. Also, another hydrodynamic property must be measured in order to determine which AER to use. According to the authors of HYDROPRO, the most accurate AER is determined by averaging the values of AER that give correct values for several properties (de la Torre, et al., 2000). Determining a useful AER will then better contribute to the determination of the orientation of the 5' nuclease domain.

Future plans

The conversion of the measured sedimentation coefficients to $S_{20,w}$ values requires knowledge of the protein's partial specific volume. So far, our only measurements of vbar were performed indirectly through equilibrium run data. Partial specific volumes will be measured using a more direct approach in order to ensure that the correct values are being used. One possible method is by taking density measurements at several concentrations of protein. The densities are then plotted versus

the protein concentration and the following relationship is used to find the partial specific volume (van Holde, 1998):

$$\rho = \rho_o + (1-\bar{v}\rho_o)C$$

ρ = density at concentration C $\bar{v}$ = partial specific volume
ρ_o = density of the solvent C = concentration in g/ml

The problem with this technique is that large quantities of protein are needed in order to take accurate measurements. A second possible method is by performing sedimentation equilibrium runs in H_2O and D_2O. The difference in the concentration distribution at equilibrium due to the change in the density of the solution by D_2O is used to determine the partial specific volume. This technique requires less protein and only two equilibrium run measurements (Edelstein and Schachman, 1967). After the vbars have been measured, any changes will be incorporated into the calculation of $S_{20,w}$ values and into the HYDROPRO predictions.

In order to solve the contradictory evidence for the 5' nuclease domain of Taq, light scattering measurements of the diffusion coefficient (D_o) and the radius of gyration (R_g) will be performed. The R_g is related to the angles of scattered light caused by the macromolecule and can be useful in determining the dimensions of a particle. According to HYDROPRO, predicted changes in the D_o and R_g between the extended and compact Taq structures are significant (7-16% for D_o and 13-20% for R_g); therefore, light scattering should be useful in determining the orientation of the 5' nuclease domain. We will also investigate whether the orientation of the domain changes as a function of solution conditions by taking measurements in various buffers.

References

Barnes, W. M. (1995) Thermostable DNA polymerase with enhanced thermostability and enhanced length and efficiency of primer extension. *United States Patent* **#5,436,149.**

Beese, L. S., Derbyshire, V., and Steitz, T. A. (1993) Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260:** 352-355.

Beese, L. S., and Steitz, T. A. (1991) Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: A two metal ion mechanism. *EMBO J.* **10(1):** 25-33.

Cantor, C., and Schimmel, P. R. (1980) *Biophysical Chemistry.* W. H. Freeman, San Francisco.

Chien, A. Edgar, D. B., and Trela, J. M. (1976) Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus. J. Bacteriol.* **127(3):** 1550-1557.

de la Torre, J. G., Huertas, M. L., and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78:** 719-730.

de la Torre, J. G., Navarro, S., Lopez Martinez, M. C., Diaz, F. G., and Lopez Cascales, J. J. (1994) HYDRO: A computer software for the prediction of hydrodynamic properties of macromolecules. *Biophys. J.* **67:** 530-531.

Edelstein, S. J. and Schachman, H. K. (1967) The simultaneous determination of partial specific volumes and molecular weights with microgram quantities. *J. Biol. Chem.* **242(2):** 306-311.

Engelke, D. R., Krikos, A., Bruck, M. E., and Ginsburg, D. (1990) Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli. Anal. Biochem.* **191(2):** 396-400.

Eom, S. H., Wang, J., and Steitz, T. A. (1996) Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature* **382:** 278-281.

Joyce, C. M. and Grindley, N. D. F. (1983) Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA Polymerase I of *Escherichia coli. Proc. Natl. Acad. Sci. USA* **80:** 1830-1834.

Kim, Y., Eom, S. H., Wang, J., Lee, D., Suh, S. W., and Steitz, T. A. (1995) Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **376:** 612-616.

Klenow, H. and Henningsen, I. (1970) Selective elimination of the exonuclease activity of the Deoxyribonucleic Acid polymerase from *Escherichi coli* B by limited proteolysis. *Proc. Natl. Acad. Sci. USA* **65(2):** 168-175.

Korolev, S., Nayal, M., Barnes, W. M., Di Cera, E., and Waksman, G. (1995) Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5-A resolution: Structural basis for thermostability. *Proc. Natl. Acad. Sci. USA* **92:** 9264-9268.

Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. (1989) Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *J. Biol. Chem.* **264(11):** 6427-6437.

Li, Y., Korolev, S., and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: Structural basis for nucleotide incorporation. *EMBO J.* **17(24):** 7514-7525.

Minkley Jr., E. G., Leney, A. T., Bodner, J. B., Panicker, M. M., and Brown, W. E. (1984) *Escherichia coli* DNA Polymerase I: Construction of a *polA* plasmid for amplification and an improved purification scheme. *J. Biol. Chem.* **259(16):** 10386-10392.

Murali, R., Sharkey, D. J., Daiss, J. L. and Krishna Murthy, H. M. (1998) Crystal structure of *Taq* DNA polymerase in complex with an inhibitory Fab: The Fab is directed against an intermediate in the helix-coil dynamics of the enzyme. *Proc. Natl. Acad. Sci. USA* **95:** 12562-12567.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313:** 762-766.

Penefsky, H. S. (1977) Reversible binding of P_i by beef heard mitochondrial adenosine triphosphatase. *J. Biol. Chem.* **252(9):** 2891-2899.

Steitz, T. A. (1999) DNA polymerases: Structural diversity and common mechanisms. *J. Biol. Chem.* **274(25):** 17395-17398.

Svedberg, T., Pederson, K. O., and Bauer, J. H. (1959) *The Ultracentrifuge.* Johnson Reprint Corp., New York.

Urs, U. K., Murali, R., and Murthy, H. M. (1999) Crystal structure of Taq DNA Polymerase shows new orientation for the structure-specific nuclease domain. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55:** 1971-1977.

van Holde, K. E., Johnson, W. C., and Ho, S. (1998) *Principles of Physical Biochemistry.* Prentice-Hall, Inc., Upper Saddle River, New Jersey. pp 192-241, 544-556, 567-572.

Voet, D., Voet, J. E., and Pratt, C. W. (1999) *Fundamentals of Biochemistry.* John Wiley and Sons, Inc., New York. pp 776-810.

Acknowledgements

First and foremost I would like to thank Dr. LiCata for his help and support during the past year and a half. Thank you, Dr. LiCata, for allowing me to be a part of your lab and for your constant enthusiasm and assistance.

I would also like to thank my committee members Dr. Russo and Dr. Stephens for their support and encouragement and for agreeing to be a part of this process.

Special thanks also to the members of the LiCata lab: Jason Bell, Kausiki Datta, Allison Joubert, Rena Karantzeni, Laurel Riley, Carmen Ruiz, Allyn Schoeffler, and Xuemei Yang. Thank you all for your guidance and support, and for making the lab an enjoyable place to be. Thank you, Carmen, Kausiki, and Irene, for so willingly giving up your proteins. Thank you, Jason, for performing many of the viscosity and density measurements. Thank you, Laurel, for teaching me how to run the ultracentrifuge.

Finally, thank you to my family for their constant support over the years, especially for listening to me worry and complain these last few weeks. I could never have gotten this far without you.