Supplementary Material

POLYMERIZATION AND NUCLEIC ACID BINDING PROPTERTIES OF HUMAN L1 ORF1 PROTEIN

Callahan, Kathryn E.[1,3], Alison B. Hickman[2], Charles E. Jones[1], Rodolfo Ghirlando[2], Anthony V. Furano[1]

From the Laboratories of Molecular and Cellular Biology[1]and Molecular Biology[2], NIDDK, National Institutes of Health, Bethesda, Maryland 20892 and the Department of Biochemistry and Molecular & Cell Biology, Georgetown University Medical Center[3], Washington, DC 20057

Experimental Procedures

Site-directed mutagenesis to generate ORF1p M121A/F and M125A/F mutants – We used the QuikChange Site-directed mutagenesis kit (Stratagene) to change both M121 and M125 to either phenylalanine or alanine. A list of the oligonucleotide primers for mutagenesis is available from the authors. Mutations were confirmed by DNA sequencing.

N-terminal sequencing – The origin of the ~24 kDa protein that co-purified with GST-ORF1p expressed in *E. coli* was identified by N-terminal sequencing. After size exclusion chromatography, ORF1p fractions were analyzed on a 4-12% NuPage gel (Invitrogen). The two major proteins were transferred to a PVDF membrane with a Novex blotting apparatus. The 24 kDa band was excised from the membrane, and analyzed on an Applied Biosystems Procise protein sequencer.

Analytical ultracentrifugation – Sedimentation velocity or sedimentation equilibrium of ORF1p or M128p (in Tris storage buffer – see Experimental procedures) was carried out in a Beckman Optima XL-A or a Beckman Coulter Proteome XL-I analytical ultracentrifuge. 2-channel, 12 mm path length cells were used for sedimentation velocity, whereas 6-channel, 12 mm path length cells were used for sedimentation equilibrium.

ORF1p Sedimentation velocity – This was carried out at 6.5 and 13.0 μM, 20.0 ºC and 55k rpm. One hundred and sixty absorbance (280 nm) scans were collected at 6.2 minute intervals using a radial spacing of 0.003 cm. Data were analyzed in SEDFIT 11.71 (1,2) in terms of a continuous c(*s*) distribution covering an $s_{20,w}$ range of 0.1 – 23 S with a resolution of 100 and a confidence level (F-ratio) of 0.68. Good fits were obtained with root mean square deviations of 0.0047–0.0077 absorbance units. The solution density (ρ) measured at 20.00 ºC (Mettler Toledo DE51 density meter) corresponded to that calculated in SEDNTERP 1.09 (3). Therefore we used the viscosity (η) calculated based on the solvent composition in SEDNTERP 1.09 (3). The partial specific volume (*v*) for ORF1p was also calculated in SEDNTERP 1.09 and sedimentation coefficients *s* were corrected to $s_{20,w}$.

ORF1p Sedimentation equilibrium – Protein at 18.5 and 37.9 μM was centrifuged at 11, 14 and 17 krpm at 4ºC. The average of four absorbance measurements at 250 and 280 nm were collected

with a radial spacing of 0.001 cm. Equilibrium was achieved within 48 hours. Data were analyzed globally using SEDPHAT 6.21 (4,5) in terms of a single ideal solute with excellent data fits.

M128p Sedimentation Velocity – This was carried out at 8.8 and 18.3 μM, 20.0 ºC and 50k rpm. One hundred and forty absorbance (280 nm) scans using a radial spacing of 0.003 cm (1,2) and analyzed in terms of a continuous c(*s*) distribution. As in the case of ORF1p, the solution density (ρ) measured at 20.00 ºC corresponded to that calculated in SEDNTERP 1.09 (3). The sedimentation velocity data collected at both loading concentrations were also analyzed globally in terms of a Lamm equation describing a monomer-trimer self-association in SEDPHAT 7.03 (4,5) with the implementation of mass conservation. Excellent fits were obtained with root mean square deviations ranging from 0.0033 – 0.0035 absorbance units.

M128p Sedimentation equilibrium – M128p at loading concentrations of 6.1, 12.2 and 18.8 μM was centrifuged at 12, 16, 20 and 24 krpm at 4ºC. The average of four absorbance measurements at 280 nm and a radial spacing of 0.001 cm were collected for each speed. Sedimentation equilibrium was achieved within 48 hours. Data collected at different speeds and loading concentrations were analyzed globally in terms of a monomer-trimer self-association in SEDPHAT 7.03 (4,5) with the implementation of mass conservation. The partial specific volume *v* was calculated based on the amino acid composition in SEDNTERP 1.09 (3), together with the extinction coefficient at 280 nm. Errors in the equilibrium constants were determined using the method of F-statistics with a confidence level of 68.3%.

