

Analysis of Linked Equilibria

JiaBei Lin, Aaron L. Lucius¹

Department of Chemistry, The University of Alabama at Birmingham, Birmingham, Alabama, USA

¹Corresponding author: e-mail address: allucius@uab.edu

Contents

1. Introduction	161
2. Determination of $L_{n,0}$ for an Assembling System	168
3. Global Fitting of Sedimentation Velocity Data as a Function of Protein Concentration Kinetic Considerations	169
4. Global Analysis Using the 1-2-4-6 Model with Rapid Dissociating Oligomers	175
5. Global Analysis Using the 1-2-4-6 Model with Slow Dissociation of Oligomers	180
6. Global Analysis with Rate Constants in the Detectable Range	181
7. Conclusions	183
References	185

Abstract

The ATPases associated with diverse cellular activities (AAA+) is a large superfamily of proteins involved in a broad array of biological processes. Many members of this family require nucleotide binding to assemble into their final active hexameric form. We have been studying two example members, *Escherichia coli* ClpA and ClpB. These two enzymes are active as hexameric rings that both require nucleotide binding for assembly. Our studies have shown that they both reside in a monomer, dimer, tetramer, and hexamer equilibrium, and this equilibrium is thermodynamically linked to nucleotide binding. Moreover, we are finding that the kinetics of the assembly reaction are very different for the two enzymes. Here, we present our strategy for determining the self-association constants in the absence of nucleotide to set the stage for the analysis of nucleotide binding from other experimental approaches including analytical ultracentrifugation.



1. INTRODUCTION

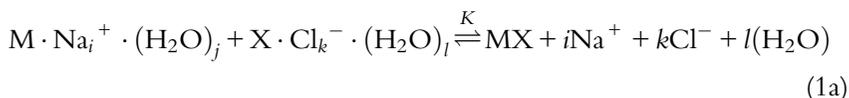
Determining a binding constant for a protein–ligand interaction is quite possibly one of the most common experiments in biophysics. Techniques such as ITC, fluorescence titrations, equilibrium dialysis, and many others are commonly used. Analysis of these data is fairly straightforward if

the protein of interest does not change its oligomeric state as the free concentration of the ligand (chemical potential of the ligand) is increased. However, if the protein does change its oligomeric state as the chemical potential of the ligand is increased, then the analysis becomes substantially more complex.

It is important to recall that a binding constant measured using any of the approaches stated above is an apparent binding constant (Alberty, 2003). Although we typically write down a binding reaction as an association between a macromolecule, M, and a ligand, X, as schematized in Eq. (1), the binding is actually much more complex.



In fact, the interaction also involves the removal of water and/or ions from the binding pocket as well as from the ligand that is entering the binding pocket. Although there could be many solution condition components, including protons, involved in this interaction, an example scheme that represents only ions and water exchange upon ligand binding is given by Eq. (1a).



These exchanges of water and ions with bulk solvent all contribute to the energetics of the binding interaction. Therefore, the binding equilibrium constant, K_{app} , given in Eq. (1) will exhibit a dependence on solution condition variables such as ion concentration, pH, water concentration. In other words, the binding constant will be thermodynamically linked to all of the components involved in the binding reaction. Thus, linkage analysis can be used to deconvolute the energetics of a “second ligand” (ion, water, protons, etc.) binding. This is accomplished by determining the apparent equilibrium constant for X binding M as a function of, for example, salt concentration. Then, this apparent equilibrium constant can be treated as a signal for a “second ligand” binding event (water, proton, ions, etc.).

With the above in mind, one can recognize that a ligand binding constant for an assembling system may exhibit a protein concentration dependence. Likewise, a protein–protein interaction constant will exhibit a ligand concentration dependence for an assembling system. That is to say, the

binding constants are thermodynamically linked to either the protein concentration or ligand concentration.

The ATPases associated with diverse cellular activities (AAA+) is a large superfamily of proteins involved in a broad array of biological processes (Neuwald, Aravind, et al., 1999). Examples include microtubule severing catalyzed by katanin (Roll-Mecak & McNally, 2010); membrane fusion involving *N*-ethylmaleimide-sensitive fusion (NSF) proteins (Yu, Jahn, & Brunger, 1999); morphogenesis and trafficking of endosomes by VPs4p (Babst, Sato, Banta, & Emr, 1997); protein disaggregation by ClpB/Hsp104; and enzyme-catalyzed protein unfolding and translocation by ClpA or ClpX for ATP-dependent proteolysis (Ogura & Wilkinson, 2001; Sauer & Baker, 2011). Another example is human VCP/p97, which has been connected to ubiquitin-dependent reactions, but is implicated in an expanding number of physiological processes (Meyer & Wehl, 2014).

Many of these proteins require nucleoside triphosphate binding to assemble into their final active hexameric form. Thus, ligand-linked assembly is an integral component of the mechanism driving assembly. For many of these examples, the oligomers present in solution will change as the chemical potential of the ligand (nucleotide) is increased.

Analytical ultracentrifugation is the technique of choice for examining the energetics of macromolecular assembly. The technique has been used extensively to do so and a large number of computer applications are available to aid in the analysis of the experimental results, and these have been discussed in many places (Cole, 2004; Demeler, Brookes, Wang, Schirf, & Kim, 2010; Schuck, 1998; Scott, Harding, Rowe, & Royal Society of Chemistry (Great Britain), 2005; Stafford & Sherwood, 2004). Moreover, an enormous body of literature exists on the application of this approach to examine assembling systems (Cole, 1996; Correia & Stafford, 2009; Schuck, 1998, 2003). However, substantially less has been done on examining ligand-linked assembly problems, but some examples can be found (Cole, Correia, & Stafford, 2011; Na & Timasheff, 1985a, 1985b; Streaker, Gupta, & Beckett, 2002; Wong & Lohman, 1995).

Here, we outline our strategy for elucidating the thermodynamic mechanism for an assembling system that exists in a complex, dynamic equilibrium of monomers, dimers, tetramers, and hexamers. This is being done in order to set the stage for an examination of the thermodynamic linkage to the nucleotide-driven hexamer formation for two example AAA+

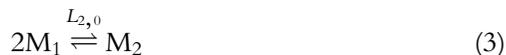
macromolecular machines, *Escherichia coli* ClpA and ClpB. We have found that both proteins form hexamers and both exist as mixtures of oligomeric states both in the presence and absence of nucleotide (Li, Lin, & Lucius, 2015; Lin & Lucius, 2015; Veronese & Lucius, 2010; Veronese, Stafford, & Lucius, 2009; Veronese, Rajendar, & Lucius, 2011). This is a difficult problem to address for a variety of reasons, some of which will be discussed here. Nevertheless, the number of systems that exhibit ligand-linked assembly is large and there is a pressing need for a set of strategies and approaches to solve the problem.

Oftentimes, without being able to predict the concentration of each species in solution, quantitatively interpreting binding and catalytic data are not possible. However, solving these problems is needed for more than just interpretation of *in vitro* studies. We are finding that the nucleotide binding affinity of these motor proteins is in the range of 10–100 μM , which is at least an order of magnitude below the concentration of nucleotide in the cell (5–10 mM). This indicates that the nucleotide binding sites on these macromolecules would likely be saturated in the cell, and thus, nucleotide is not likely to be a regulatory molecule. On the other hand, the concentration of these proteins in the cell (Dougan, Reid, Horwich, & Bukau, 2002; Farrell, Grossman, & Sauer, 2005; Mogk et al., 1999) have been found to be similar to their ligand-linked assembly dissociation equilibrium constant, indicating that the linkage between nucleotide binding and assembly may be an important regulatory component of their function.

