

*promoting access to White Rose research papers*



**Universities of Leeds, Sheffield and York**  
**<http://eprints.whiterose.ac.uk/>**

---

This is an author produced version of a paper published in **Journal of Molecular Biology**.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/4152/>

---

**Published paper**

Tuma, R., Tsuruta, H., French, K.H. and Prevelige, P.E. (2008) *Detection of intermediates and kinetic control during assembly of bacteriophage P22 procapsid*. *Journal of Molecular Biology*, 382 (5), pp. 1395-1406.

DOI: [10.1016/j.jmb.2008.06.020](https://doi.org/10.1016/j.jmb.2008.06.020)

---

# Detection of intermediates and kinetic control during assembly of bacteriophage P22 procapsid

Roman Tuma<sup>1,2\*</sup>, Hiro Tsuruta<sup>3</sup>, Kenneth H. French<sup>1</sup>, and Peter E. Prevelige<sup>1</sup>

<sup>1</sup>Department of Microbiology University of Alabama at Birmingham, USA

<sup>2</sup>Astbury Centre for Structural Molecular Biology, University of Leeds, U.K. and Institute of Biotechnology, University of Helsinki, Finland

<sup>3</sup>Stanford Synchrotron Radiation Laboratory, Stanford Linear Accelerator Center, Stanford University, Menlo Park, CA94025-7015, USA

Running title: Intermediates of dsDNA virus assembly

Keywords: self-assembly, virus, X-ray, kinetics, structure, coat protein, scaffolding protein

\*To whom correspondence should be addressed: r.tuma@leeds.ac.uk

## Abstract

Bacteriophage P22 serves as a model for the assembly and maturation of other icosahedral double-stranded DNA viruses. P22 coat and scaffolding proteins assemble in vitro into an icosahedral procapsid, which then expands during DNA packaging (maturation). Efficient in vitro assembly makes this system suitable for design and production of monodisperse spherical nanoparticles (diameter  $\approx 50$  nm). In this work we explore the possibility of controlling the outcome of assembly by scaffolding protein engineering. The scaffolding protein exists in monomer-dimer-tetramer equilibrium. We address the role of monomers and dimers in assembly by using three different scaffolding proteins with altered monomer-dimer equilibrium (weak dimer, covalent dimer, monomer). The progress and outcome of assembly was monitored by time-resolved X-ray scattering which allowed us to distinguish between closed shells and incomplete assembly intermediates. Binding of scaffolding monomer activates the coat protein for assembly. Excess dimeric scaffolding protein resulted in rapid nucleation and kinetic trapping yielding incomplete shells. Addition of monomeric wild type scaffold with excess coat protein completed these metastable shells. Thus, the monomeric scaffolding protein plays an essential role in the elongation phase by activating the coat and effectively lowering its critical concentration for assembly.

## Introduction

A typical viral capsid is composed of an outer protein shell that contains and protects the viral nucleic acid. Simple viral shells may be made of multiple copies of a single coat protein<sup>1</sup>. Virus assembly is a process in which well organized, usually highly symmetric (icosahedral or helical), viral capsids are built by association of the protein subunits<sup>2</sup>. The advent of nanotechnology has brought renewed interest in viruses and their assembly<sup>3</sup>. While understanding virus assembly is crucial to tackling viral diseases<sup>4; 5</sup> it also provides a basis for the exploitation of viruses in the design of new nanostructures<sup>6; 7</sup>. Current nanotechnology applications rely on modified viral coat proteins that assemble native-like icosahedral shells and serve as attachment scaffolds for other molecules or assemblies. However, advanced applications of viral capsids are likely to require purposeful control over the assembly process. For example, one would like to assemble capsids within nanofabricated compartments (e.g. lab-on-chip applications) in a controllable fashion. This will require careful control of assembly initiation by chemical or physical means, e.g. gradual build up of concentration or surface immobilized nucleation sites. Similarly, control over the outcome of assembly may be needed when partially assembled shells are desired or if incorporation of a bulky cargo inside the viral shell would require slower, gradual polymerization. Achieving such a level of control depends on a detailed understanding of the assembly process.

Several bacteriophages and viruses had been reconstituted *in vitro* from purified components and the morphogenetic pathways for many bacteriophages had been delineated using ingenious genetic schemes<sup>8</sup>. However, detailed kinetic analyses have been achieved for only a few of the simplest capsids which are usually composed of one type of subunit<sup>9; 10</sup>. Among the more complex systems, in which two or more proteins co-assemble, bacteriophage P22 assembly is perhaps the best understood. The assembly of bacteriophage P22 has been extensively studied *in vivo* by both genetic and structural means<sup>11; 12; 13</sup>. These studies have

revealed that the capsid assembles in two steps. First, an empty procapsid results from co-assembly of the major coat protein (Gp5, 47 kDa, abbreviated CP) with the scaffolding protein (Gp8, 34 kDa, abbreviated SP) and several minor capsid proteins<sup>14</sup>. Then the procapsid shell undergoes a large conformational change and expands during DNA packaging<sup>15; 16</sup>. An *in vitro* system for procapsid reconstitution had been established<sup>17</sup> and made possible detailed kinetic studies of the assembly process<sup>18</sup>. The *in vitro* studies demonstrated that CP and SP are essential as well as sufficient for the formation of isometric procapsid shells (Fig. 1A). Kinetic analyses indicated that PC assembly proceeds in two phases, a rate-limiting nucleation phase is followed by fast growth leading to completion of the shells. The putative nucleation complex was proposed to consist of a coat protein pentamer together with several SP molecules. Subsequent studies with different scaffolding protein mutants pointed out the prominent role of scaffolding dimer in assembly<sup>19</sup>.

Scaffolding protein exists in a monomer-dimer-tetramer equilibrium in solution<sup>20</sup> and their respective roles in assembly were proposed on the basis of the final PC structure<sup>21; 22; 23</sup>. The scaffolding protein can be extracted from the procapsid by either mild denaturant (GuHCl, urea)<sup>17</sup>, salt<sup>24</sup> or heat<sup>25</sup> treatment without affecting the structure<sup>23</sup>. This suggests that the scaffolding protein's role is in the formation of transient complexes during nucleation and growth phases. Given that the final structure provides only limited information on the early steps of assembly the role of different scaffolding protein forms in these steps remains elusive. In particular, the role of scaffolding protein monomer has not been elucidated.

This study assesses the role of different SP oligomers in assembly by time-resolved small angle X-ray scattering (TR-SAXS). This technique together with singular value decomposition (SVD)<sup>26</sup> enabled us to resolve complete shells from intermediates and aberrant products and to follow assembly under stoichiometric as well as sub-stoichiometric conditions. In addition, we take advantage of various engineered scaffolding proteins (Fig.

1B) and delineate the role of the monomeric and the dimeric scaffolding protein in assembly. The results demonstrate that assembly rate and outcome can be controlled by manipulating the scaffolding protein dimerization equilibrium and coat protein concentrations.

## **Results**

### ***Wild type scaffolding protein***

#### **SAXS based assembly assay**

Monomeric coat protein and wild type scaffolding protein were manually mixed in two molar ratios (CP:SP 1:1 and 2:1, respectively) and their assembly process was monitored by SAXS (Fig. 2A and 3). The manual mixing procedure and activating the X-ray hutch interlock system resulted in about 1 min dead time. The time-dependence of the scattering revealed the gradual emergence of maxima and minima characteristic of closed shells. After 25 min the reaction reached a stationary phase with little further change. The observed scattering curve was modeled by calculating the scattering from hollow spheres with outer and inner radii 310 and 160 Å, respectively (Fig. 2B). The model fits well for  $q < 0.04 \text{ \AA}^{-1}$  above which the departure of the icosahedral shell from the spherical model becomes apparent. The thickness of the shell (150Å) includes the density from both coat and scaffolding protein. The overall dimensions are consistent with those derived from cryo-EM<sup>22; 23</sup>.

#### **Detection and structural characterization of assembly intermediates**

The TR-SAXS data were plotted as a three dimensional surface (Fig. 2A). Given the relatively high concentrations of CP and SP required for SAXS experiments any intermediates would be expected to appear early during the assembly and would quickly grow into complete shells which would dominate scattering during the later stages of assembly. Due to their low abundance and small size any scattering arising from intermediates would be quite weak<sup>18</sup>. In

order to examine whether the TR SAXS data contains contributions from any putative intermediates we employed singular value decomposition (SVD) analysis. This mathematical procedure yields the number of linearly independent and orthogonal scattering components in TR-SAXS at a given noise level<sup>27</sup> (See Supplementary Methods for detailed description). If there were no intermediates then the data set would be fully described (to within experimental noise) by two components, namely scattering from the initial mixture of CP and SP and from the final shells, respectively. Such least squares procedure failed to reproduce the data and significant, time-resolved residuals remained (Supplementary Methods Fig. S1). Hence, the data contained a significant scattering contribution from one or more assembly intermediate in addition to the scattering contributed by the initial and final species. In fact, three components (corresponding to the three highest SVD eigenvalues, Supplementary Methods Fig. S2) were necessary to reproduce the experimental data to within the noise level (Supplementary Methods Fig. S3). Because SVD yields orthogonal scattering and concentration factors they alternate between positive and negative values (Supplementary Methods Fig. S4) and do not directly represent physically meaningful scattering curves or concentrations. However, the SVD scattering vectors represent a convenient orthogonal basis for obtaining a plausible and physically meaningful set of scattering curves as a linear combination. Constrained least-squares method and manual rotation were used to obtain scattering factors and concentration loadings which were non-negative to within the noise level (see <sup>28</sup> and Supplementary Methods for further details).