Results

Human ORF1p products synthesized in *E. coli* - Full length human ORF1p expressed in *E. coli* co-purified with a 24 kDa protein. N-terminal sequencing of this protein revealed two signals: a strong signal corresponding to MEDEMNEM and a weak one, XNXMK (where X represents an unidentifiable amino acid) (Figure S1A). These results indicated that the 24 kDa band consists of two proteins fragments derived from ORF1p, one beginning with Met121 and the second with Met125. As non-canonical Shine-Dalgarno sequences (Figure S1B) are 5' of these methionines, it seemed likely that internal initiation at these sites accounted for the 24 kDa ORF1p fragments. The fact that mutation of Met121 and Met125 to phenylalanine or alanine reduced the amount of internally initiated ORF1p (Figure S1C) supports this conclusion. However, even for these double mutants some 24 kDa fragment persists, the product of a third internal initiation product at Met128.

Human ORF1p is a trimer. Others recently showed that human ORF1p expressed in *E. coli* is a trimer (6). Here we determined that human ORF1p expressed in eukaryotic cells is also trimeric. Sedimentation velocity experiments indicated the presence of a major species having an average $s_{20,w}$ of 4.60 ± 0.05 S, which accounted for at least 90% of the soluble material loaded (Figure S2A). Sedimentation equilibrium experiments confirm that the ORF1p is a mono-disperse species and an analysis in terms of a single ideal solute returns a molecular mass of 122.4 ± 1.3 kDa (n = 3.05 ± 0.03), which corresponds to an ORF1p trimer, with excellent data fits (Figure S2B & C). Assuming a hydration contribution of 0.3 g water per g of protein, the measured sedimentation coefficient corresponds to a frictional ratio, *f/fo*, of 1.76, indicating that the protein

is asymmetric in shape. This finding is consistent with results previously reported for the mouse protein that has a frictional ratio of 1.69 (7) and with atomic force microscopy data that showed the mouse protein resembles an asymmetric dumbbell (8).

The carboxy terminal half of ORF1p (M128p) forms a temperature sensitive trimer - M128p retains 2.5 heptads of the predicted (9) 13 heptad coiled coil domain (Figures 1 & S1), too few to form a coiled coil per the prediction program. Nonetheless, M128p elutes as a multimer during size exclusion chromatography (see Experimental Procedures) at 4 ºC (data not shown). Sedimentation velocity experiments at 20 °C at two different protein concentrations revealed the presence of two species (Figure S3A). The slower sedimenting species observed at 2.0 S represents the M128p monomer as the f/f_o obtained from the analysis returned molecular masses of 20 – 24 kDa (the calculated mass of the monomer M128p is 24 kDa). The faster sedimenting species at 2.9 – 3.2 S represents an M128p oligomer. As the relative amount of this species increases at the higher protein concentration, we presume that the monomer and oligomer are in equilibrium. Analysis of the sedimentation equilibrium data (Figure S3B-D) in terms of a reversible monomer-trimer returned a K_a of 8.4 x 10^{10} M^{-2} (range of 5.9 – 12.8 x 10^{10} M^{-2}) corresponding to a K_d of 3.4 ± 0.7 μM. This analysis allowed us to deconstruct the raw equilibrium data in terms of mixtures of monomers and trimers (Figure S3E-G).

Based on these results, we then modeled the sedimentation velocity data in terms of a Lamm equation describing a self-associating monomer-trimer. We obtained a K_a of 8.3 x $10^8 M^{-2}$. This corresponds to a K_d of 35 μM, ~10 times that of the K_d obtained at 4ºC from the sedimentation equilibrium data (see above). This result indicates that the monomer-trimer equilibrium is strongly temperature dependent greatly favoring the monomer at 20ºC. In addition, the data fit returns sedimentation coefficients of 2.0 S and 4.0 S for the M128p monomer and trimer, respectively. The data are additionally consistent with fast reaction kinetics as an estimated k_{off} of 7 x 10^{-4} s^{-1} is determined for the M128p trimer.