To begin to quantitatively address the linkage of ligand binding to macromolecular assembly, we can start by writing down a partition function, Q , that represents the sum of all of the macromolecular states for an arbitrary solution containing monomers (M_1), dimers (M_2), tetramers (M_4), and hexamers (M_6) as follows:

$$Q = [M_1] + [M_2] + [M_4] + [M_6] \quad (2)$$

We can define self-association equilibrium constants for each thermodynamic state. Here, we will use L for stoichiometric or overall protein–protein interaction constants and K for both stepwise protein–protein interaction constants and for ligand binding constants. For the monomer, dimer, tetramer, and hexamer system, the following three reactions given by Eqs. (3)–(5) can define the equilibria





The three stoichiometric equilibrium constants that result are given by Eqs. (6)–(8)

$$L_{2,0} = \frac{[M_2]}{[M_1]^2} \quad (6)$$

$$L_{4,0} = \frac{[M_4]}{[M_1]^4} \quad (7)$$

$$L_{6,0} = \frac{[M_6]}{[M_1]^6} \quad (8)$$

where the first subscript represents the oligomeric state and the second subscript represents the nucleotide ligation state, in this case the zero represents no ligand bound. The partition function given in Eq. (2) can be simplified by algebraically solving Eqs. (6)–(8) for each oligomer and substituting the solutions into Eq. (2) to yield Eq. (9)

$$Q = [M_1] + L_{2,0}[M_1]^2 + L_{4,0}[M_1]^4 + L_{6,0}[M_1]^6 \quad (9)$$

where the partition function given by Eq. (9) is only a function of the free monomer concentration and the equilibrium constants for each oligomer formed. If each of the oligomers can bind ligand, then each term in the partition function will be multiplied by a partition function for binding of ligand to that particular oligomer to yield Eq. (10)

$$Q = [M_1]P_1 + L_{2,0}[M_1]^2P_2 + L_{4,0}[M_1]^4P_4 + L_{6,0}[M_1]^6P_6 \quad (10)$$

where P_1 , P_2 , P_4 , and P_6 represent the partition functions for the binding of ligand to the monomer, dimer, tetramer, and hexamer, respectively. The steps for deriving Eq. (10) are straightforward and will not be reproduced here; for a review, see [Wyman & Gill \(1990\)](#). For simplicity, if the monomer binds one ligand, X, then the partition function for ligand binding is given by the sum of all of the monomeric states normalized to the unligated state given by Eq. (11)

$$P_1 = \frac{[M_1] + [M_1X]}{[M_1]} \quad (11)$$

If we define an equilibrium constant for nucleotide binding to the monomer as:

$$K_{1,1} = \frac{[M_1X]}{[M_1]} \quad (12)$$

where the first subscript represents the oligomeric state and the second subscript represents the number of ligands bound. We can simplify Eq. (11) to be:

$$P_1 = 1 + K_{1,1}[X] \quad (13)$$

If we assume that the monomer has n -independent and identical binding sites, then the partition function given by Eq. (13) would be expressed as Eq. (14)

$$P_1 = (1 + K_1[X])^{n_1} \quad (14)$$

where n_1 represents the number of binding sites per monomer. If we assume that each oligomer binds n number of ligands independently and identically, then the partition function for the entire system is given by

$$Q = [M_1](1 + K_1[X])^{n_1} + L_{2,0}[M_1]^2(1 + K_2[X])^{n_2} + L_{4,0}[M_1]^4(1 + K_4[X])^{n_4} + L_{6,0}[M_1]^6(1 + K_6[X])^{n_6} \quad (15)$$

where n_1 , n_2 , n_4 , and n_6 represent the number of binding sites on the monomer, dimer, tetramer, and hexamer, respectively, and K_1 , K_2 , K_4 , and K_6 represent the average binding constant for ligand binding to monomers, dimers, tetramers, and hexamers, respectively, where the average binding constant has been corrected with statistical factors (Wyman & Gill, 1990). Since, in this model, all of the binding sites are assumed to be the same, there is no second subscript on K_x to denote the number of ligands bound.

For any experiment that would seek to examine ligand binding to such a complex system, whether it be ITC, fluorescence titrations, or some other approach, the signal is typically proportional to ligand bound divided by total macromolecule, which is defined as “extent of binding” or \bar{X} . The importance of expressing the partition function is that the partition function can now be used to derive an equation that represents the extent of binding, ligand bound over total macromolecule, which could be used to analyze ligand binding data. The extent of binding is given by Eq. (16) (Wyman & Gill, 1990)

$$\frac{[X]_{\text{Bound}}}{[M]_{\text{Total}}} = \bar{X} = \frac{dQ/d \ln [X]}{dQ/d \ln [M]} = \frac{[X]}{[M]} \frac{dQ/d [X]}{dQ/d [M]} \quad (16)$$

If the derivatives in Eq. (16) are applied to the partition function given by Eq. (15), then the extent of binding equation is given by Eq. (17)

$$\bar{X} = \frac{[M_1]n_1K_1[X](1 + K_1[X])^{n_1-1} + L_{2,0}[M_1]^2n_2K_2[X](1 + K_2[X])^{n_2-1} + L_{4,0}[M_1]^4n_4K_4[X](1 + K_4[X])^{n_4-1} + L_{6,0}[M_1]^6n_6K_6[X](1 + K_6[X])^{n_6-1}}{[M_1](1 + K_1[X])^{n_1} + 2L_{2,0}[M_1]^2(1 + K_2[X])^{n_2} + 4L_{4,0}[M_1]^4(1 + K_4[X])^{n_4} + 6L_{6,0}[M_1]^6(1 + K_6[X])^{n_6}} \quad (17)$$

Clearly, Eq. (17) would have entirely too many parameters to apply to a single binding isotherm collected with any technique. However, several important predictions can be made from inspection of Eq. (17) or from inspection of the partition function given in Eq. (15). First, these equations are functions of the self-association constants in the absence of ligands, $L_{n,0}$, the ligand binding constants to each oligomer, K_n , the free monomer concentration $[M_1]$, and the free ligand concentration $[X]$. Second, Eq. (17), tells the experimentalist that there is a need to first define the self-association equilibrium constants in the absence of nucleotide, $L_{n,0}$. Third, unlike binding isotherms for simple systems, a binding system that is linked to macromolecular assembly will exhibit a dependence on the free protein concentration. This tells the experimentalist that binding studies will have to be executed over a range of protein concentrations.

Figure 1 shows a series of isotherms simulated using Eq. (17) with several different total macromolecule concentrations ranging from 1 to 10 μM . In this example, $L_{2,0} = 1 \times 10^4 \text{ M}^{-1}$, $L_{4,0} = 1 \times 10^{14} \text{ M}^{-3}$, and $L_{6,0} = 1 \times 10^{24} \text{ M}^{-5}$, each monomer is considered to bind one ligand so that $n_1 = 1$, $n_2 = 2$, $n_4 = 4$, and $n_6 = 6$, the ligand binding constants for the monomers through tetramers are all identical so that $K_1 = K_2 = K_4 = 1 \times 10^5 \text{ M}^{-1}$ and the hexamer binding constant is an order of magnitude tighter, $K_6 = 1 \times 10^6 \text{ M}^{-1}$.