Semi-quantitative scaling of scattering intensities (i.e. relative to each other) could be achieved by extrapolating the scattering factors to zero angle using the Guinier approximation<sup>29; 30</sup>. In order to hold for shell-sized scattering objects the Guinier approximation would require a significant portion of the experimental data to extend below  $q < 0.005 \text{ \AA}^{-1}$  (i.e. within the so called Guinier region,  $R_g \cdot q \leq 2$ ). This was precluded by the limited experimental

range ( $q > 0.008$  for manual mixing data and  $q > 0.012$  for stopped-flow experiments). In order to overcome this limitation we assume that the observed assembly intermediates are partial shells with rotational symmetry around the z-axis (i.e. spherical section inscribed by a cone with vertex angle  $2\Theta$ , Fig. 4A). This is a reasonable approximation since other smaller species would not dominate in the scattering signal due to their sizes and would be less stable<sup>31</sup>. Then the scattering curves for such partial spherical shells were simulated for various degree of completion (i.e.  $\Theta$  from 0 to 180 deg, Fig. 4B). The true and apparent  $R_g$  values for each degree of completion were computed from these curves using the two experimentally accessible  $q$  ranges (Fig. 4C). Comparison of the simulated curves show that the extended Guinier region may be used for intermediates up to 45% completion i.e. apparent  $R_g$  values up to 190 Å. Hence, we used the extended Guinier region to estimate the  $R_g$  and the degree of completeness, i.e. the relative apparent mass of the intermediate (from Fig. 4C). This relative mass was then used to scale the intensities of the scattering factors and amplitudes of relative mass concentrations (Fig. 3).

Note that the assignment of single apparent  $R_g$  and mass to a given scattering factor does not mean this represents a single intermediate. This could hardly be expected for a reaction of such complexity and consequently the values characterize an ensemble of intermediates. The apparent mass represents a weight average value while the apparent  $R_g$  is the Z-average across the distribution<sup>32</sup>. Both of these averages are biased towards larger species and thus these parameters effectively represent the upper bound of the distribution.

An apparent  $R_g$  of  $155 \pm 25 \text{Å}$  (lower initial coat concentration, 1:1 CP:SP ratio) and  $125 \pm 30 \text{Å}$  (higher coat concentration with 2:1 CP:SP ratio, i.e similar to the stoichiometry found in native procapsids) was obtained for intermediates from the data in Fig. 3A and Fig. 3B, respectively. The apparent  $R_g$  values as well as the scattering curves agreed within experimental error for the two conditions. This suggests that in both cases the largest

intermediates were approximately of the same size corresponding to about 25-35% completeness (Fig. 4C). The scattering curves of the intermediates (red curves in Fig 3) are also similar to the 30%-complete shell model (Fig. 4B). However, the kinetics exhibited distinct behavior for the two CP concentrations, namely the maximum in the intermediate concentration (representing 5% total mass fraction) was attained at about 65-100 s for the higher CP concentration while the peak (8% total mass fraction) was delayed to 210 s for the lower concentration. Because of their relatively large size and preferential sensitivity of SAXS for larger species these intermediates are still detectable despite their low abundance. On the other hand their occurrence precluded detection of smaller species.

The 30%-complete shells contain roughly 86 CP subunits, i.e. much larger number than the pentameric nucleation complex<sup>18</sup>. Similarly the  $R_g$  values are also larger than those expected for the nucleation complex (~60-70Å, estimated  $R_g$  for a pentameric complex). This indicates that smaller intermediates are either not significantly populated during assembly or are populated only during the dead time of manual mixing experiments. Hence we resorted to rapid mixing techniques to resolve early complexes of coat and scaffolding proteins which may lead to the formation of assembly nuclei.

### Stopped-flow kinetics

Stopped-flow mixing was achieved with a 2-syringe apparatus with an incorporated scattering flow cell<sup>33</sup>. In order to achieve the necessary sensitivity towards weakly scattering early intermediates (e.g. pre-nucleation complexes such as CP<sub>2</sub>:SP<sub>2</sub> dimer) high concentrations of both CP (27.2 μM, the maximum achievable concentration for the monomeric, assembly competent CP) and SP (22.9 μM) were used. The first 5 min of the assembly kinetics were sampled using 5 to 10 s exposures (Fig. 5A). SVD analysis revealed 3 significant factors which were then rotated to obtain non-negative, physically meaningful scattering factors (Fig. 5B top) and the associated relative concentrations (Fig. 5B bottom).

The time course for the appearance of completed shells exhibited a noticeable lag time (70 s) previously assigned to the nucleation phase<sup>18</sup>. During this phase a considerable amount of coat protein was incorporated into intermediates. An apparent  $R_g$  of  $170 \pm 50$  Å was obtained from the initial slope ( $q$  values  $0.012$ - $0.016$  Å<sup>-1</sup>) and the scattering factor in Fig 5B resembled that of shells with completeness between 20-40%. Once again, the size of these intermediates represents the upper bound of intermediate ensemble due to the  $Z$ -averaging nature of scattering intensities. This suggests that the rate limiting nucleation step is followed by a rapid addition of coat/scaffold subunits to the partial shells until the number of sites for addition becomes large and the decreasing concentration of CP limits the rate of addition of coat protein to the existing partial shells (elongation rate) and prevents shell closure. Under those conditions the reaction is prone to kinetic trapping. Substantial accumulation of these intermediates suggests that at the initial high CP and SP concentrations nucleation dominates over the elongation.

After the lag phase the production of shells increased linearly over time until approximately 250 s. During this phase the intermediate concentration peaked at about 150 s (at 10% total mass abundance) after which the intermediates were completed into shells while nucleation of new assemblies was limited by the decreased CP and SP concentrations. Finally, the pools of free coat protein and partially assembled intermediates become exhausted and the assembly reaction enters steady state phase after approximately 250 s.

Comparison of the TR-SAXS data collected under different protein concentrations (Figs. 3 and 5) revealed that an increase in CP concentration speeds up the rate of elongation while increasing SP concentration leads to a longer lag phase for complete shell formation and thus favors nucleation.

### ***Scaffolding protein dimer***

The scaffolding protein dimer has been implicated as the active species in P22 assembly. Here we took advantage of a reversible disulfide crosslink in a SP mutant (R74C/L177I, Fig. 1B) which leads to a covalent dimer formation<sup>19</sup>. We address the question of whether the SP dimer alone was sufficient for assembly. Although the assembly reactions were done under similar concentrations as in the case of wt SP it is evident that the dimer alone is not sufficient for closed shell assembly in vitro (Fig 6A). In a control experiment the reduced form of the mutant SP exhibited assembly similar to the wt SP (Fig. 6D). This indicates that the defect is due to SP dimerization and not due to the mutation alone.

From the analysis of the assembly reaction, the final products had apparent  $R_g$  ranging from 80 to 180 Å depending on the initial CP concentration (Fig. 6B-C). Although higher CP concentrations led to apparently larger assemblies and kinetics exhibiting apparent intermediates (Fig. 6C), the final product did not contain any closed shells. The maxima and minima typical of scattering from spherical shells were not detected in any of the scattering factors. Under both CP concentrations the largest assembly product exhibited an apparent  $R_g$  of 180Å which corresponds to 40% complete shells (Fig. 4C) provided these products were intermediates on the procapsid assembly pathway. Alternatively these products could be aberrant assemblies.

To distinguish between these two possibilities the distribution of assembled species obtained under different conditions were compared using sucrose gradients and electron microscopy (Fig. 7). In comparison with wt SP the dimer was more efficient in inducing coat protein assembly (c.f. CP amount in the top, 5% sucrose gradient fractions for rows 1 and 2 on the gel). However, the dimer induced assembly of broadly-distributed, smaller (lighter) products which appeared to be incomplete shells in negative stain electron micrograph (Fig. 7 right). Addition of wt SP to the dimer reaction restored the product distribution to almost the

wild type pattern (c.f. rows 1 and 3). The addition of excess CP and simultaneous decrease of SP dimer concentration had a similar effect (row 5) while reducing both CP and SP concentrations yielded less complete shells (row 4). The incomplete assemblies were rescued by the addition of fresh CP after reaching the stationary phase with the dimer SP (not shown). This demonstrates that the products of assembly reaction with R74C dimer are PC assembly intermediates.

The SAXS and sucrose gradient results suggest that the SP dimer is more effective in nucleating assembly than the wt SP (monomer/dimer equilibrium) and decreases the critical concentration of coat protein (most CP had assembled in the presence of R74C SP). Hence, the incomplete shells constitute kinetically trapped intermediates which are produced by over-nucleation of assembly by the dimeric scaffold. From the rescue of assembly by wt SP it is clear that there is a role for the scaffold monomer, most likely during the elongation phase.

### ***Scaffolding protein monomer***

The C-terminal 38 amino acid residues of the scaffolding protein encompass the coat protein binding site<sup>34</sup>. The C-terminal fragment (CTF) lacks the dimerization domain and thus remains monomeric<sup>19; 35</sup>. Despite the lack of dimerization the monomeric CTF induced rapid assembly (Fig. 8A). However, the products were polydisperse and lacked scattering features typical of complete shells, i.e. maxima and minima. A broad range of products was found by sucrose gradient analysis (Fig. 8B). This shows that SP monomer switches the coat protein into an assembly competent form and effectively decreases the critical concentration for assembly.

## Discussion

### *Three phases in assembly*

In our TR-SAXS study, we have resolved a nucleation phase during which rapid initiation of assembly leads to accumulation of partial shell intermediates. For wt SP this phase leads to the establishment of steady state during which the nucleation rate matches the rate of shell completion and the amount of PC increases approximately linearly with time. Eventually, coat protein gets depleted and the reaction enters stationary phase during which there is no net increase in assembly products. Stationary phase is particularly pronounced for the kinetically trapped assembly reactions of SP covalent dimer (Fig. 6). This phase should not be confused with the final equilibrium since kinetically trapped intermediates can be completed either by addition of coat protein (as shown here in Fig. 7) or by slow disassembly and reassembly<sup>36;37</sup>.