The effect of oligonucleotide length and concentration on the recovery of cross-linked ORF1p by gel electrophoresis - This experiment was carried similarly to that shown in Figure 3A, main paper. ORF1p was incubated with different concentrations of d120_c or d60_c (Table 1) or with 0.5 M NaCl for 20 minutes, and then incubated with 1 mM EGS for one hour before being subject to gel electrophoresis under denaturing conditions. The results are shown in Figure S4 and discussed in the Results section of the main paper in the context of the results shown in Figure 3A.

Nucleic acid binding by the M128p monomers - Monomeric M128p bound RNA with much lower affinity than the full-length ORF1p: the $[ORF1p]_{0.5FB}$ for RNA is ~47 nM (Figure S5 A1) Vs ~1 nM for full-length ORF1p (Figure 5B). These RNA binding data are analogous to those found with the carboxy-terminal 1/3 (C-⅓) of the mouse ORF monomer, although our measured affinity is ~10 fold lower than the apparent K_d for the mouse protein. The fact that the mouse C-⅓ protein lacks the amino-terminal half of the RRM (10) could contribute to this difference. M128p either inconsistently bound DNA or produced DNA complexes that were not stable enough to survive the filter-binding assay (Figure S5 A2 & A3).

A similar difference between the binding by the full-length ORF1p trimer and M128p monomer

was found using an EMSA to measure the binding of a single strand 53-mer DNA (Figure S5B).

References:

1. Lebowitz, J., Lewis, M.S. and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.*, **11**, 2067-2079.
2. Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.*, **320**, 104-124.
3. Hayes, D.B., Laue, T. & Philo, J. (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods in Cell Biology*, **84**, 143-179.
4. Dam, J., Velikovsky, C.A., Mariuzza, R.A., Urbanke, C. and Schuck, P. (2005) Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions c(s). *Biophys J.*, **89**, 619-634.
5. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J.*, **78**, 1606-1619.
6. Khazina, E. and Weichenrieder, O. (2009) Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proc. Natl. Acad. Sci. U S A*, **106**, 731-736.
7. Martin, S.L., Branciforte, D., Keller, D. and Bain, D.L. (2003) Trimeric structure for an essential protein in L1 retrotransposition. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 13815-13820.
8. Basame, S., Wai-lun Li, P., Howard, G., Branciforte, D., Keller, D. and Martin, S.L. (2006) Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition. *J. Mol. Biol.*, **357**, 351-357.
9. Wolf, E., Kim, P.S. and Berger, B. (1997) MultiCoil: A program for predicting two-and three-stranded coiled coils. *Protein Science*, **6**, 1179-1189.
10. Martin, S.L., Li, J. and Weisz, J.A. (2000) Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1. *J. Mol. Biol.*, **304**, 11-20.
11. Januszyk, K., Li, P.W.-l., Villareal, V., Branciforte, D., Wu, H., Xie, Y., Feigon, J., Loo, J.A., Martin, S.L. and Clubb, R.T. (2007) Identification and Solution Structure of a Highly Conserved C-terminal Domain within ORF1p Required for Retrotransposition of Long Interspersed Nuclear Element-1. *J. Biol. Chem.*, **282**, 24893-24904.

Figure Legends:

Figure S1 ORF1p products synthesized in *E. coli* - **(A)** The human ORF1p amino acid sequence. The heptads (yellow and green rectangles) of the predicted (9) coiled coil domain (red), and Edman degradation products (blue and green) are highlighted. The RNA recognition motif (RRM) and C-terminal domain (CTD) are also indicated (respectively refs., 6,11). **(B)** Amino acid residues 114-128 from human ORF1p and the corresponding DNA sequence. Non-canonical Shine-Dalgarno sites are indicated. **(C)** SDS PAGE of purified preparations (described in the

Experimental Procedures) of wild type ORF1p and both double methionine mutant ORF1p proteins, M121F & M125F, and M121A & M125A. The lanes correspond to sequential fractions eluted from a HiLoad 26/60 Superdex-200 size exclusion column. The full-length 40 kDa ORF1p and internally initiated 24 kDa products are indicated.

Figure S2 Analytical ultracentrifugation of ORF1p - The protein purified from baculovirus infected insect cells as described in the Experimental Procedures is a monodisperse trimer**. (A)** The c(*s*) distributions obtained for ORF1p (at 6.5 μM - green, and 13.0 μM -red) are based on sedimentation velocity absorbance data collected at 55 krpm and 20.0 ºC. **(B & C)** Sedimentation equilibrium profiles for ORF1p at 4ºC plotted as a distribution of the absorbance at 280 nm *vs.* radius at equilibrium. Data were collected at 11 (orange), 14 (yellow) and 17 (brown) krpm and loading concentrations of 18.5 μM (B) and 37.9 μM (C). The solid lines show the best-fit analysis in terms of a single ideal solute with the corresponding residuals shown in the plots above.