The most salient feature of the binding isotherms shown in Fig. 1 is that there is a shift of the midpoint to lower free ligand concentration as the macromolecule concentration increases. This is the consequence of the fact that as the macromolecule concentration increases, there is a corresponding increase in the concentration of hexamers. This observation is general because the thermodynamic driving force for an assembly reaction is the chemical potential of the free monomer. Thus, there will always be an increase in the

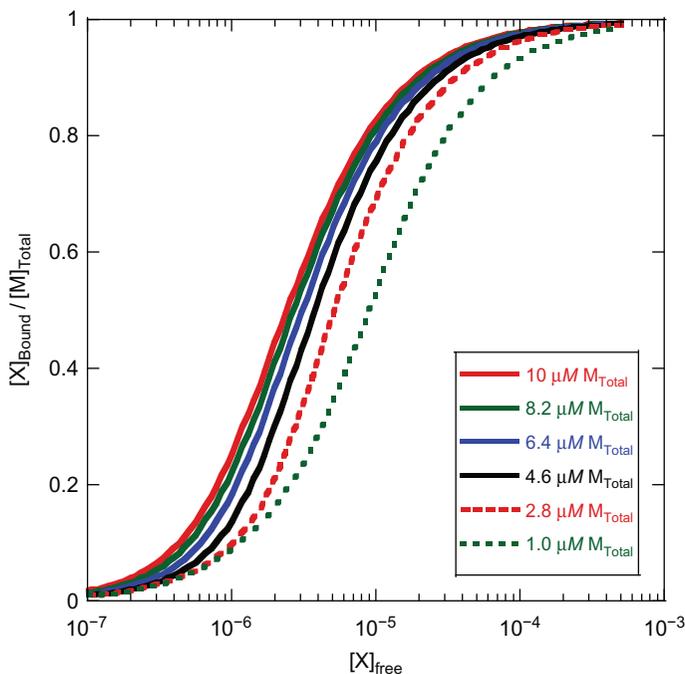
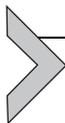


Figure 1 Predicted binding isotherms from Eq. (17) with $L_{2,0}=1 \times 10^4 M^{-1}$, $L_{4,0}=1 \times 10^{14} M^{-3}$, and $L_{6,0}=1 \times 10^{24} M^{-5}$, each monomer is considered to bind one ligand so that $n_1=1$, $n_2=2$, $n_4=4$, and $n_6=6$, the ligand binding constants for the monomers through tetramers are all identical so that $K_1=K_2=K_4=1 \times 10^5 M^{-1}$ and the hexamer binding constant is an order of magnitude tighter, $K_6=1 \times 10^6 M^{-1}$.

population of higher-order oligomers with increasing protein concentration. The apparent increase in ligand binding affinity as the macromolecule concentration is increased is because the ligand binds to the hexamer with an affinity constant one order of magnitude tighter than the other oligomers.

When examining ligand binding to a macromolecule, it is always advisable to perform multiple titrations at several total macromolecule concentrations (Lohman & Mascotti, 1992). If such a strategy is invoked and an apparent change in the affinity constant is observed as a function of macromolecule concentration, then this would be the first indicator that a ligand-linked assembly process is occurring.



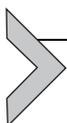
2. DETERMINATION OF $L_{n,0}$ FOR AN ASSEMBLING SYSTEM

If assembly is suspected from an experiment such as that illustrated by Fig. 1, then Eq. (17) suggests that the first objective would be to determine

the self-association equilibrium constants in the absence of any ligand, $L_{n,0}$, so that this parameter could be constrained in the examination of the titration curves. To determine $L_{n,0}$, we perform sedimentation velocity experiments over a range of protein concentrations.

Sedimentation velocity has been our experiment of choice over sedimentation equilibrium. This is because we have found that the nucleotide bound hexamers are not stable over the timescale required for sedimentation equilibrium, which is often several days. In contrast to a sedimentation equilibrium experiment, a sufficient number of sedimentation boundaries can be collected within 2–3 h in a sedimentation velocity experiment. Further, deconvoluting three or more species from exponential fitting performed on sedimentation equilibrium boundaries can be difficult.

Naturally, one would not know the self-association equilibrium constants, so knowing what concentrations to examine will not be initially certain. However, if a series of isotherms were collected like those shown in Fig. 1, one could judge that the assembly state is making a transition over the range of 1–10 μM and this would be a reasonable starting point.



3. GLOBAL FITTING OF SEDIMENTATION VELOCITY DATA AS A FUNCTION OF PROTEIN CONCENTRATION KINETIC CONSIDERATIONS

Sedimentation velocity data are potentially sensitive to the assembly kinetics if the dissociation rate constants for the species are in the range of 10^{-2} – 10^{-5} s^{-1} (Correia, Alday, Sherwood, & Stafford, 2009; Dam, Velikovskiy, Mariuzza, Urbanke, & Schuck, 2005; Demeler et al., 2010; Stafford & Sherwood, 2004). Consequently, there are two ways to globally analyze sedimentation velocity data. The first is to assume that the system is always at thermodynamic equilibrium. This assumption holds if the dissociation rate constants are faster than 10^{-2} s^{-1} . The second is that the dissociation rate constant is found to be within or slower than the empirical range 10^{-2} – 10^{-5} s^{-1} . If the dissociation rate constants are within this empirical range, modeling the reaction kinetics is required. Otherwise, for dissociations that are slower than 10^{-5} s^{-1} , the components can be considered as noninteracting discrete species.

To illustrate the impact of the assembly kinetics, we simulated sedimentation velocity experiments with reverse rate constants of either 1 or 10^{-6} s^{-1} . Dissociation rate constants of $k_r=1$ or 10^{-6} s^{-1} correspond to half-lives of 0.7 s and 192 h, respectively. It is not difficult to conclude that reaction kinetics occurring with these half-lives would be outside of the

detectable range in a sedimentation velocity experiment since the boundaries are typically collected on the minutes timescale. Although one could enhance the temporal resolution with interference, experiments performed on only a single cell since these could be collected every 8 s.

To simulate the sedimentation boundaries using SEDANAL, the concentration of each species needs to be modeled. Here, we will use the language of the Gibbs phase rule. A component is defined as a chemical component and a species is made up of products of reactions between the components. Thus, in an experiment containing a single protein, “M,” that reacts to form dimers, tetramers, and hexamers, we define “M” as the component and monomers, dimers, tetramers, and hexamers as species. Therefore, this is a single component—four-species system.

The first step in the simulation is to relate the total loading concentration of the protein to the concentration of each species. This is done by writing down the conservation of mass equation given by Eq. (18)

$$[M_1]_T = [M_1] + 2[M_2] + 4[M_4] + 6[M_6] \quad (18)$$

where $[M_1]_T$ is the total monomer concentration, $[M_1]$, $[M_2]$, $[M_4]$, and $[M_6]$ are the equilibrium concentrations of monomers, dimers, tetramers, and hexamers. The coefficients of 2, 4, and 6 are present because the total concentration is expressed in monomer units, e.g., 2 monomers in a dimer. Equation (18) can be expressed in terms of the equilibrium constants given by Eqs. (6)–(8) and the free monomer concentration, $[M_1]$, to yield Eq. (19)

$$[M_1]_T = [M_1] + 2L_{2,0}[M_1]^2 + 4L_{4,0}[M_1]^4 + 6L_{6,0}[M_1]^6 \quad (19)$$

In a sedimentation velocity experiment, there are two stages of equilibrium that one needs to consider. Those stages are before the force of centrifugation is applied and while the force is present. When the force is present, the equilibrium can be perturbed. However, before the force is applied, the system should be at equilibrium. This “pre-equilibrium” is achieved if the experimentalist has given sufficient time for the system to fully relax to equilibrium after any perturbations, i.e., preparation of samples, dilutions, temperature equilibration. Thus, if an experiment is intended to be performed at, for example, $[M_1]_T = 1 \mu M$, then the experimentalist would make up this solution in some vessel. In that vessel, the system will distribute itself into free monomers, dimers, tetramers, and hexamers, where the population of each could be defined by Eq. (19) if the equilibrium constants are known. However, an important control would be to allow this

sample to incubate for increasing amounts of time before performing the run. Then, the results could be compared after, for example, a 6 h versus 12 h preincubation time. If the system is at equilibrium, one would expect to see results that are independent of incubation time. If incubation time-dependent differences are observed, then the incubation time should be extended until differences are no longer observed.

To determine the concentration of each species for this system, one needs to solve Eq. (19), which is a sixth-order polynomial in the free monomer concentration. However, the free monomer concentration, $[M_1]$, is not known. What is known to the experimentalist is the total monomer concentration, $[M_1]_T$. Thankfully, numerically solving a polynomial is not a difficult task.