Experimental limitations (poor signal-to-noise at low protein concentrations) precluded examination of broader protein concentration range. However, qualitative conclusions can be drawn from the concentrations sampled. The duration of the nucleation phase, measured as the time at which the concentration of intermediates peaks, decreases with increasing initial concentration of CP (c.f. Fig. 3A and B) and increases upon doubling the SP concentration (i.e. increase in the fraction of dimer from 10 to 18%,<sup>20</sup>, c.f. Fig. 3 B and Fig. 5B). The covalent SP dimer accelerated nucleation, completely eliminated the steady-state phase and yielded only kinetically trapped partial shells. This could be due to its higher affinity for the shells. However, the covalent SP dimer also decreased the apparent critical coat concentration for assembly (note the lower amount of unassembled coat on the top of sucrose gradients in Fig. 7 for the crosslinked R74C) pointing out its direct role in nucleation

The partial shells could be completed by addition of excess coat protein or by including even small amount of wt SP in the reaction (i.e. monomeric SP). This suggests that efficient elongation is accelerated by monomeric SP. Our observation that the monomeric scaffolding fragment CTF promotes coat aggregation suggests that SP monomer may activate CP for assembly.

### ***Activation of coat for assembly by scaffolding monomer***

The activation of CP by SP constitutes a step common to both the nucleation and elongation phases. This activation relies on a direct CP-SP interaction which is predominantly electrostatic<sup>24</sup>. Low salt would be expected to promote CP-SP interactions and thus lead to accelerated assembly. In fact, excessive nucleation and rapid assembly of partial shells was observed under low salt conditions<sup>38</sup>. These authors concluded that strengthening the electrostatic interactions between CP and SP promoted rapid nucleation and growth of shells<sup>38</sup>. In addition, they showed that low ionic strength led to conformational changes in CP but not in SP. The altered conformation may, perhaps, be more amenable to assembly. PC assembly is accompanied by significant change in CP secondary structure<sup>39</sup>. This and the partially unfolded character of CP monomer<sup>40</sup> suggest that CP monomer may exist in solution as an ensemble of conformations of which only a fraction is assembly competent. Binding of SP monomer may sway the equilibrium in favor of assembly competent conformations and thus activate the coat protein. This is reminiscent of bacteriophage MS2 coat activation for assembly by specific RNA binding<sup>41</sup>. However, in this case the conformational switching is involved in early stages of assembly and RNA packaging.

The assembly phenotypes observed by Parent et al at low ionic strength are remarkably similar to those observed here for the SP covalent dimer under standard ionic strength. This suggests that perhaps, in addition to affecting CP-SP interactions, the low salt

conditions may also affect the monomer-dimer equilibrium of the highly polar and charged SP and promote nucleation by excess of scaffolding dimer.

### ***Role of scaffolding monomer-dimer equilibrium in assembly***

It is clear from the above discussion that monomeric SP is required for assembly, most likely in the later stages. In fact, addition of wt SP to the reaction mixture (as low as 0.05 mg/ml i.e. 95% monomeric) also alleviated the over-nucleation induced kinetic trap caused by the covalent dimer (Fig. 7 and data not shown). This sub-stoichiometric effect suggests that the monomeric SP may transiently bind to CP, activate it and then dissociate instead of remaining stably associated with the resulting PC shell.

The monomeric CTF activates CP for assembly but unlike the dimer it is not stably incorporated into the products (c.f. Fig. 7 and Fig. 8B). This suggests that the affinity of the coat protein for the monomeric SP is low or may even decrease during the course of assembly while the affinity for the dimer is higher. Indeed, the procapsid exhibits two distinct types of binding sites for wt SP with different affinities which were assigned to monomeric and dimeric states, respectively <sup>42</sup>. Thus, the monomeric SP most likely transiently interacts with CP and activates it for assembly, effectively increasing the concentration of assembly competent CP and increasing the elongation rate (Fig. 9). This also competes with nucleation which is promoted by the SP dimer. Thus, the SP monomer-dimer equilibrium is ultimately linked to the nucleation-elongation balance and may be effective in steering the course and outcome of the assembly reaction (Fig. 9).

### ***P22 assembly differs from the structurally related bacteriophage HK97***

P22 coat protein has a similar fold to that of another dsDNA bacteriophage HK97 <sup>16</sup>. However, the latter utilizes a two-step assembly pathway in which relatively stable pentamers and hexamers serve as the building blocks for the prohead (procapsid equivalent in HK97)

formation<sup>1</sup>. Thus, coat protein fold similarity and identical capsid architecture may not automatically result in the same assembly pathway. The different assembly pathways may be due to the radically different interaction between coat and scaffolding proteins. While P22 SP interacts with CP non-covalently, the HK97 has its scaffolding element, the N-terminal 110-residue  $\Delta$ -domain, integrated within the coat protein. The  $\Delta$ -domain is proteolytically cleaved within the assembled prohead and subsequently released. Very little is known about the interactions between the putative scaffolding domains in HK97. On the other hand, coat-coat interactions, which are weak for P22, dominate the early assembly of HK97 (i.e. pentamer/hexamer formation). Hence, the final structure of a virus may give us little information about the assembly mechanism unless a quantitative survey of protein-protein interactions is performed and related to the solution conditions. The latter are important because assembly is governed by a set of relatively weak protein-protein interactions<sup>43</sup> and these in turn are extremely sensitive to solution conditions such as ionic strength and pH.

### ***Kinetic control over assembly in vitro and implications for in vivo***

The detectable amounts of P22 assembly intermediates also demonstrate that under the conditions used here the assembly reactions are far from equilibrium. Even under those conditions viral shells are being produced when wild type SP is present indicating that the assembly process is quite robust. However, the results presented here and elsewhere<sup>38 31</sup> suggest that virus assembly is prone to kinetic trapping under conditions of strong protein-protein interactions. The common cause is either high concentration of subunits which leads to over-nucleation or imbalance of inter-molecular interactions (e.g. mutations). Specifically, in a system with several co-assembling components, such as P22, balance of the interactions and concentrations for all components plays essential role in successful assembly. Hence, the scaffolding protein association is relatively weak and dynamic, and coat protein associates only upon interaction with SP. In addition, the scaffolding protein concentration is being

maintained inside the infected cell by a translational control in which the N-terminal domain acts as auto-repressor<sup>44</sup>.

Recent theoretical studies demonstrate that the assembly mechanisms could abruptly switch from a productive mode to a non-productive one upon a slight shift in concentrations or strength of subunit interactions<sup>45</sup>. In general, conditions of slower growth are generally less prone to kinetic trapping<sup>31</sup> and weak protein-protein interactions represent one way to achieve this<sup>46</sup>. Alternatively, low protein concentrations need to be maintained but such conditions would interfere with the need for high protein production in a short time during infection, a condition applicable to most lytic viruses.

Although not desirable during the course of productive infection, kinetic trapping may be useful in the manufacture of virus-based nanoparticles. The step-wise completion of the shells in vitro as demonstrated here could be exploited for e.g. encapsidating guest material or incorporating selectively labeled subunits that may otherwise interfere with the nucleation step of assembly. In the P22 system the kinetic trap may be “programmed” at ease by manipulating the self-association of the highly soluble scaffolding protein while using the hard-to-obtain coat protein in its wild type, assembly-competent, form.

## **Materials and Methods**

### ***Protein purification***

Coat protein was purified fresh just before the assembly experiments to assure monodispersity as described previously<sup>39</sup>. The wt scaffolding protein was extracted from procapsids expressed in *E. coli* and purified as described<sup>20</sup>. Engineered scaffolding proteins were expressed in *E. coli* and purified as described<sup>19; 20</sup>. All proteins were centrifuged for 15 min at 14,000 g at 4°C just before use. The R74C/L177I disulfide dimer was reduced by 3.3 mM glutathione.

### ***In vitro assembly reactions***

Assembly reactions were started by mixing (either manually or by a stopped-flow mixer) a small amount (ca. 5-10 $\mu$ l) of concentrated scaffolding protein into excess of coat protein. All reactions were done in buffer B (50 mM Tris, 25 mM NaCl, 1 mM EDTA, pH 7.6) at 20°C. If needed assembly products were analyzed by sedimentation velocity on 5-20% (w/v) sucrose gradient (50 $\mu$ l loading per tube). The gradient was run in SW55 rotor at 20°C for 55 min and fractionated into 250 $\mu$ l fractions which were subsequently analyzed on SDS-PAGE.

### ***TR SAXS data collection and processing***

All SAXS data collection was performed at SSRL beam line 4-2 as previously described<sup>39</sup>. Stopped-flow time-resolved experiments were conducted using the experimental procedures previously published<sup>47</sup> with the exceptions of different exposure times and slightly different  $q$  range covered. The magnitude of momentum transfer  $q$  is defined as  $4\pi\sin(\theta)/\lambda$ , where  $\theta$  is one half the scattering angle and  $\lambda$  is the x-ray wavelength.

A limited region of  $q$ -values containing most information (0.008-0.15  $\text{\AA}^{-1}$  for the manual mixing runs and 0.011-0.15  $\text{\AA}^{-1}$  for stopped flow experiments) was selected for singular value decomposition (SVD) analysis of time-resolved scattering. SVD was performed in Matlab by a standard protocol<sup>27</sup>. The number of significant scattering factors was determined by ranking the singular values<sup>48</sup> and by comparing the residuals between the reconstructed and experimental data with the experimental noise level<sup>26</sup> (see Supplementary Methods, Fig. S2 and S3).

The raw eigenvectors (scattering curves, scattering factors) and their cognate loading factors (time-dependent concentrations) as produced by SVD are orthogonal and thus generally do not have a direct physical meaning, e.g. often giving negative values (See Supplementary Methods for further discussion). Thus, linear combinations (rotations) of the

SVD eigenvectors were generated to obtain physically meaningful scattering curves and their corresponding time-dependent concentrations<sup>28; 49; 50</sup>. First, the scattering curves of the initial (I) and the final (F) state were approximated by linear combinations of the SVD eigenvectors in such a way that their loading factors were non-negative. Then an appropriate linear combination approximately orthogonal to both I and F was found to yield a non-negative scattering curve and the corresponding non-negative loading factors (Fig 3). This procedure yielded similar TR-SAXS result at two different CP concentrations (c.f. Fig 3, A and B). After scaling of intensities and concentrations we obtained a unique solution to the problem (see Supplementary Methods for additional details).