Figure S3. Analytical centrifugation of M128p - The protein purified from *E. coli* as described in the Experimental Procedure self-associates reversibly to form a trimer. **(A)** The c(*s*) distributions obtained for M128p (8.8 μM – blue, and 18.3 μM - red) are based on sedimentation velocity absorbance data collected at 50 krpm and 20.0ºC. **(B-D)** Sedimentation equilibrium profiles obtained for M128p shown in terms of A_{280} versus the radius (R) for data collected at loading concentrations of 6.1 μM (**B, E**), 12.2 μM (**C, F**) and 18.8 μM (**D, G**). Orange, yellow, green, and red symbols (**B-E**) represent the data collected at 12, 16, 20 and 24 krpm respectively. The black lines are the results of fitting these data in terms of a reversible monomer-trimer self-association with mass conservation and the corresponding residuals are shown above their respective sedimentation equilibrium profile data. (**E-G**), the contributions to the absorbance profiles calculated for the M128p trimer (solid lines) and monomer (dashed lines)

Figure S4 ORF1p polymerization as a function of oligonucleotide length and concentration. The stated molecular weights (kDa) of the commercial protein markers, Mark12™ (Invitrogen) are indicated. The percent protein recovered on the gel was determined by densitometry using NIH ImageJ and normalized to the amount recovered as trimer cross linked in 0.5 M NaCl, which was set to 100%. These values are given as "% on gel relative to 0.5 M NaCl".

Figure S5 Nucleic acid binding by M128p. **(A)** Replicate filter binding assays are shown in each panel. The sequences of the oligonucleotides are given in Table 1. **(A1)** The 30-mer RNA is r30_3'utr-a **(A2)** The 29-mer single strand DNA is d29_c. **(A3)** The 29-mer mismatched duplex is d29:[^{32}P]-d29_cmm. **(B)** Results of an EMSA with ORF1p or M128p and 58-mer single strand DNA [^{32}P]- d58_pT-b. Protein concentrations are in terms of monomer.