In SEDANAL, the roots of Eq. (19) and thus the concentrations of each species are determined by numerical methods, specifically the Newton–Raphson method. In practice, this is achieved in SEDANAL by going to “preferences” choosing “Control extended,” “Kinetics/equilibrium control” and under “Initial equilibration” one chooses “No analytic solution” under the Newton–Raphson column. Again, since this model requires solving a sixth-order polynomial, there is no analytic solution and therefore the equation must be solved numerically.

What was just described represents the determination of the concentration of each species in the reaction vessel before the sample is subjected to the force of sedimentation. Since there is no force, there is no perturbation of the equilibrium, and therefore, the concentrations of each species are fixed. The next task at hand is to define the concentrations of each species upon application of force. Since, in a sedimentation velocity experiment, the experimentalist is observing the time-dependent separation of each species, the experiment is potentially sensitive to the reaction kinetics.

To simulate the sedimentation boundaries, one needs to again determine the concentration of each species and then model the sedimentation of each species by passing the determined concentration to the Lamm equation, which defines the movement of the particle under the force of centrifugation and back diffusion. For the monomer, dimer, tetramer, and hexamer reaction, we assume that all species are being formed through bimolecular interactions defined by the following reactions given by Eqs. (20)–(22)





where $k_{f,n}$ is the bimolecular association rate constant with units of $M^{-1} s^{-1}$ and $k_{r,n}$ is the dissociation rate constant with units of s^{-1} . We assume that all species are formed through bimolecular interactions because trimolecular reactions and above are highly improbable. The equilibrium constants for the reactions in Eqs. (20)–(22) are given by Eqs. (23)–(25)

$$K_2 = \frac{k_{f2}}{k_{r2}} = \frac{[M_2]}{[M_1]^2} \quad (23)$$

$$K_4 = \frac{k_{f4}}{k_{r4}} = \frac{[M_4]}{[M_2]^2} \quad (24)$$

$$K_6 = \frac{k_{f6}}{k_{r6}} = \frac{[M_6]}{[M_2][M_4]} \quad (25)$$

It is important to note that if the system is an equilibrium system, then the stepwise equilibrium constants given by Eqs. (23)–(25) can be related to the stoichiometric interaction constants given by Eqs. (6)–(8) as follows:

$$L_{2,0} = K_2 \quad (26)$$

$$L_{4,0} = K_2^2 \cdot K_4 \quad (27)$$

$$L_{6,0} = K_2^3 \cdot K_4 \cdot K_6 \quad (28)$$

Since the equilibrium is potentially being perturbed, determining the concentration of each species is accomplished by numerically solving the following system of coupled differential equation given by Eqs. (29)–(32)

$$\frac{d[M_1]}{dt} = [M_2]k_{r2} - [M_1]^2k_{f2} = [M_2]k_{r2} - [M_1]^2K_2k_{r2} \quad (29)$$

$$\begin{aligned} \frac{d[M_2]}{dt} &= [M_1]^2k_{f2} - [M_2]k_{r2} - [M_2]^2k_{f4} + [M_4]k_{r4} - [M_2][M_4]k_{f6} + [M_6]k_{r6} \\ &= [M_1]^2K_2k_{r2} - [M_2]k_{r2} - [M_2]^2K_4k_{r4} + [M_4]k_{r4} - [M_2][M_4]K_6k_{r6} + [M_6]k_{r6} \end{aligned} \quad (30)$$

$$\begin{aligned} \frac{d[M_4]}{dt} &= [M_2]^2k_{f4} - [M_4]k_{r4} - [M_2][M_4]k_{f6} + [M_6]k_{r6} \\ &= [M_2]^2K_4k_{r4} - [M_4]k_{r4} - [M_2][M_4]K_6k_{r6} + [M_6]k_{r6} \end{aligned} \quad (31)$$

$$\frac{d[M_6]}{dt} = [M_2][M_4]k_{f6} - [M_6]k_{r6} = [M_2][M_4]K_6k_{r6} - [M_6]k_{r6}, \quad (32)$$

where the right-hand side of Eqs. (29)–(32) has been simplified to include only the equilibrium constant and reverse rate constants for each reaction. To model the sedimentation of each species, the system of coupled differential equations is numerically integrated to determine the concentration of each species as a function of time based on the values of the equilibrium constants and dissociation rate constants for each reaction.

In the extreme of rapid dissociation, $k_r > 0.01 \text{ s}^{-1}$ ($t_{1/2} < 1.2 \text{ min}$), the system is considered to be in rapid equilibrium. That is to say, for each infinitely small radial slice of solution the macromolecules that rapidly dissociate within this slice are considered to rapidly reassociate, and therefore, the differential equations given in Eqs. (29)–(32) are all equal to zero, $d[M_n]/dt = 0$. This indicates that the concentration of each oligomer in each infinitely small radial slice is constant and thus at equilibrium.

In contrast, if the dissociation rate constants are on the other extreme, $k_r < 10^{-5} \text{ s}^{-1}$ ($t_{1/2} > 19 \text{ h}$), then the differential equations are again equal to zero and the concentrations of each species are again considered to be fixed. However, in this scenario there is no reequilibration at each infinitely small radial slice because no dissociation is occurring during the course of the entire experiment. Thus, the species are being separated when force is applied. In this scenario, the system can be modeled as noninteracting discrete species because they will not appear to react on the timescale of sedimentation.

To illustrate these points, we simulated a series of sedimentation boundaries from the monomer, dimer, tetramer, and hexamer model given by Eqs. (20)–(22) using SEDANAL for three total protein concentrations, $[M_1]_T = 1, 9$ and $15 \mu\text{M}$. We refer to this model as the “1-2-4-6” model. For the first simulation, all reverse rate constants were considered to be fast relative to sedimentation, $k_{r2} = k_{r4} = k_{r6} = 1 \text{ s}^{-1}$ (see Table 1). In a second simulation, all reverse rate constants were considered to be slow, $k_{r2} = k_{r4} = k_{r6} = 10^{-6} \text{ s}^{-1}$ (see Table 1). Both extremes of the rate constants were considered to be outside of 10^{-2} – 10^{-5} s^{-1} , which has been previously reported to be the range over which one would expect to be able to extract meaningful measures of the rate constants from sedimentation velocity data (Dam et al., 2005; Demeler et al., 2010; Stafford & Sherwood, 2004).

Our preferred first level of analysis is to analyze the sedimentation boundaries using SedFit to generate $c(s)$ distributions (Peter Shuck, NIH) (Schuck, 1998). Figure 2A shows the results of a $c(s)$ analysis on the simulated data assuming all dissociation rate constants are 1 s^{-1} . In red is the $c(s)$ distribution from the analysis of the simulation with $[M_1]_T = 1 \mu\text{M}$ and a single

Table 1 Parameters Used to Simulate the Sedimentation Velocity Data for 1–15 μM Macromolecule Concentration

Model Used for Simulation	Parameters Used for Fast Dissociation		Parameters Used for Slow Dissociation	
	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$
$2B \xrightleftharpoons[k_{r1}]{k_{f1}} B_2$	6×10^4	1	6×10^4	1×10^{-6}
$2B_2 \xrightleftharpoons[k_{r4}]{k_{f4}} B_4$	1.5×10^5	1	1.5×10^5	1×10^{-6}
$B_2 + B_4 \xrightleftharpoons[k_{r6}]{k_{f6}} B_6$	1.5×10^6	1	1.5×10^6	1×10^{-6}

The stoichiometric equilibrium constants for dimerization, tetramerization, and hexamerization can be calculated using Eqs. (25)–(27) and are: $L_{2,0} = 6 \times 10^4 M^{-1}$, $L_{4,0} = 5.4 \times 10^{14} M^{-3}$, and $L_{6,0} = 4.86 \times 10^{25} M^{-5}$. The sedimentation coefficients used for simulations are $s = 3.0, 5.6, 8.86,$ and 11.9 for monomer, dimer, tetramer, and hexamer, respectively. The standard deviation of noise added to the data is 0.005 absorbance unit.