### ***Modeling of incomplete shells***

Incomplete hollow shells with inner and outer radius 160Å and 310Å, respectively were modeled as spherical sections inscribed by a cone with angle  $\theta$ . The distance distributions  $H(R, \theta)$  for these partial shells were obtained computationally by direct enumeration for each value of  $\theta$  (using steps of 0.9°). Scattering curves were obtained from the distance distributions by Fourier transform<sup>29</sup>.

### ***Acknowledgments***

R.T. was in part supported by a research fellowship from Academy of Finland (1118462). PEP was supported by grants from the National Institutes of Health- GM47980 and National Science Foundation- DBI-9726698. X-ray scattering studies were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program Grant Number 5 P41 RR001209. The contents of this work are

solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH

## References

1. Hendrix, R. W. (2005). Bacteriophage HK97: assembly of the capsid and evolutionary connections. *Adv Virus Res* 64, 1-14.
2. Reddy, V. S. & Johnson, J. E. (2005). Structure-derived insights into virus assembly. *Adv Virus Res* 64, 45-68.
3. Wang, Q., Lin, T., Johnson, J. E. & Finn, M. G. (2002). Natural supramolecular building blocks. Cysteine-added mutants of cowpea mosaic virus. *Chem Biol* 9, 813-9.
4. Stray, S. J., Bourne, C. R., Punna, S., Lewis, W. G., Finn, M. G. & Zlotnick, A. (2005). A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. *Proc Natl Acad Sci U S A* 102, 8138-43.
5. Sakalian, M., McMurtrey, C. P., Deeg, F. J., Maloy, C. W., Li, F., Wild, C. T. & Salzwedel, K. (2006). 3-O-(3',3'-dimethylsuccinyl) betulinic acid inhibits maturation of the human immunodeficiency virus type 1 Gag precursor assembled in vitro. *J Virol* 80, 5716-22.
6. Wang, Q., Lin, T., Tang, L., Johnson, J. E. & Finn, M. G. (2002). Icosahedral virus particles as addressable nanoscale building blocks. *Angew Chem Int Ed Engl* 41, 459-62.
7. Falkner, J. C., Turner, M. E., Bosworth, J. K., Trentler, T. J., Johnson, J. E., Lin, T. & Colvin, V. L. (2005). Virus crystals as nanocomposite scaffolds. *J Am Chem Soc* 127, 5274-5.
8. Abendon, S. T. & Calendar, R. L. (2005). *The bacteriophages*. 2nd edit, Oxford University Press, New York.
9. Zlotnick, A., Aldrich, R., Johnson, J. M., Ceres, P. & Young, M. J. (2000). Mechanism of capsid assembly for an icosahedral plant virus. *Virology* 277, 450-6.
10. Zlotnick, A., Johnson, J. M., Wingfield, P. W., Stahl, S. J. & Endres, D. (1999). A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* 38, 14644-52.
11. Casjens, S. & King, J. (1974). P22 morphogenesis. I: Catalytic scaffolding protein in capsid assembly. *J Supramol Struct* 2, 202-24.
12. Earnshaw, W., Casjens, S. & Harrison, S. C. (1976). Assembly of the head of bacteriophage P22: x-ray diffraction from heads, proheads and related structures. *J Mol Biol* 104, 387-410.
13. King, J., Botstein, D., Casjens, S., Earnshaw, W., Harrison, S. & Lenk, E. (1976). Structure and assembly of the capsid of bacteriophage P22. *Philos Trans R Soc Lond B Biol Sci* 276, 37-49.
14. Prevelige, P. E., Jr. & King, J. (1993). Assembly of bacteriophage P22: a model for ds-DNA virus assembly. *Prog Med Virol* 40, 206-21.
15. Prasad, B. V., Prevelige, P. E., Marietta, E., Chen, R. O., Thomas, D., King, J. & Chiu, W. (1993). Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. *J Mol Biol* 231, 65-74.
16. Jiang, W., Li, Z., Zhang, Z., Baker, M. L., Prevelige, P. E., Jr. & Chiu, W. (2003). Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat Struct Biol* 10, 131-5.

17. Prevelige, P. E., Jr., Thomas, D. & King, J. (1988). Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells in vitro. *J Mol Biol* 202, 743-57.
18. Prevelige, P. E., Jr., Thomas, D. & King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys J* 64, 824-35.
19. Parker, M. H., Casjens, S. & Prevelige, P. E., Jr. (1998). Functional domains of bacteriophage P22 scaffolding protein. *J Mol Biol* 281, 69-79.
20. Parker, M. H., Stafford, W. F., 3rd & Prevelige, P. E., Jr. (1997). Bacteriophage P22 scaffolding protein forms oligomers in solution. *J Mol Biol* 268, 655-65.
21. Thuman-Commike, P. A., Greene, B., Malinski, J. A., King, J. & Chiu, W. (1998). Role of the scaffolding protein in P22 procapsid size determination suggested by T = 4 and T = 7 procapsid structures. *Biophys J* 74, 559-68.
22. Thuman-Commike, P. A., Greene, B., Jakana, J., Prasad, B. V., King, J., Prevelige, P. E., Jr. & Chiu, W. (1996). Three-dimensional structure of scaffolding-containing phage p22 procapsids by electron cryo-microscopy. *J Mol Biol* 260, 85-98.
23. Thuman-Commike, P. A., Greene, B., Malinski, J. A., Burbea, M., McGough, A., Chiu, W. & Prevelige, P. E., Jr. (1999). Mechanism of scaffolding-directed virus assembly suggested by comparison of scaffolding-containing and scaffolding-lacking P22 procapsids. *Biophys J* 76, 3267-77.
24. Parker, M. H. & Prevelige, P. E., Jr. (1998). Electrostatic interactions drive scaffolding/coat protein binding and procapsid maturation in bacteriophage P22. *Virology* 250, 337-49.
25. Galisteo, M. L. & King, J. (1993). Conformational transformations in the protein lattice of phage P22 procapsids. *Biophys J* 65, 227-35.
26. Fowler, A. G., Foote, A. M., Moody, M. F., Vachette, P., Provencher, S. W., Gabriel, A., Bordas, J. & Koch, M. H. (1983). Stopped-flow solution scattering using synchrotron radiation: apparatus, data collection and data analysis. *J Biochem Biophys Methods* 7, 317-29.
27. Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations*. 3rd edit, The Johns Hopkins University Press, Baltimore.
28. Lanman, J., Tuma, R. & Prevelige, P. E., Jr. (1999). Identification and characterization of the domain structure of bacteriophage P22 coat protein. *Biochemistry* 38, 14614-23.
29. Svergun, D. I., Semenyuk, A. V. & Feigin, L. A. (1988). Small-angle scattering data treatment by the regularization method. *Acta Cryst.* A44, 244-250.
30. Guinier, A. (1939). Diffraction of X Rays of Very Small Angles-Application to the Study of Ultramicroscopic Phenomena. *Ann. Phys.* 12, 161.
31. Endres, D. & Zlotnick, A. (2002). Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys J* 83, 1217-30.
32. Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, Plenum Press, New York.
33. Kihara, H. (1994). Stopped-Flow Apparatus for X-ray Scattering and XAFS. *J. Synchrotron Rad.* 1, 74-77.
34. Sun, Y., Parker, M. H., Weigele, P., Casjens, S., Prevelige, P. E., Jr. & Krishna, N. R. (2000). Structure of the coat protein-binding domain of the scaffolding protein from a double-stranded DNA virus. *J Mol Biol* 297, 1195-202.
35. Parker, M. H., Jablonsky, M., Casjens, S., Sampson, L., Krishna, N. R. & Prevelige, P. E., Jr. (1997). Cloning, purification, and preliminary characterization by circular dichroism and NMR of a carboxyl-terminal domain of the bacteriophage P22 scaffolding protein. *Protein Sci* 6, 1583-6.

36. Parent, K. N., Zlotnick, A. & Teschke, C. M. (2006). Quantitative analysis of multi-component spherical virus assembly: scaffolding protein contributes to the global stability of phage P22 procapsids. *J Mol Biol* 359, 1097-106.
37. Parent, K. N., Suhanovsky, M. M. & Teschke, C. M. (2007). Phage P22 procapsids equilibrate with free coat protein subunits. *J Mol Biol* 365, 513-22.
38. Parent, K. N., Doyle, S. M., Anderson, E. & Teschke, C. M. (2005). Electrostatic interactions govern both nucleation and elongation during phage P22 procapsid assembly. *Virology* 340, 33-45.
39. Tuma, R., Tsuruta, H., Benevides, J. M., Prevelige, P. E., Jr. & Thomas, G. J., Jr. (2001). Characterization of subunit structural changes accompanying assembly of the bacteriophage P22 procapsid. *Biochemistry* 40, 665-74.
40. Kang, S. & Prevelige, P. E., Jr. (2005). Domain study of bacteriophage p22 coat protein and characterization of the capsid lattice transformation by hydrogen/deuterium exchange. *J Mol Biol* 347, 935-48.
41. Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W. & Ashcroft, A. E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J Mol Biol* 369, 541-52.
42. Parker, M. H., Brouillette, C. G. & Prevelige, P. E., Jr. (2001). Kinetic and calorimetric evidence for two distinct scaffolding protein binding populations within the bacteriophage P22 procapsid. *Biochemistry* 40, 8962-70.
43. Zlotnick, A. (2003). Are weak protein-protein interactions the general rule in capsid assembly? *Virology* 315, 269-74.
44. Wyckoff, E. & Casjens, S. (1985). Autoregulation of the bacteriophage P22 scaffolding protein gene. *J Virol* 53, 192-7.
45. Zhang, T. & Schwartz, R. (2006). Simulation study of the contribution of oligomer/oligomer binding to capsid assembly kinetics. *Biophys J* 90, 57-64.
46. Ceres, P. & Zlotnick, A. (2002). Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* 41, 11525-31.
47. Canady, M. A., Tsuruta, H. & Johnson, J. E. (2001). Analysis of Rapid, Large-Scale Protein Quaternary Structural Changes: Time-Resolved X-ray Solution Scattering of Nudaurelia capensis o Virus (NoV) Maturation,. *J. Mol. Biol.* 311, 803-814.
48. Quinn, G. P. & Keough, M. J. (2002). *Experimental Design and Data Analysis for Biologists.*, Cambridge University Press, Cambridge.
49. Paatero, P. & Tapper, U. (1993). Analysis of different modes of factor analysis as least squares fit problems. *Chemometrics and Intelligent Laboratory Systems* 18, 183-194.
50. Paatero, P. & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5, 111-126.