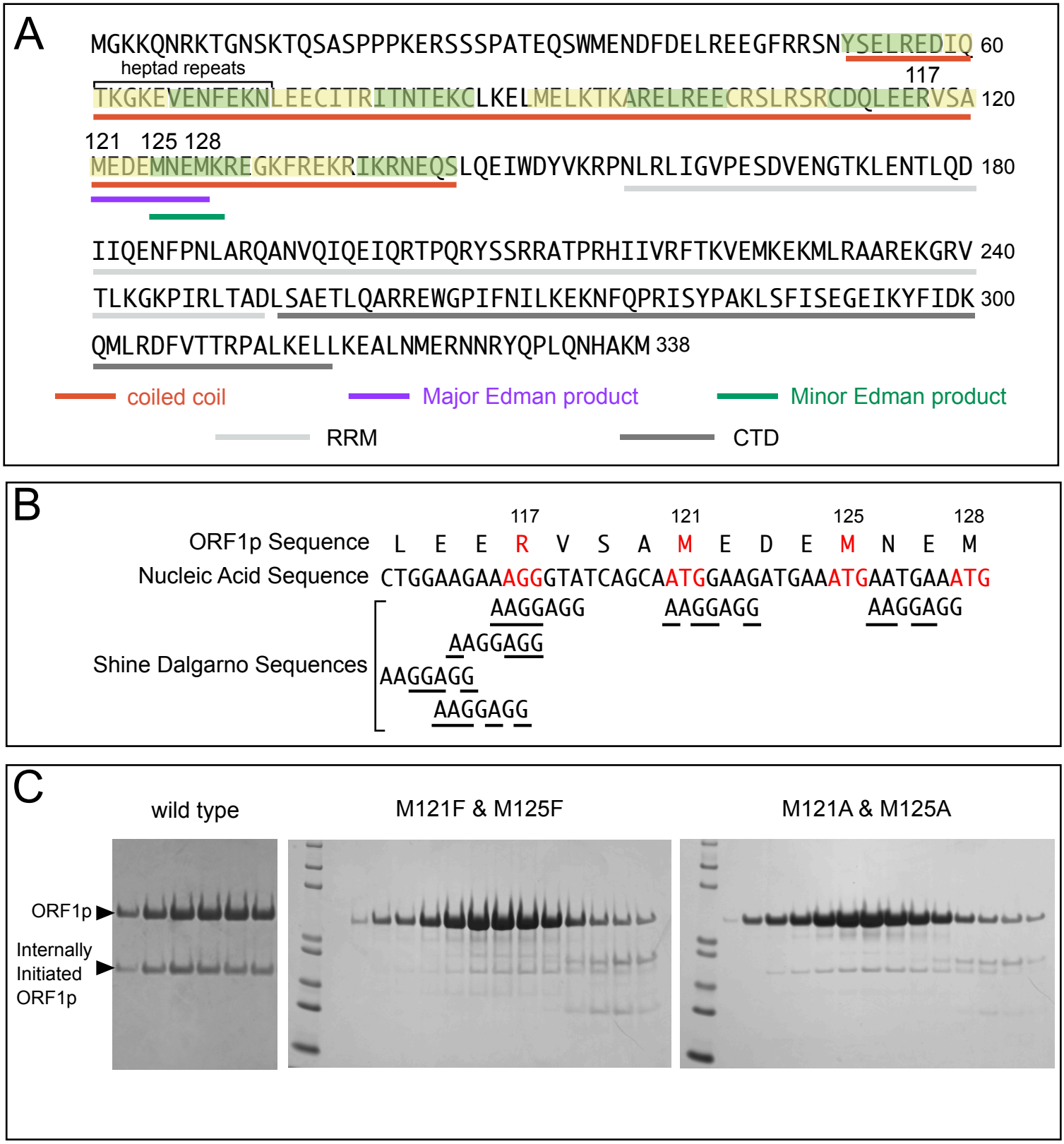

Figure S1 ORF1p products synthesized in *Escherichia coli*

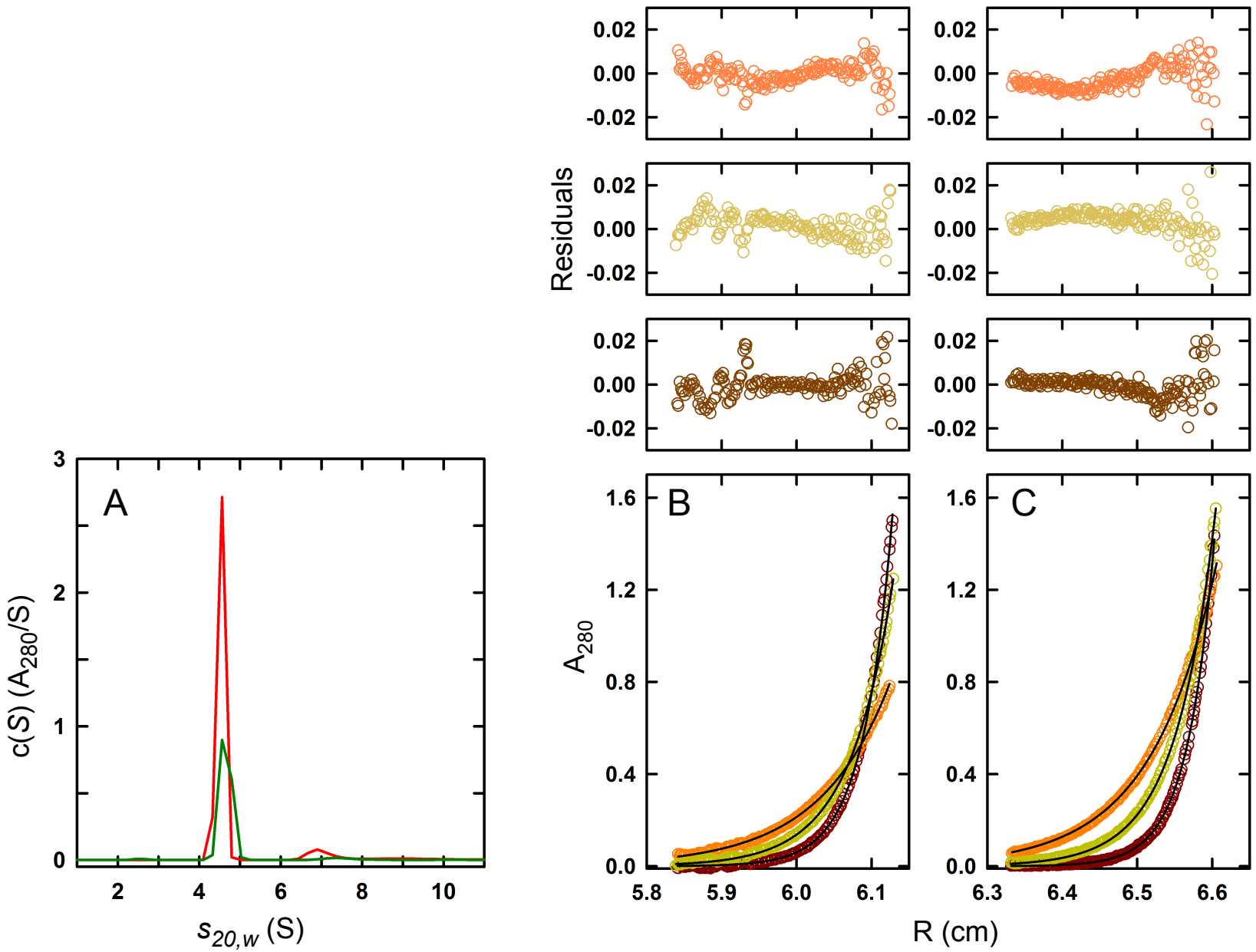