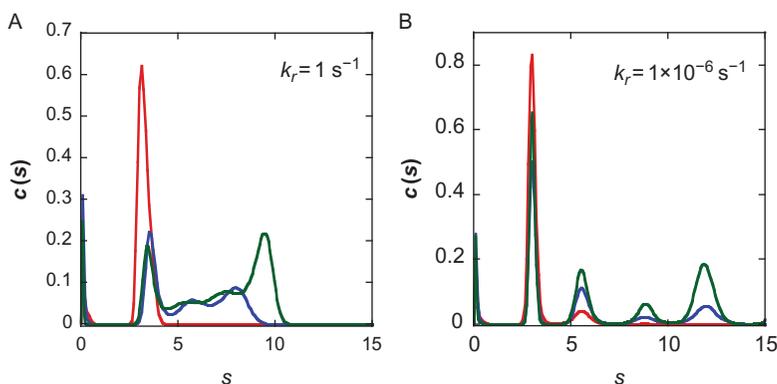


Figure 2 $c(s)$ distributions resulting from $c(s)$ analysis of data simulated from 1, 9, and 15 μM protein. (A) Parameters are given in Table 1 and (B) parameters are given in Table 2. In both cases, the sedimentation coefficients used to simulate the data were $s = 3.0, 5.6, 8.86,$ and 11.9 for monomer, dimer, tetramer, and hexamer, respectively. The extinction coefficient used in the simulation is $4.2 (\text{mg/ml})^{-1} \text{cm}^{-1}$ at 230 nm and $0.45 (\text{mg/ml})^{-1} \text{cm}^{-1}$ at 280 nm. Time between simulated scans is 4 min.

peak is observed that corresponds to the monomer. In blue is the $c(s)$ analysis for $[M_1]_T = 9 \mu\text{M}$. The peak corresponding to monomer shifts to the right slightly and a broad distribution from ~ 5 to 8 S emerges. In green is the $c(s)$ distribution from the simulation with $[M_1]_T = 15 \mu\text{M}$ and what is observed is

a clear shifting of the peaks further to the right. The peak shifting can be taken as the first indication that the system is exhibiting fast reaction kinetics (Dam & Schuck, 2005; Dam et al., 2005). Thus, a preliminary $c(s)$ analysis that shows this type of broad $c(s)$ distribution may serve to be the first indicator that rapid dissociation may be occurring on the timescale of sedimentation, and this hypothesis would warrant further testing.

Figure 2B shows a $c(s)$ analysis resulting from analysis of simulations performed with all dissociation rate constants $k_r = 10^{-6} \text{ s}^{-1}$. At $1 \mu\text{M}$ total protein, two peaks are observed that correspond to the sedimentation coefficient of monomers and dimers. In contrast, at $15 \mu\text{M}$ all four peaks appear that correspond to the monomers, dimers, tetramers, and hexamers, respectively. Under these conditions of slow dissociation, the oligomers sediment as noninteracting discrete species. Under such conditions, the area under each of these peaks would represent the equilibrium concentrations of each species and could be analyzed to yield the equilibrium constants.



4. GLOBAL ANALYSIS USING THE 1-2-4-6 MODEL WITH RAPID DISSOCIATING OLIGOMERS

If the $c(s)$ plot given by Fig. 2A was experimentally observed, our next step would be to globally fit all protein concentrations by direct boundary analysis in SEDANAL. Again, the goal here is to determine the self-association equilibrium constants, $L_{n,0}$. Since the $c(s)$ plot suggests there may be rapid dissociation, the first strategy in this analysis would be to float each of the reverse rate constants starting with a guess of around 0.01 s^{-1} . Thus, the global floating fitting parameters are the loading concentrations, K_2 , K_4 , K_6 , k_{r2} , k_{r4} , and k_{r6} as given by Eqs. (23)–(25).

As with the simulations described above to fit the data accounting for the kinetics, there are several steps that have to be executed in SEDANAL. We always assume that the system is at equilibrium at the start of the run. In practice, this is achieved in SEDANAL by going to “preferences” choosing “Control extended,” “Kinetics/equilibrium control” and under “Initial equilibration” one chooses “No analytic solution” under the Newton–Raphson column. What this accomplishes is numerically solving Eq. (19) to determine the free monomer concentration and thus the concentration of each oligomer is determined based on the initial guesses of the equilibrium constants. Next, under “Equilibration during run,” we choose either BulSt or SEulEx under “Kinetic integrator” in the “No analytic solution” row.

The choice of kinetic integrator primarily impacts the speed of executing the fit and the differences have been discussed elsewhere (Stafford & Sherwood, 2004), including the SEDANAL manual. What this accomplishes is numerically solving the system of coupled differential equations given by Eqs. (29)–(32) for the concentrations of each species as a function of time and passes this to the Lamm equation to define the movement of the oligomer in the field.

The other parameters required for this analysis are the molar mass, the sedimentation coefficient, density increment, and mass extinction coefficient. The molar mass would be known from sequence information and thus the molar mass of each oligomer is calculated. Estimates of the sedimentation coefficients are needed and solution conditions can usually be modified to acquire reasonable estimates of the monomer and largest oligomers. One can estimate the intermediate sedimentation coefficients based on the $s_n = s_1(n)^{2/3}$ with the assumption that the frictional ratio of the oligomers are the same as monomer's, where s_n is the sedimentation coefficient for the oligomer containing n protomer units and s_1 is the sedimentation coefficient for the monomer (Correia, 2000). Alternatively, if data on the three-dimensional structure are available, then one can approximate hydrodynamic information using applications like Hydropro (Ortega, Amoros, & Garcia de la Torre, 2011; Chapter “Hydrodynamic Modeling and Its Application in AUC” by Rocco and Byron). The extinction coefficient should be rigorously determined by denaturing the protein in 6 M guanidine (Edelhoch, 1967; Gill & von Hippel, 1989; Pace, Vajdos, Fee, Grimsley, & Gray, 1995). The density increments, $d\rho/dc$, can be determined experimentally, but for most cases it is acceptable to substitute $(1 - \bar{v}\rho)$, where \bar{v} is the partial specific volume of the protein and ρ is the density of the buffer. These parameters are typically calculated with SednTerp (David Hayes, Magdalen College; Tom Laue, University of New Hampshire; and John Philo, Alliance Protein Laboratories) (Laue, Shah, Ridgeway, & Pelletier, 1992).

Global NLLS analysis of the data simulated from the 1-2-4-6 model with all the dissociation rate constants $k_{r,n} = 1 \text{ s}^{-1}$ yields good estimates of the equilibrium constants (compare values used to generate data in Table 1 to fitted values in Table 2, first column). The difference curves and their associated fits are shown in Fig. 3. Values of the rate constants float somewhere between ~ 0.3 and 2 s^{-1} (see Table 3). We have interpreted this to indicate that the kinetic parameters are unconstrained because the value used to simulate the data of 1 s^{-1} results in a half-life of $\sim 0.7 \text{ s}$. Thus, there is little

Table 2 Examination of Data Simulated with Fast Dissociation Using the “1-2-4-6” Model

Model Used for Fitting	Fit with Kinetic Integrator Allowing $k_{r,n}$ to Float		Fit with Kinetic Integrator, Constraining $k_{r,n} = 0.01 \text{ s}^{-1}$		Fit with Newton–Raphson, No $k_{r,n}$ Input	
	RMSD = 5.003×10^{-3}		RMSD = 5.929×10^{-3}		RMSD = 5.003×10^{-3}	
	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$
$2B_2 \xrightleftharpoons[k_{r,1}]{k_{f,1}} B_2$	6.00×10^4	0.29	5.87×10^4	0.01	6.00×10^4	NA
$2B_2 \xrightleftharpoons[k_{r,4}]{k_{f,4}} B_4$	1.49×10^5	1.52	2.02×10^5	0.01	1.49×10^5	NA
$B_2 + B_4 \xrightleftharpoons[k_{r,6}]{k_{f,6}} B_6$	1.51×10^6	1.94	1.02×10^6	0.01	1.52×10^6	NA