## Figure legends

**Figure 1.** (A) Scaffolding protein assisted *in vitro* assembly of bacteriophage P22 procapsid. (B) Different scaffolding protein constructs as used in this study.

**Figure 2.** Manual mixing (dead time about 1 min) and TR SAXS using wt scaffold (14.7 $\mu$ M) and coat(24.9 $\mu$ M). (A) TR SAXS. (B) Final product scattering (after 25 min, solid red line) fitted to a hollow shell model (dotted blue):  $R_{\text{inner}} = 160 \text{ \AA}$   $R_{\text{outer}} = 311 \text{ \AA}$ .

**Figure 3.** Analysis of TR SAXS from assembly reactions by SVD. Top panels are the factor loadings which represent relative distribution of protein mass between the species corresponding to the scattering factors in the bottom panels. (A) 14 $\mu$ M CP + 14.7 $\mu$ M SP (B) 24.9 $\mu$ M CP + 14.7 $\mu$ M SP. Black traces (solid line) correspond to the relative abundance or scattering factor of free coat and scaffolding protein (reactants). Red traces (dashed) correspond to those of the intermediates. Green traces (dotted) indicate the product concentrations and scattering factors.

**Figure 4.** Dependence of SAXS on the completeness of the shell. (A) Schematic illustration of the partial shell model (conical wedge, left and a two-dimensional projection on the right) with rotational symmetry used for simulating the scattering curves in panel B. The inner and outer radii obtained for the final state from the experimental data in Fig. 2B were used in the computations. The spherical  $\Theta$  was varied from 0 to 180 degree i.e. from 0 to 100% completeness (B) Representative simulated scattering curves for various degrees of completion. (C) Dependence of the radius of gyration ( $R_g$ ) on completeness. The solid black trace corresponds to the exact value computed directly from the model in panel A. The red dotted curve corresponds to the apparent  $R_g$  estimated from the simulated scattering curves in panel B using the  $q$  range 0.008-0.012  $\text{\AA}^{-1}$ . The blue dash-dot line shows the apparent  $R_g$  estimated from the simulated scattering curves using the  $q$  range 0.011-0.016  $\text{\AA}^{-1}$ .

**Figure 5.** P22 assembly kinetics monitored by stopped-flow TR-SAXS. (A) TR SAXS curves rendered to 10 sec resolution, 24.9 $\mu$ M CP + 14.7 $\mu$ M SP. (B) The scattering factors (top) and relative mass concentrations (bottom) computed from the data in A by SVD and factor rotation as described in Methods.

**Figure 6.** Assembly of CP with R74C/L177I disulfide dimer SP (A) TR-SAXS at 24.5  $\mu$ M CP 7.4  $\mu$ M SP dimer (monomer concentration 14.7  $\mu$ M) (B-C) Scattering factors (bottom) and relative concentrations (top) for two initial conditions: (B) 12.8  $\mu$ M CP, 7.4  $\mu$ M SP dimer (C) 24.5  $\mu$ M CP, 7.4  $\mu$ M SP dimer. The dashed lines indicate that only the end point at 1500 s was collected after the finish of the time-resolved series at 900 s. (D) Assembly with reduced R74C: TR-SAXS on the left, scattering factors on the right, 16.6  $\mu$ M CP, 18.5  $\mu$ M SP (monomer concentration).

**Figure 7.** Scaffolding monomer is required for assembly of complete shells. R74C dimer promotes nucleation. Left panel shows SDS-PAGE analysis of fractions after sedimentation of assembly products on a 5-20% sucrose gradient. Right panel shows negatively stained assemblies from the peak fraction of the scaffold dimer reaction. Protein concentrations ( $\mu$ M) are shown in columns on the left.

**Figure 8.** Assembly of CP with SP CTF. (A) Stopped-flow TR SAXS (B) Assembly products after 60 min were separated by a sucrose gradient and fractions were analyzed by 13% SDS-PAGE.

**Figure 9.** Role of different scaffolding protein oligomers in assembly of bacteriophage P22 procapsid.

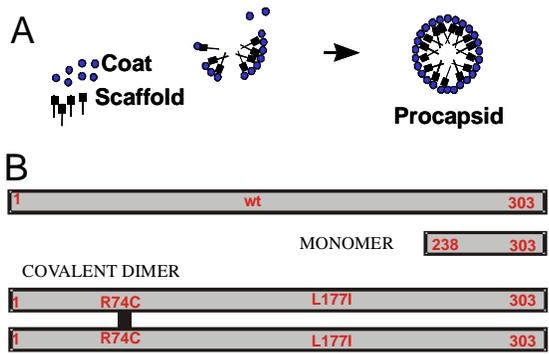


Figure 1

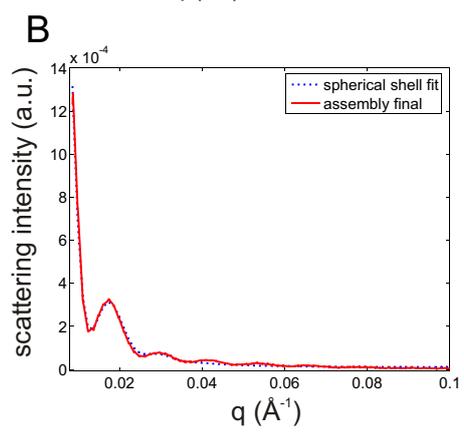
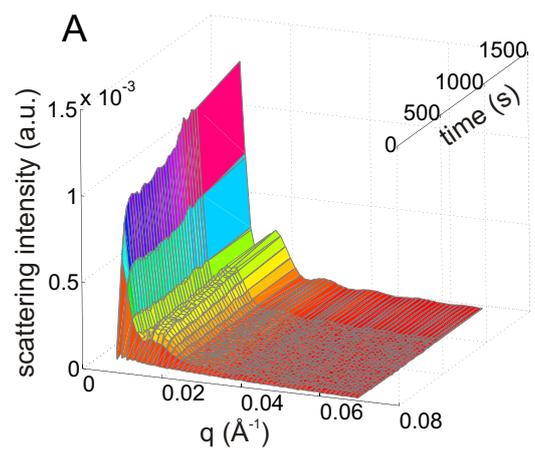