Figure S2 Analytical centrifugation of ORF1p

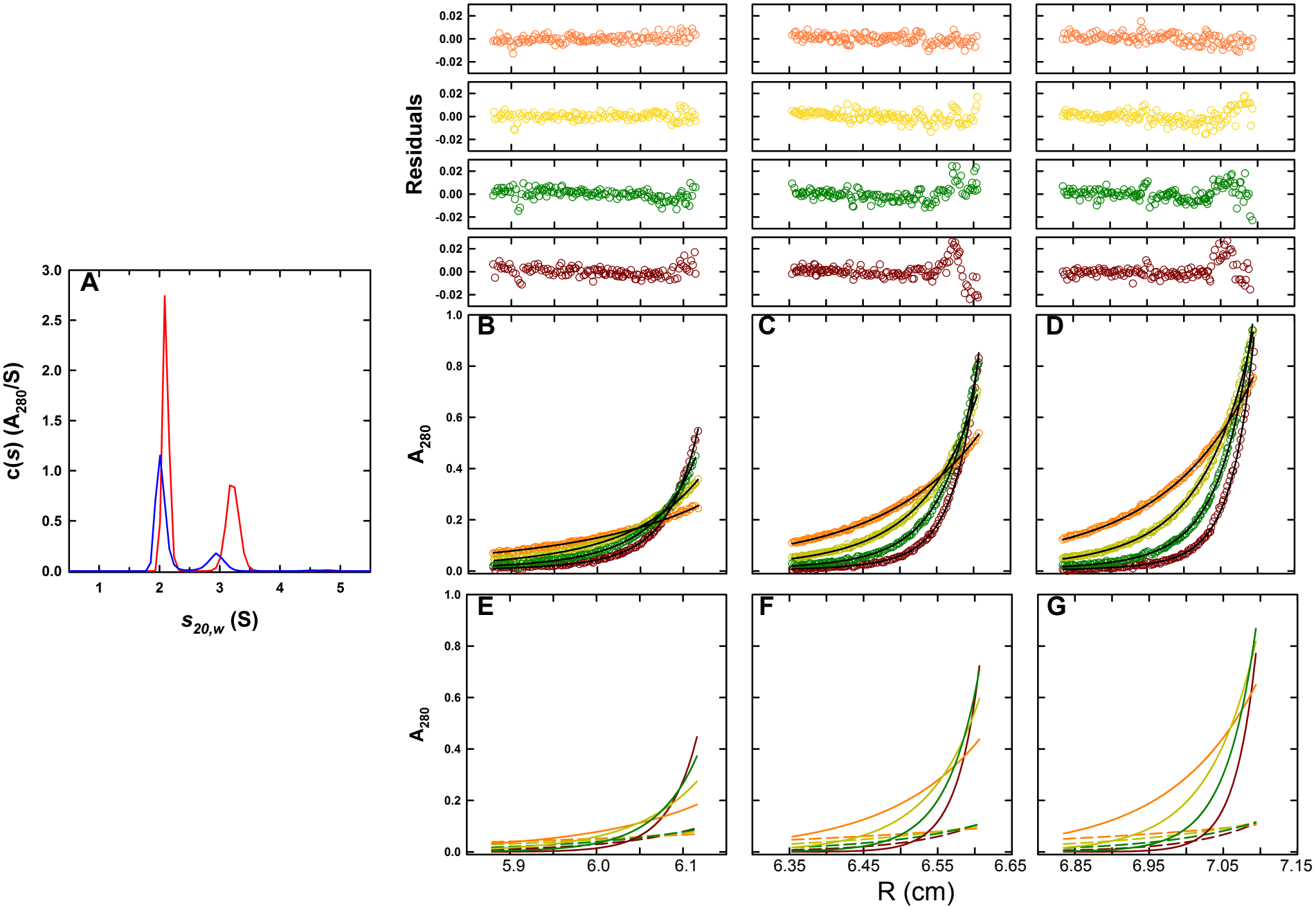

Figure S3 Analytical Centrifugation of the truncated M128p monomer