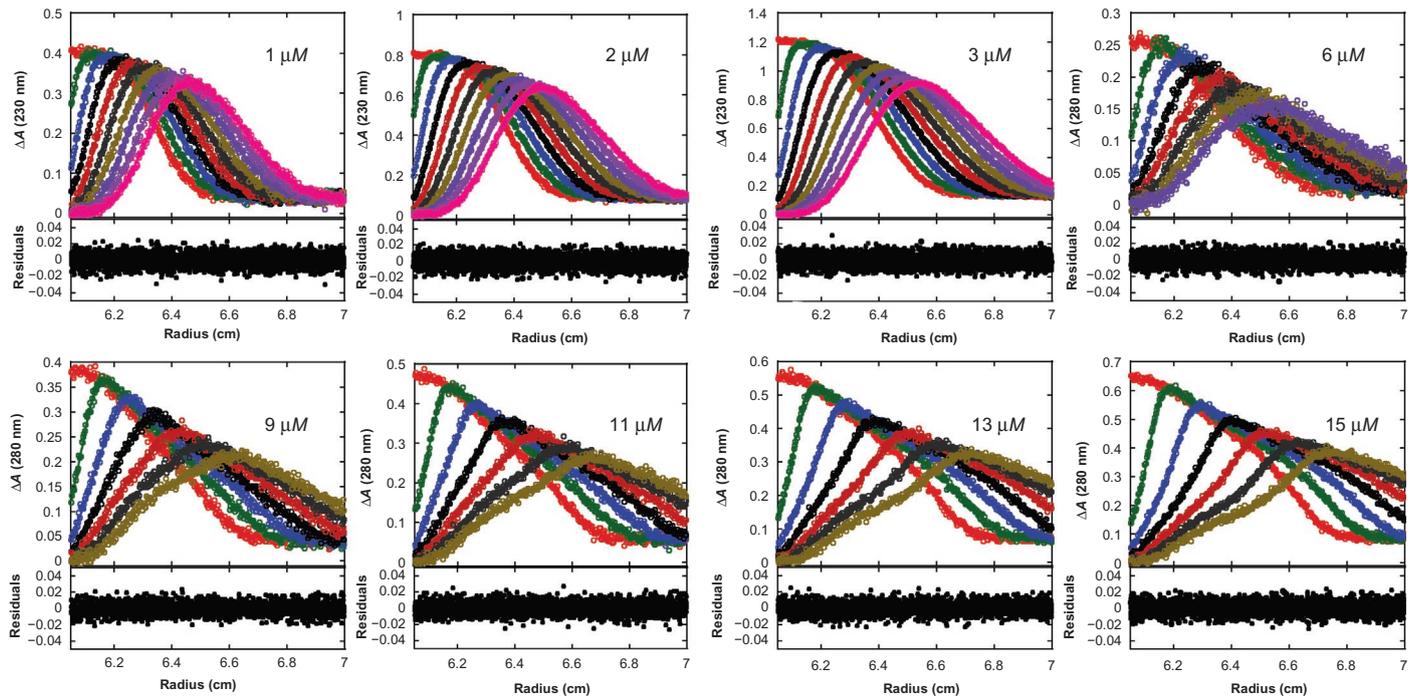


Figure 3 Global analysis of simulated sedimentation velocity data using 1-2-4-6 model by allowing dissociation rate constant, $k_{r,D}$, to float. The data were simulated using parameters presented in [Table 1](#). The fitting results are presented in [Table 3](#). The concentrations of protein are indicated on the plots. Every fourth difference curve is presented.

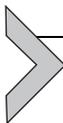
Table 3 Examination of Data Simulated with Slow Dissociation Using the “1-2-4-6” Model

Model Used for Fitting	Fit with Kinetic Integrator Allowing $k_{r,n}$ to Float		Fit with Kinetic Integrator, Constraining $k_{r,n} = 1 \times 10^{-5} \text{ s}^{-1}$		Fit with Newton-Raphson, No $k_{r,n}$ Input	
	RMSD = 5.000×10^{-3}		RMSD = 5.188×10^{-3}		RMSD = 3.178×10^{-2}	
	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$
$2B \xrightleftharpoons[k_{r1}]{k_{f1}} B_2$	6.00×10^4	1.16×10^{-6}	5.91×10^4	1×10^{-5}	5.89×10^3	NA
$2B_2 \xrightleftharpoons[k_{r4}]{k_{f4}} B_4$	1.52×10^5	4.86×10^{-9}	1.65×10^5	1×10^{-5}	6.45×10^6	NA
$B_2 + B_4 \xrightleftharpoons[k_{r6}]{k_{f6}} B_6$	1.49×10^6	5.13×10^{-7}	1.43×10^6	1×10^{-5}	7.00×10^7	NA

information in the sedimentation boundaries that yield constraints on the values of these rate constants.

The next step in our analysis strategy would be to constrain the dissociation rate constants to a value of 0.01 s^{-1} and repeat the analysis. In doing this on the same simulated data as above, we acquire estimates of the equilibrium constants that have the correct order of magnitude but are as much as 25% higher than the values used to generate the data (Table 2, second column). Moreover, the RMSD is significantly larger than the RMSD determined when allowing the rate constants to float. This cannot simply be the consequence of having three additional floating parameters because the degrees of freedom (DOFs) between the fits are essentially identical. That is to say, since the DOF is the difference between the number of data points and the number of parameters and here the number of data points is $\sim 80,000$, the additional three parameters do not influence this number enough to account for the deviation in the RMSD.

Despite the fact that constraining the rate constants to the empirical upper bound for “fast dissociation” leads to an $\sim 25\%$ overestimate of the equilibrium constant, the observation that both fits lead to the correct order of magnitude in the equilibrium constant suggests that the data could be modeled with a purely thermodynamic model. This leads to the suggestion that a path-independent thermodynamic model could be used to describe the data. To test this, the data were fit by choosing Newton–Raphson under the “Equilibration during run” in the “No analytic solution” row. What this accomplishes is numerically solving the sixth-order polynomial in the free monomer concentration to yield the concentration of species, thereby modeling the system under the assumption that each infinitely small slice of radial position is at equilibrium. This analysis yields equilibrium constants that are in good agreement with the values used to generate the data and the RMSD is identical to the value acquired when allowing the rate constants to float as fitting parameters (see Table 2, third column).



5. GLOBAL ANALYSIS USING THE 1-2-4-6 MODEL WITH SLOW DISSOCIATION OF OLIGOMERS

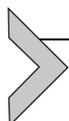
Figure 2B shows a $c(s)$ analysis from sedimentation velocity experiments simulated for the 1-2-4-6 model with all dissociation rate constants of 10^{-6} s^{-1} . Again, our first level of global analysis is to allow both the kinetic parameters, $k_{r,n}$, and the equilibrium constants, K_n , to float as fitting parameters. As seen in Table 3, the values of the equilibrium constants are in

good agreement with those used to simulate the data (compare Table 1 for simulated values to Table 3 for fitted values). However, the rate constants are as much as three orders of magnitude slower than the values used to simulate the data. This is not surprising because a dissociation rate constant of 10^{-6} s^{-1} yields a half-life of $\sim 192 \text{ h}$ if one assumes a simple first-order reaction. Thus, on the timescale of sedimentation, there is no appreciable dissociation that occurs for the simulated data with a given 10^{-6} s^{-1} rate constant.