Figure 2

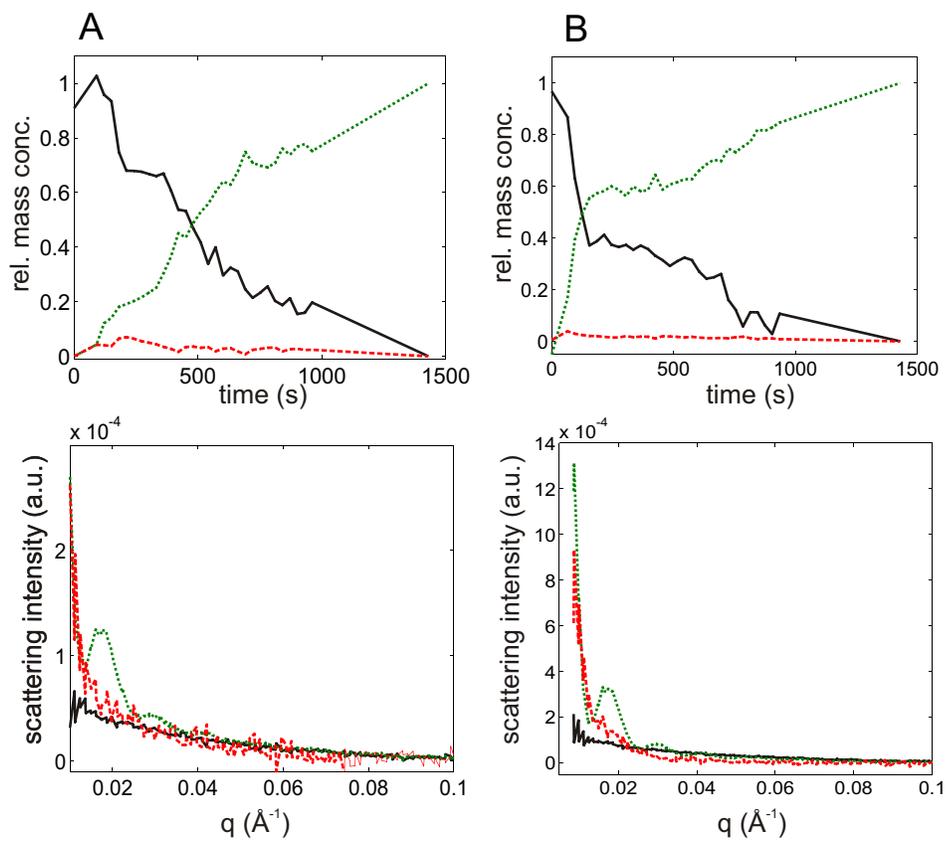


Figure 3

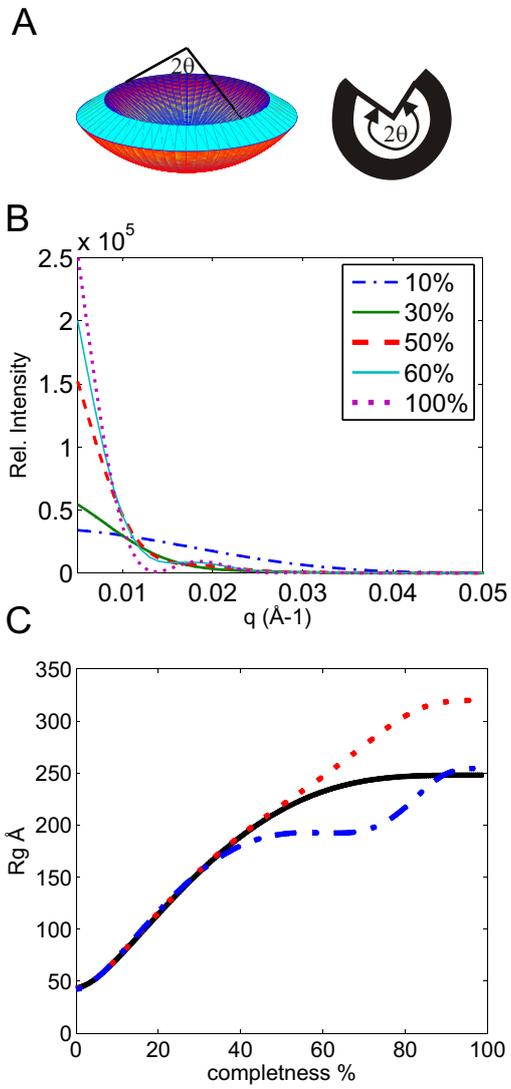


Figure 4

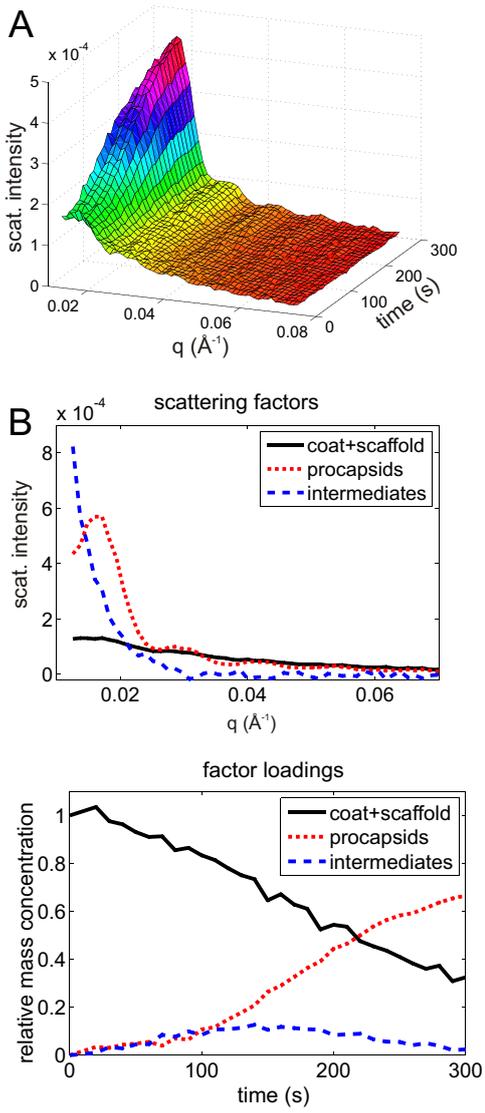


Figure 5

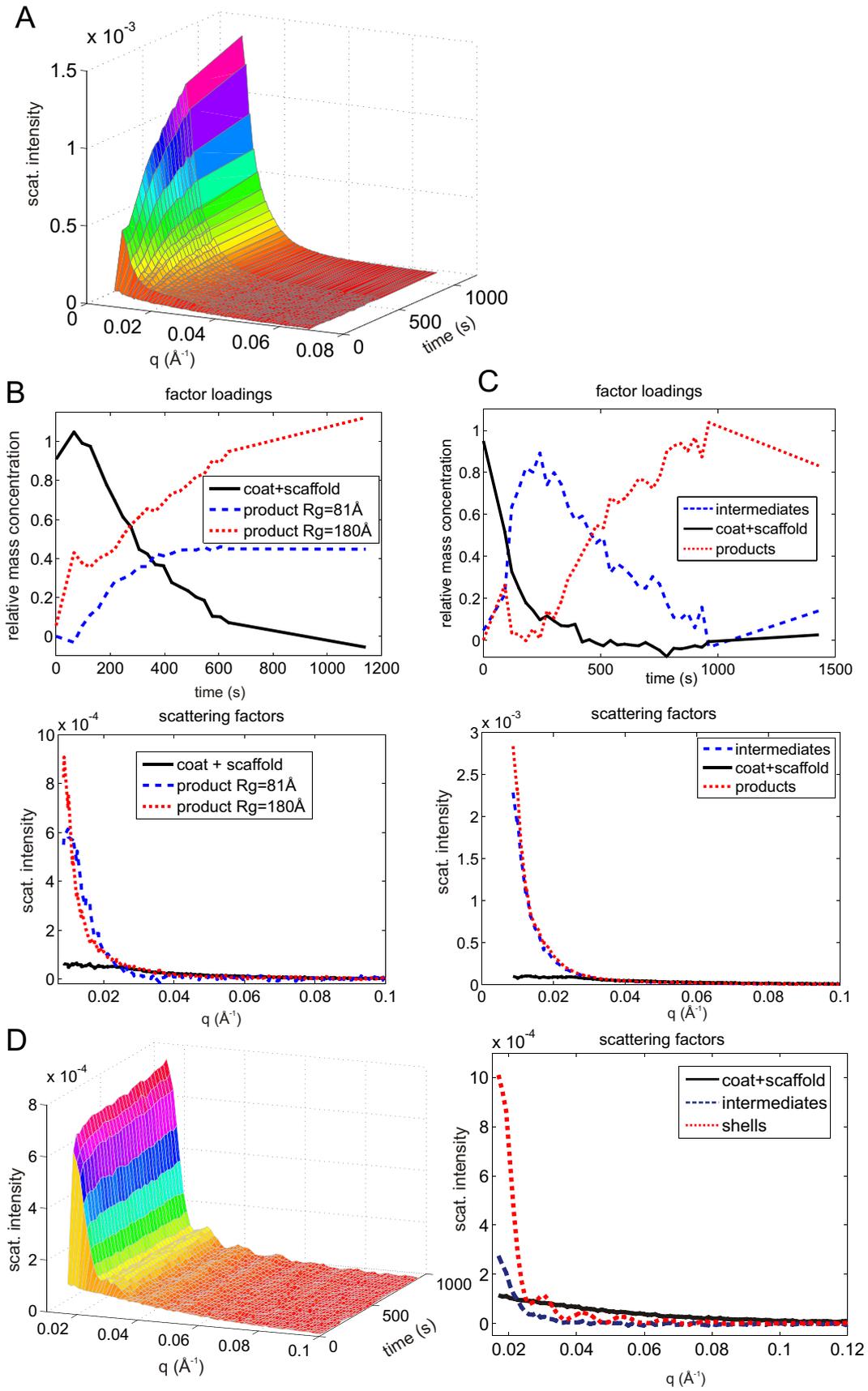


Figure 6

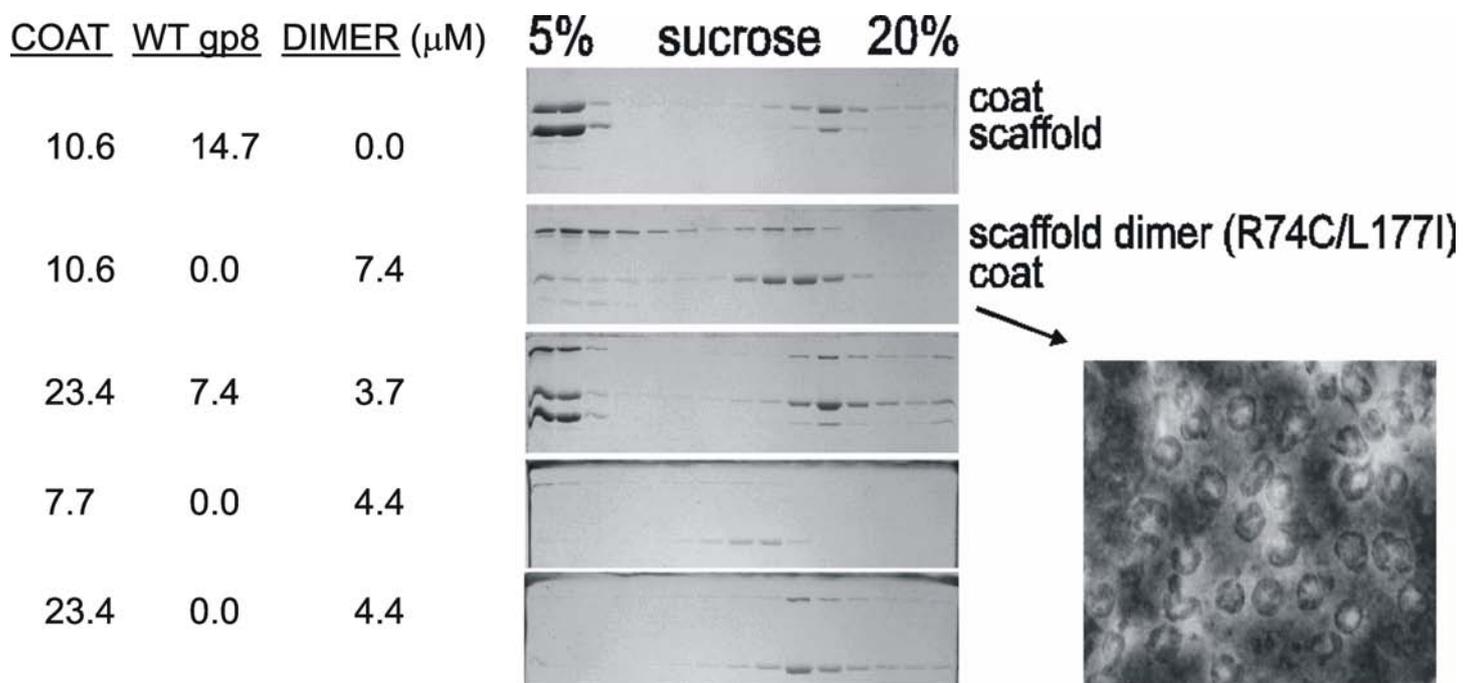


Figure 7

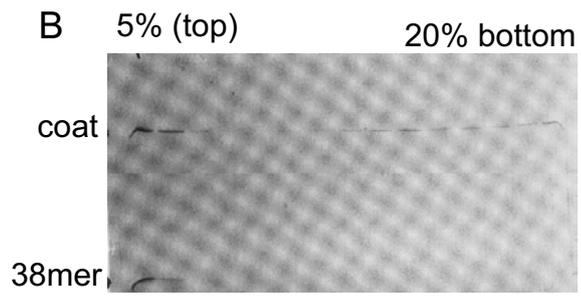
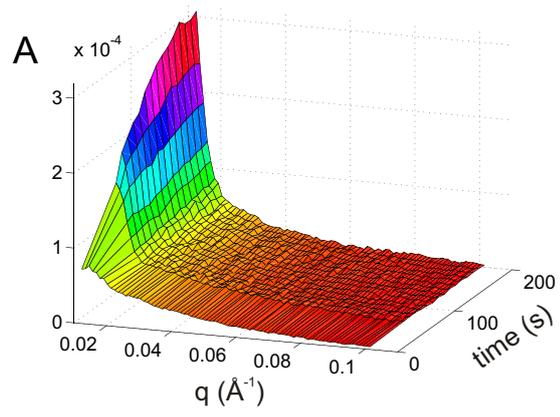


Figure 8

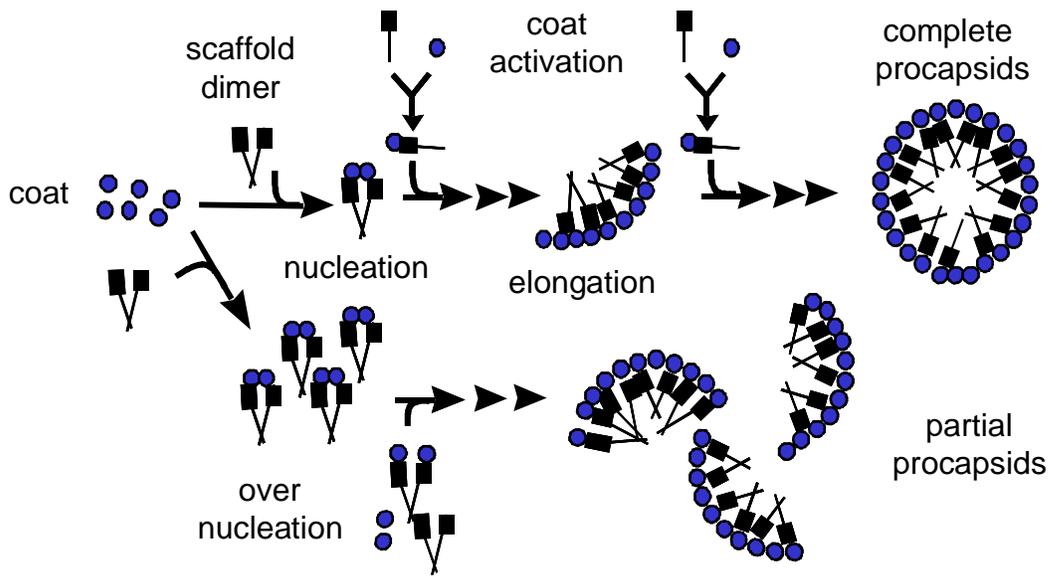


Figure 9

## Supplementary methods

Here we illustrate the SVD data processing using the data shown in Fig. 2 as an example. Unless stated otherwise, other data sets were processed in similar fashion.

### Detection of intermediates-two states versus three or more states

In the absence of detectable intermediates the composition of a dynamic system at any time can be described as a superposition of the starting and the resulting state. In our particular case this means that each of the scattering curves within the TR-SAXS envelope shall be a linear combination of the starting ( $t=0$ ) scattering curve (coat + scaffold) and the end point (ideally  $t=\infty$  but in practice the stationary scattering curve at  $t=1500$  s for data in Fig. 2A). In mathematical terms:

$$I(t,q)=c^0(t)\cdot I(0,q)+c^1(t)\cdot I(1500,q) \quad \text{SEq. 1}$$

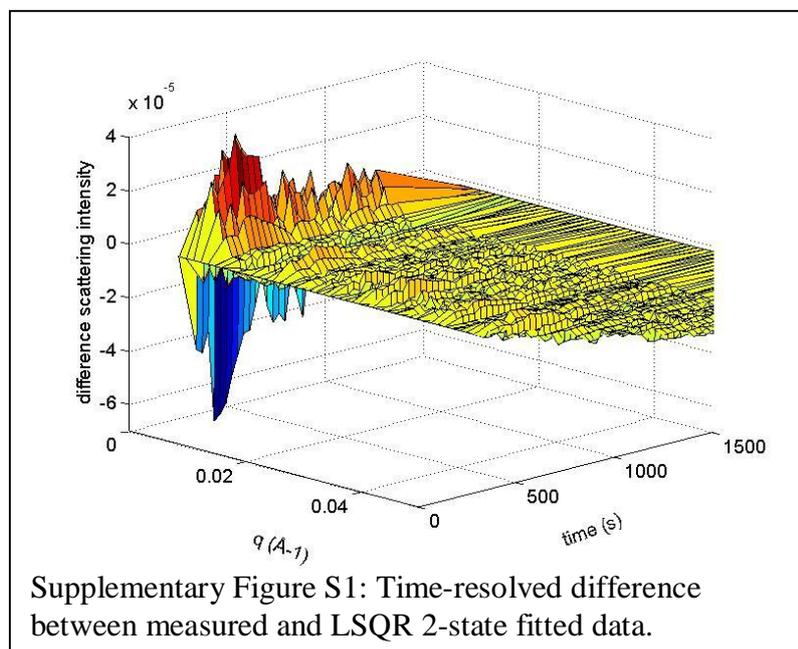
Where  $I(t,q)$  is the scattering intensity at time  $t$  and scattering vector and  $c^0(t)$  resp.  $c^1(t)$  are the time dependent concentrations of initial mixture and product, respectively. In matrix notation we have:

$$I=C\cdot I^{0\infty} \quad \text{SEq. 2}$$

Where matrix  $I^{0\infty}$  contains the two row vectors corresponding to  $I(0)$  and  $I(\infty)$ ,  $C$  is the unknown matrix containing the respective concentrations and  $I$  contains the measured TR-SAXS data. The unknown concentrations can be obtained by standard linear least-squares (LSQR) procedure ( $T$  indicates matrix transposition)<sup>1</sup>:

$$C=I\cdot I^{0\infty T}\cdot(I^{0\infty}\cdot I^{0\infty T})^{-1} \quad \text{SEq. 3}$$