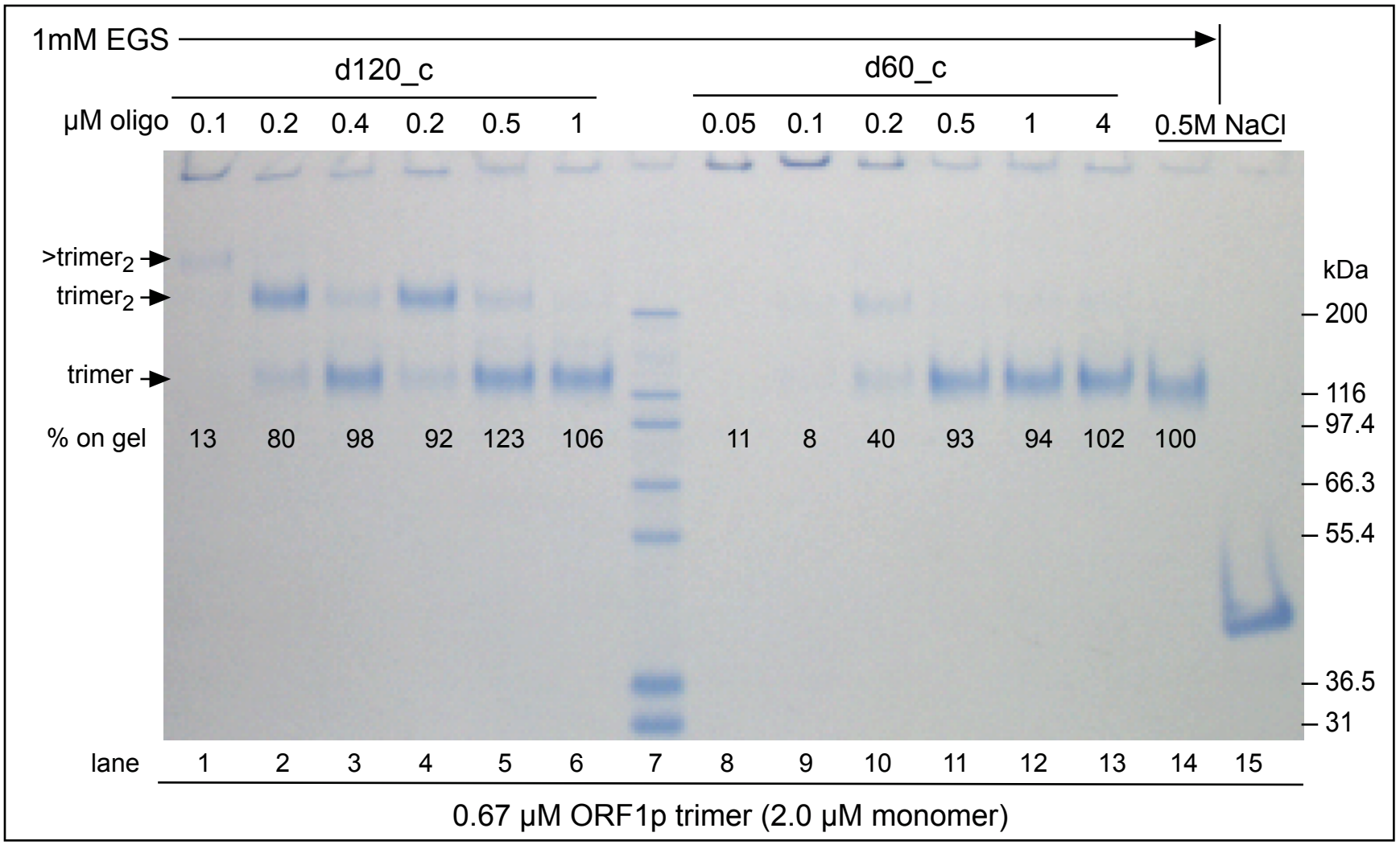

Figure S4 - ORF1p polymerization as a function of oligonucleotide length & concentration