When the data are analyzed with the rate constants constrained to 10^{-5} s^{-1} , the values of the equilibrium constants determined are in agreement with the values used to simulate the data within 2–9% errors (see Table 3). However, the RMSD is significantly worse than the value when the rate constants are allowed to float as fitting parameters. This is surprising because with such a slow dissociation rate constant, one does not expect the data to contain any information on these parameters. One possible explanation is that the empirical bound of 10^{-5} s^{-1} may be on the edge of values that would define no dissociation for the simulated system.

To probe this further, the data were analyzed by constraining all $k_{r,n} \leq 10^{-6} \text{ s}^{-1}$. As shown by the analysis performed when allowing $k_{r,n}$ to float, $k_{r,n} = 10^{-6} \text{ s}^{-1}$ is slow enough to indicate no dissociation for the simulated system. Therefore, rate constants smaller than 10^{-6} s^{-1} should be able to represent reactions with no dissociation and describe the simulated data adequately well. Our results show that an $\text{RMSD} = 5.000 \times 10^{-3}$ was determined when constraining all $k_{r,n} = 10^{-6} \text{ s}^{-1}$ and an $\text{RMSD} = 5.003 \times 10^{-3}$ was determined when constraining all $k_{r,n} = 10^{-7} \text{ s}^{-1}$. Both analyses yield identical equilibrium constants and are identical to the value used to generate the data.

Above all, the analyses suggest that the empirical boundary for defining “no dissociation” is in the range of 10^{-6} – 10^{-5} s^{-1} . We suspect that for systems with different levels of complexity and sizes of species, the empirical boundaries for rapid and slow dissociation may be slightly larger than the reported 10^{-2} – 10^{-5} s^{-1} range. Based on these observations, we recommend testing the reaction kinetics by allowing the rate constants to float as fitting parameters in the analysis.



6. GLOBAL ANALYSIS WITH RATE CONSTANTS IN THE DETECTABLE RANGE

The next and most obvious question is how well can the rate constants be determined if they fall between the empirical bound of 10^{-2} – 10^{-5} s^{-1} ?

Table 4 Parameters Used to Simulate the Sedimentation Velocity Data for 1–15 μM Macromolecule with Rate Constants Indicated in the Table

Model Used for Simulation	Values Used for Simulation		Parameters Resulting from Analysis	
	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$	$K_n (M^{-1})$	$k_{r,n} (s^{-1})$
$2B \xrightleftharpoons[k_{r1}]{k_{f1}} B_2$	6×10^4	3×10^{-3}	6.00×10^4	3.04×10^{-3}
$2B_2 \xrightleftharpoons[k_{r4}]{k_{f4}} B_4$	1.5×10^5	4×10^{-4}	1.50×10^5	3.98×10^{-4}
$B_2 + B_4 \xrightleftharpoons[k_{r6}]{k_{f6}} B_6$	1.5×10^6	7×10^{-4}	1.50×10^6	7.01×10^{-4}

The stoichiometric equilibrium constants for dimerization, tetramerization, and hexamerization can be calculated using Eqs. (25)–(27) and are: $L_{2,0} = 6 \times 10^4 M^{-1}$, $L_{4,0} = 5.4 \times 10^{14} M^{-3}$, and $L_{6,0} = 4.86 \times 10^{25} M^{-5}$.

To address this question, we simulated data using the 1–2–4–6 model with equilibrium constants and rate constants given in Table 4. As shown in Table 4, the rate constants are in the range of 10^{-3} – $10^{-4} s^{-1}$. Table 4 shows the results of globally fitting these simulated data to the 1–2–4–6 model. It is not surprising that the data are well described by the correct model and the predicted parameters are well within range of the values used to simulate the data. Although not surprising, one might expect that the information on the kinetic rate constants could be lost due to the complexity of the model. Thus, what this analysis does show is that the global fitting strategy is able to detect these rate constants for this model.

Thermodynamics is path independent. Sedimentation velocity data where the oligomers are either in rapid dissociation or slow dissociation do not contain information about path. Although the specific strategy for doing so is different due to the process of sedimentation, in the two extremes fast and slow kinetics the data can be described by either stoichiometric or stepwise equilibrium constants when allowing the kinetic parameters to float and the results are the same. This is not the case for a system that exhibits dissociation with rate constants in the range of 10^{-2} – $10^{-5} s^{-1}$. This is shown in Table 5 where the simulated data were analyzed assuming that each oligomer forms in a single step from monomers as given by Eqs. (3)–(8) where monomers form dimers, tetramers, or hexamers in a single kinetic step. The analysis of these data does not accurately predict the equilibrium constants or

Table 5 Analysis of Data Simulated in Table 4 Using “1-2, 1-4, 1-6” Model to Analyze the Data

Model Used for Fitting	Fit with Kinetic Integrator Allow $k_{r,n}$ to Float		
	RMSD = 5.539×10^{-3}		
	$L_{n,0}$	Calculated $K_n (M^{-1})$	$k_{r,n} (s^{-1})$
$L_{2,0}$ $2B \rightleftharpoons B_2$	$6.24 \times 10^4 M^{-1}$	6.24×10^4	1.88×10^{-3}
$L_{4,0}$ $4B \rightleftharpoons B_4$	$3.91 \times 10^{14} M^{-3}$	1.00×10^5	1.16×10^{-4}
$L_{6,0}$ $6B \rightleftharpoons B_6$	$5.64 \times 10^{25} M^{-5}$	2.31×10^6	1.80×10^{-4}

the rate constants, with the exception of dimer formation which is not different for the two models. What this analysis reveals is that when the rate constants are in the detectable range, there is information on the path and the data cannot be modeled by simple path-independent thermodynamic models. Thus, under these conditions care needs to be taken to write down appropriate path-dependent models. Although it may take more computational power to do so, we assert that this should be done by assuming all reactions occur through bimolecular interactions.

7. CONCLUSIONS

Similar simulations and examination of the kinetics as discussed here have also been discussed elsewhere (Correia & Stafford, 2009; Dam et al., 2005). The advance presented here is an analysis of a more complex model. Most previous discussions have centered on a monomer–dimer equilibrium. Here, we have focused on the monomer, dimer, tetramer, and hexamer equilibrium since this is what we are experimentally observing for two AAA+ molecular motors, *E. coli* ClpA (Veronese & Lucius, 2010; Veronese et al., 2011; Veronese et al., 2009) and ClpB (Li et al., 2015; Lin & Lucius, 2015). Most importantly, there are some general principles that have been derived from previous studies on less complex systems that do not seem to generalize to the more complex systems we are studying. That is to say, the empirical bound of $0.01\text{--}10^{-5} s^{-1}$ may be a bit wider depending on the size of molecules and the rotor speed.

One observation that holds for all values of the kinetic rate constants is that floating the values always seems to lead to acquisition of the values used

to simulate the data. So why not always float the kinetic rate constants? The answer is the lack of computational power. Indeed, 20 years ago, globally analyzing upward of 80,000 data points by numerically solving the Lamm equation combined with numerically solving a complex system of coupled differential equations was outside of our computational reach. However, we now have desktop computers that can accomplish these tasks. Nevertheless, fitting to such complex models is very slow and it can be accelerated when limiting assumptions can be made.

One limiting factor, which is true for all NLLS minimization routines, is that each computation is dependent on the outcome of the last. Thus, such approaches are not amenable to parallel computing. In contrast, routines such as the genetic algorithm are gaining interest because of the independence of each calculation and thus the ease with which such routines can be parallelized. Ultrascan is an application for the analysis of sedimentation velocity data that take advantage of the genetic algorithm and thus takes advantage of advances in parallel computing (Demeler et al., 2010).

Sedphat does not allow the user to write their own models for global analysis. Thus, the experimentalist is constrained to use prewritten models that assume, for example, that tetramers would form in a single kinetic step. One could make an argument for trimolecular interactions but anything higher than three bodies colliding in a single kinetic step is unrealistic. Although Ultrascan allows the user to write down their own models, it does not allow the product of one reaction to be the reactant of another. As shown by the simulations presented here, there is a need to be able to write down kinetic models that reduce the reactions to bimolecular steps when kinetic information is present.