And the reconstructed data  $I'$  could be obtained by substituting the LSQR concentration matrix into SEq. 2. Fig. S1 shows the difference  $I-I'$  which exhibits a distinct set of time-dependent minima and maxima at low scattering angles, i.e. in the region sensitive to the appearance of larger species. In comparison with noise (estimated from higher  $q$  values) these maxima and minima are significant, they also seem to appear and peak at around 200 s and then gradually subside. This is an indication of significant amount of additional species, i.e. assembly intermediates.



then gradually subside. This is an indication of significant amount of additional species, i.e. assembly intermediates. The total number of detectable species is equal to the effective rank of the data matrix  $I$  which can be determined by SVD<sup>1</sup>.

### Singular value decomposition

The SVD theorem states that for a real  $m$  by  $n$  matrix (such as  $I$ ) there exist orthogonal matrices  $U$  ( $m\times m$ ) and  $V$  ( $n\times n$ ) such as:

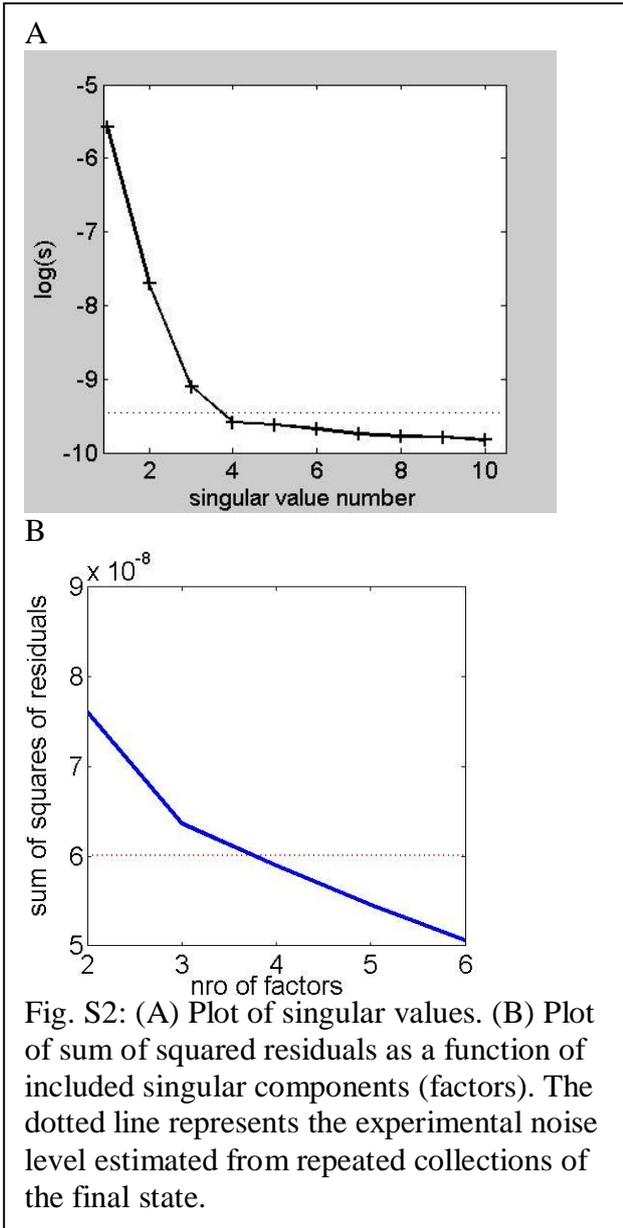


Fig. S2: (A) Plot of singular values. (B) Plot of sum of squared residuals as a function of included singular components (factors). The dotted line represents the experimental noise level estimated from repeated collections of the final state.

values and setting the rest to zero suffices to approximate  $I$  with a matrix  $I_k$  of rank  $k$ . The effective rank is equal to the number of linearly independent components that are sufficient to approximate the data set within the experimental error, i.e. noise level. Here, we determined the effective rank by first plotting out the  $s$  values and using only the ones in the region above the noise plateau (above the “elbow”), i.e.  $k=3$  for the data set in Fig.2 (Fig. S2A)<sup>2</sup>. Then we checked that this number of components described the data set to within experimental noise level (Fig S2B) and that no systematic deviations were apparent in residuals Fig. S3A). Including 4 components did not improve the fit significantly (Fig. S3B). Given that 2 components are insufficient to describe the data (see above and Fig. S1) the data set was described by the 3 most significant components (scattering factors and cognate concentrations).