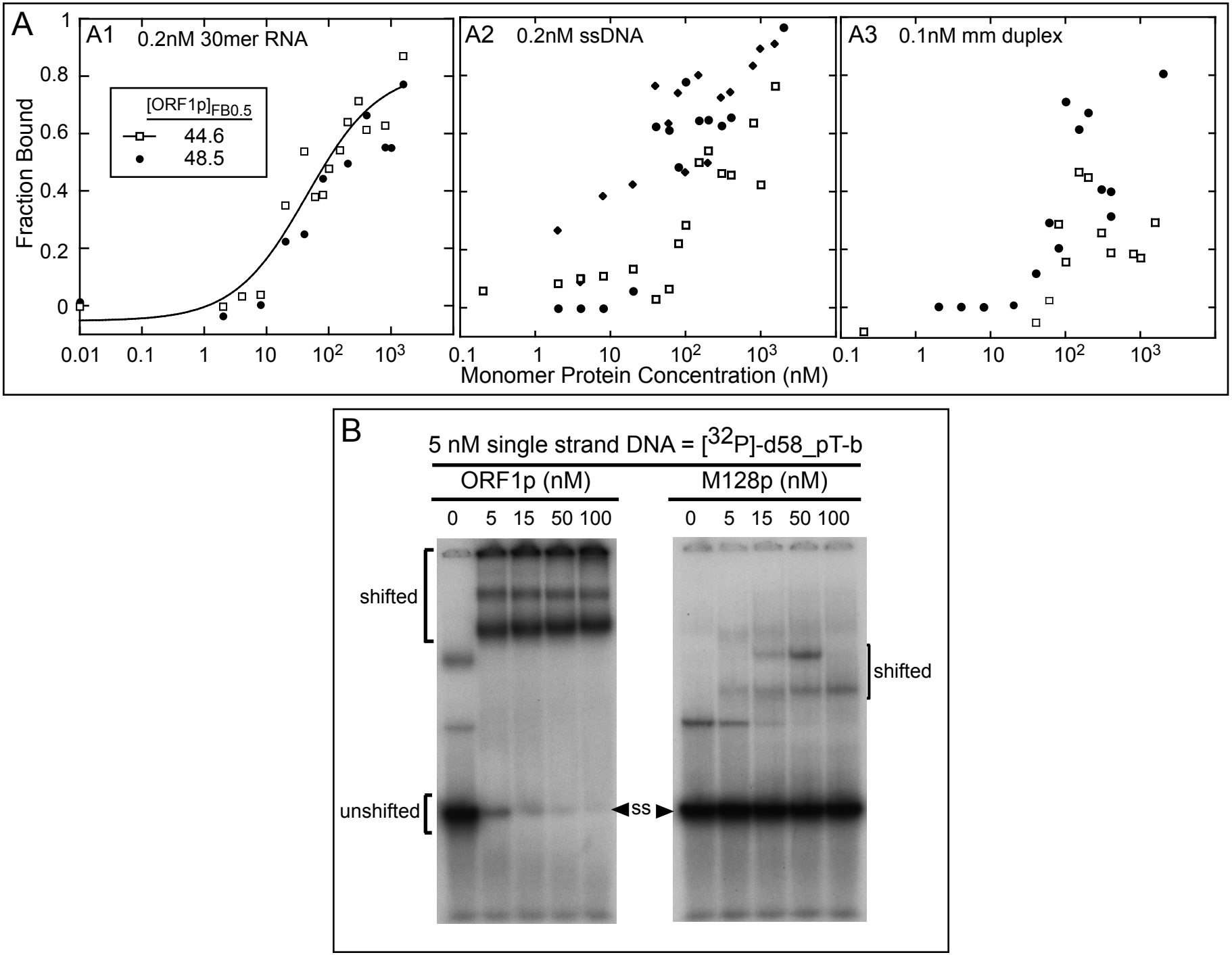

Figure S5 Nucleic acid binding by the M128p monomer