To our knowledge, the only software application that allows the experimentalist to construct their own models and directly fit boundaries by numerically solving the Lamm equation is SEDANAL (Stafford & Sherwood, 2004). More importantly, SEDANAL is numerically solving a system of coupled differential equations describing the reaction. Thus, the rate constants that emerge have the potential to contain a great deal more information about the reaction than arbitrary rate constants that are being applied in other applications. We assert that the gravity of what SEDANAL is doing behind the scenes is not fully appreciated. Indeed, sedimentation velocity would not be the experiment of choice if one were trying to thoroughly deconvolute a kinetic mechanism. However, not adequately accounting for the kinetics in the analysis of these data may lead to significant errors in the equilibrium parameters of interest. Moreover, it is clear that

sedimentation velocity experiments can lead the experimentalist to pursue more direct kinetic approaches if something interesting is revealed in the modeling.

REFERENCES

- Alberty, R. A. (2003). *Thermodynamics of biochemical reactions*. Hoboken, NJ: Wiley Interscience.
- Babst, M., Sato, T. K., Banta, L. M., & Emr, S. D. (1997). Endosomal transport function in yeast requires a novel AAA-type ATPase, Vps4p. *The EMBO Journal*, *16*(8), 1820–1831.
- Cole, J. L. (1996). Characterization of human cytomegalovirus protease dimerization by analytical centrifugation. *Biochemistry*, *35*(48), 15601–15610.
- Cole, J. L. (2004). Analysis of heterogeneous interactions. *Methods in Enzymology*, *384*, 212–232.
- Cole, J. L., Correia, J. J., & Stafford, W. F. (2011). The use of analytical sedimentation velocity to extract thermodynamic linkage. *Biophysical Chemistry*, *159*(1), 120–128.
- Correia, J. J. (2000). Analysis of weight average sedimentation velocity data. *Methods in Enzymology*, *321*, 81–100.
- Correia, J. J., Alday, P. H., Sherwood, P., & Stafford, W. F. (2009). Effect of kinetics on sedimentation velocity profiles and the role of intermediates. *Methods in Enzymology*, *467*, 135–161.
- Correia, J. J., & Stafford, W. F. (2009). Extracting equilibrium constants from kinetically limited reacting systems. *Methods in Enzymology*, *455*, 419–446.
- Dam, J., & Schuck, P. (2005). Sedimentation velocity analysis of heterogeneous protein-protein interactions: Sedimentation coefficient distributions $c(s)$ and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophysical Journal*, *89*(1), 651–666.
- Dam, J., Velikovskiy, C. A., Mariuzza, R. A., Urbanke, C., & Schuck, P. (2005). Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions $c(s)$. *Biophysical Journal*, *89*(1), 619–634.
- Demeler, B., Brookes, E., Wang, R., Schirf, V., & Kim, C. A. (2010). Characterization of reversible associations by sedimentation velocity with UltraScan. *Macromolecular Bioscience*, *10*(7), 775–782.
- Dougan, D. A., Reid, B. G., Horwich, A. L., & Bukau, B. (2002). ClpS, a substrate modulator of the ClpAP machine. *Molecular Cell*, *9*(3), 673–683.
- Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, *6*(7), 1948–1954.
- Farrell, C. M., Grossman, A. D., & Sauer, R. T. (2005). Cytoplasmic degradation of ssrA-tagged proteins. *Molecular Microbiology*, *57*(6), 1750–1761.
- Gill, S. C., & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry*, *182*(2), 319–326.
- Laue, T. M., Shah, B. D., Ridgeway, T. M., & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In A. J. Rowe, S. E. Harding, & J. C. Horton (Eds.), *Analytical ultracentrifugation in biochemistry and polymer science*. Cambridge: Royal Society of Chemistry.
- Li, T., Lin, J., & Lucius, A. L. (2015). Examination of polypeptide substrate specificity for *Escherichia coli* ClpB. *Proteins*, *83*(1), 117–134.
- Lin, J., & Lucius, A. L. (2015). Examination of the dynamic equilibrium for *E. coli* ClpB. Submitted to *Proteins*.
- Lohman, T. M., & Mascotti, D. P. (1992). Nonspecific ligand-DNA equilibrium binding parameters determined by fluorescence methods. *Methods in Enzymology*, *212*, 424–458.

- Meyer, H., & Wehl, C. C. (2014). The VCP/p97 system at a glance: Connecting cellular function to disease pathogenesis. *Journal of Cell Science*, 127(Pt. 18), 3877–3883.
- Mogk, A., Tomoyasu, T., Goloubinoff, P., Rudiger, S., Roder, D., Langen, H., et al. (1999). Identification of thermolabile *Escherichia coli* proteins: Prevention and reversion of aggregation by DnaK and ClpB. *The EMBO Journal*, 18(24), 6934–6949.
- Na, G. C., & Timasheff, S. N. (1985a). Measurement and analysis of ligand-binding isotherms linked to protein self-associations. *Methods in Enzymology*, 117, 496–519.
- Na, G. C., & Timasheff, S. N. (1985b). Velocity sedimentation study of ligand-induced protein self-association. *Methods in Enzymology*, 117, 459–495.
- Neuwald, A. F., Aravind, L., Spouge, J. L., & Koonin, E. V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Research*, 9(1), 27–43.
- Ogura, T., & Wilkinson, A. J. (2001). AAA+ superfamily ATPases: Common structure–diverse function. *Genes to Cells*, 6(7), 575–597.
- Ortega, A., Amoros, D., & Garcia de la Torre, J. (2011). Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophysical Journal*, 101(4), 892–898.
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Science*, 4(11), 2411–2423.
- Roll-Mecak, A., & McNally, F. J. (2010). Microtubule-severing enzymes. *Current Opinion in Cell Biology*, 22(1), 96–103.
- Sauer, R. T., & Baker, T. A. (2011). AAA+ proteases: ATP-fueled machines of protein destruction. *Annual Review of Biochemistry*, 80, 587–612.
- Schuck, P. (1998). Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. *Biophysical Journal*, 75(3), 1503–1512.
- Schuck, P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Analytical Biochemistry*, 320(1), 104–124.
- Scott, D. J., Harding, S. E., Rowe, A. J., & Royal Society of Chemistry (Great Britain). (2005). *Analytical ultracentrifugation: Techniques and methods*. Cambridge, UK: RSC Publications.
- Stafford, W. F., & Sherwood, P. J. (2004). Analysis of heterologous interacting systems by sedimentation velocity: Curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. *Biophysical Chemistry*, 108(1–3), 231–243.
- Streaker, E. D., Gupta, A., & Beckett, D. (2002). The biotin repressor: Thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific DNA binding. *Biochemistry*, 41(48), 14263–14271.
- Veronese, P. K., & Lucius, A. L. (2010). Effect of temperature on the self-assembly of the *Escherichia coli* ClpA molecular chaperone. *Biochemistry*, 49(45), 9820–9829.
- Veronese, P. K., Rajendar, B., & Lucius, A. L. (2011). Activity of *Escherichia coli* ClpA bound by nucleoside di- and triphosphates. *Journal of Molecular Biology*, 409(3), 333–347.
- Veronese, P. K., Stafford, R. P., & Lucius, A. L. (2009). The *Escherichia coli* ClpA molecular chaperone self-assembles into tetramers. *Biochemistry*, 48(39), 9221–9233.
- Wong, I., & Lohman, T. M. (1995). Linkage of protein assembly to protein-DNA binding. *Methods in enzymology*, 259, 95–127.
- Wyman, J., & Gill, S. J. (1990). *Binding and linkage: Functional chemistry of biological macromolecules*. Mill Valley, CA: University Science Books.
- Yu, R. C., Jahn, R., & Brunger, A. T. (1999). NSF N-terminal domain crystal structure: Models of NSF function. *Molecular Cell*, 4(1), 97–107.