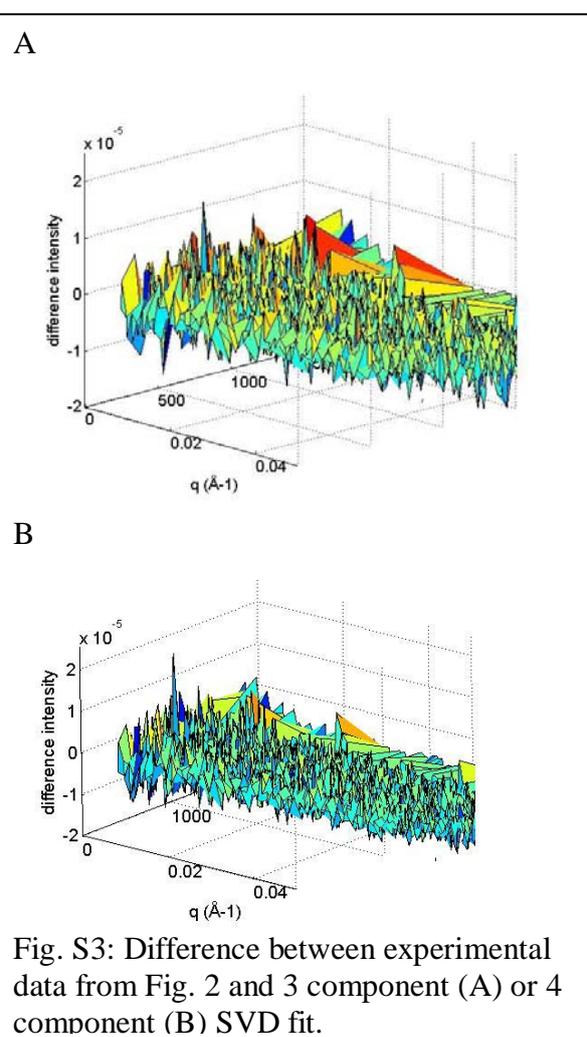


Fig. S3: Difference between experimental data from Fig. 2 and 3 component (A) or 4 component (B) SVD fit.

$$U^T \cdot I \cdot V = S \quad \text{SEq. 4}$$

Where  $S$  is a diagonal matrix with decreasing diagonal values  $s_1 \geq s_2 \geq \dots \geq s_p \geq 0$  ( $p = \min(m, n)$ ). Selecting only a limited number ( $k$ ) of  $s$

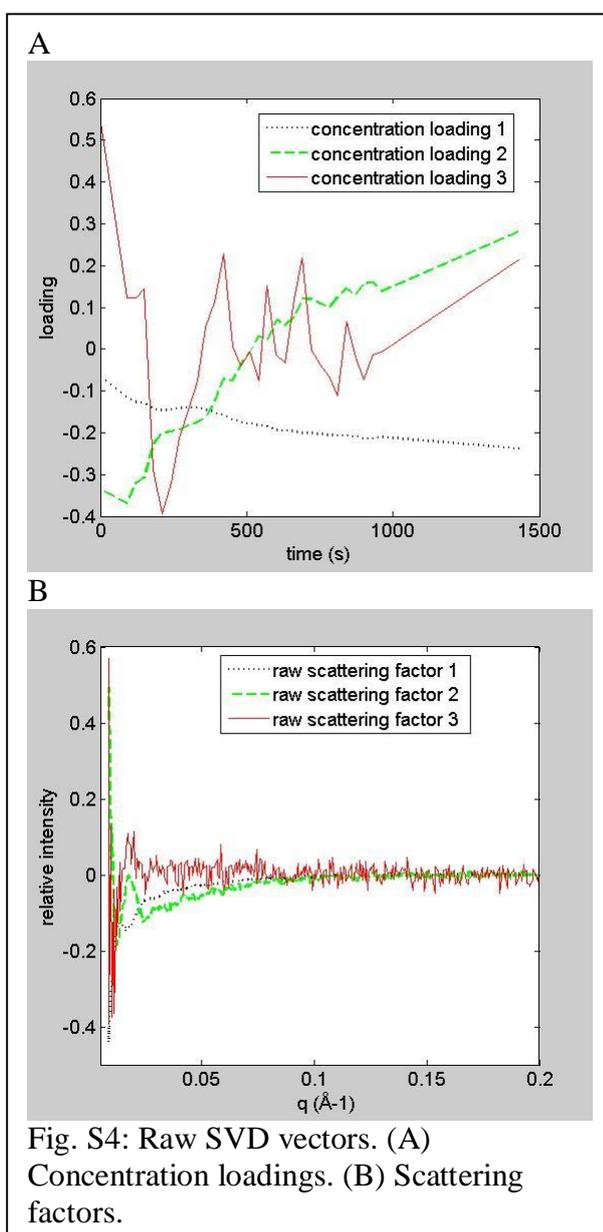
## From SVD vectors to physically meaningful scattering factors and concentrations

The vectors of  $U$  and  $V$  form orthogonal basis for linear approximation of the experimental data. In our case of  $I$  (32 time points  $\times$  310  $q$  values) the matrix  $U$  (32 $\times$ 32) is related to concentrations (factor loadings) while  $V$  (310 $\times$ 310) is related to the scattering factors. The first three (significant) vectors of  $U$  and  $V$  matrix are shown in Fig. S4. Given that SVD is a mathematical procedure it is not surprising that the raw scattering factors and concentration loadings do not directly represent the scattering curves of the initial, final and intermediates states. As such, the raw vectors are clearly unphysical in terms of being negative, and they become meaningful only in a suitable linear combination (note that linear transformation with non-singular matrix do not add or remove any information).

However, the orthogonal, SVD-derived vectors provide a convenient starting point (expansion basis) for reconstructing the structure factors and cognate concentrations, including the unknown intermediate. This is done by transforming (rotating) the significant raw SVD vectors (in matrices  $U'$  and  $V'$ ) into a new set:

$$Y = A \cdot S \cdot V'^T$$

SEq. 5



Where  $A$  is the yet unknown (3 $\times$ 3) regular rotation matrix and  $S'$  is the truncated (3 $\times$ 3) diagonal matrix containing the significant singular values. The cognate concentrations are determined accordingly:

$$C = U' \cdot A^T$$

SEq. 6

In case 2 out of the 3 physically meaningful scattering factors are known (i.e. the initial and final scattering curve) the matrix  $A$  was determined by a least squares procedure as described previously<sup>3,4</sup> and determined up to a scaling factor, i.e. the concentrations may be multiplied by a non-zero number while the scattering curve is divided by the same number. These were determined from scaling of scattering factors as described below. In the cases that the final state is polydisperse i.e. two or more product components were present (e.g. data in Fig. 6) the rotation could not be sufficiently constrained by the initial and final state alone. Hence, the concentration loadings of the (known) initial and final states were determined first. The rest of the scattering factors and concentrations were obtained by iterative manual rotation while applying additional constraints (non-negativity of both scattering factors and concentrations within experimental error)<sup>5</sup>. However, this procedure did not yield unique scattering factors and concentrations and many linear

combinations were still possible. Additionally, scaling of the scattering factors and relative mass concentrations was employed to further constrain the linear combinations.

Scaling of scattering factors and concentrations.

The scattering intensity extrapolated to the zero angle  $I(0)$  is proportional to the product of mass squared and number concentration of the scattering species<sup>6</sup>. The relative masses of partial shell species were estimated from the apparent  $R_g$  using degree of completion from Fig. 4C. The scattering factors were then multiplied to achieve the right ratio between  $I(0)$  values of different species. At the same time, the magnitude of the scaling factors was selected such as the relative mass concentrations (factor loadings) were adding up to unity at all time points (i.e. conservation of mass). In cases of product heterogeneity (e.g. data in Fig. 6), the latter condition was not automatically fulfilled i.e. scattering factors were still mixed up. In such a case, another round of iterative factor rotation was applied to obtain scaled relative concentrations.

### Supplementary References

1. Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations*. 3rd edit, The Johns Hopkins University Press, Baltimore.
2. Quinn, G. P. & Keough, M. J. (2002). *Experimental Design and Data Analysis for Biologists.*, Cambridge University Press, Cambridge.
3. Paatero, P. & Tapper, U. (1993). Analysis of different modes of factor analysis as least squares fit problems. *Chemometrics and Intelligent Laboratory Systems* 18, 183-194.
4. Lanman, J., Tuma, R. & Prevelige, P. E., Jr. (1999). Identification and characterization of the domain structure of bacteriophage P22 coat protein. *Biochemistry* 38, 14614-23.
5. Paatero, P. & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5, 111-126.
6. Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, Plenum Press, New York.