Supporting Material for

The Conformational Ensembles of α-Synuclein and Tau: Combining Single-Molecule FRET and Simulations

Abhinav Nath[1]*, Maria Sammalkorpi[2,3], David C. DeWitt[1], Adam J. Trexler[1], Shana Elbaum-Garfinkle[1], Corey S. O'Hern[4,5]*, and Elizabeth Rhoades[1,5]*

[1] Department of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, P. O. Box 208114, New Haven, CT 06520-8114
[2] Department of Chemical & Environmental Engineering, Yale University, P. O. Box 208286, New Haven, CT 06520-8286
[3] Department of Chemistry, Aalto University, P.O. Box 16100, FI-00076 Aalto, Finland
[4] Department of Mechanical Engineering & Materials Science, Yale University, P. O. Box 208286, New Haven, CT 06520-8286
[5] Department of Physics, Yale University, P. O. Box 208120, New Haven, CT 06520-8120

Single-molecule FRET Procedures

SmFRET measurements of αS and tau in solution were performed as described previously (1, 2). Briefly, αS and tau were expressed in and purified from *E. coli* with pairs of cysteine residues introduced at selected labeling positions using two rounds of site-directed mutagenesis. Constructs were labeled with donor (Alexa Fluor 488) and acceptor (Alexa Fluor 594) using thiol-maleimide chemistry under conditions which optimize the fraction of double-labeled molecules. Fluorescence anisotropy and lifetime measurements were performed at each dye position to ensure that energy transfer efficiency (ET_{eff}) values were not affected by hindered dye rotation or local environment artifacts. smFRET measurements were performed on an Olympus IX-71 inverted microscope (Olympus America, Center Valley, PA): a 488 nm laser (Newport, Franklin, MA) was focused into a dilute (~50 pM) solution of double-labeled protein, and emitted fluorescence was split by a 585nm long-pass dichroic mirror (Chroma, Bellows Falls, VT) into donor and acceptor channels. Each channel was transmitted through 100 μm optical fibers to avalanche photodiode detectors (PerkinElmer, Waltham, MA), and photon traces were collected in 1 ms time bins. Bursts were identified based on a total photon threshold, and ET_{eff} values were calculated taking into account the cross-talk between donor and acceptor channels, and differences in detection efficiency and quantum yield between the two dyes. Presented values are the means of at least 3 independent measurements.

Calculation of Distance Constraints

IDPs such as αS and tau rapidly sample a range of conformations in solution, with intra-chain contact rates in the 1-10 μs range (4, 5). A single smFRET measurement thus represents a time-average of ET_{eff} over the timescale of a single transit through the observation volume (~1 ms). To properly account for the relevant distribution of distances between donor- and acceptor-labeled residues, energy transfer efficiency is determined by a convolution of the well-known Förster equation with the inter-residue distance distribution $P(r)$, as follows:

$$ET_{eff} = \int_0^\infty \frac{P(r)}{1+(r/R_0)^6}\,dr \tag{S1}$$

Here, R_0 is the characteristic Förster distance at which the energy transfer efficiency from donor to acceptor is 50%. To determine the form of $P(r)$ for IDPs, we performed unconstrained excluded-volume MC simulations of αS and tau, which represent the self-avoiding random-coil limit for these systems. Notably, for a given mean inter-residue distance, these excluded-volume ensembles have substantially narrower separation distributions $P(r)$ than those predicted for Gaussian chains (Figure S1a, inset):

$$P_G(r) = \frac{4\pi r^2}{\left(\frac{2}{3}\pi r_{RMS}^2\right)^{3/2}} \exp\left(-\frac{3r^2}{2r_{RMS}^2}\right), \tag{S2}$$

where r_{RMS} is the root-mean-square end-to-end distance of a Gaussian chain. As a point of comparison, simulations of self-avoiding random walks on 3-dimensional cubic lattices (6, 7) indicate that the analogous distributions for excluded-volume chains take the form:

$$P_{EV}(r) \propto \left(\frac{r\sqrt{3}}{r_{RMS}}\right)^{\theta} \exp\left(-D\left(\frac{r\sqrt{3}}{r_{RMS}}\right)^{\delta}\right) \tag{S3}$$

with $\theta \approx 0.27$, $D \approx 0.33$ and $\delta \approx 2.5$, so that $P_{EV}(r)$ is more symmetric and narrower than $P_G(r)$ for a given value of r_{RMS}. The $P(r)$ distributions generated by excluded-volume MC are qualitatively closer to the $P_{EV}(r)$ form, but are not directly comparable both because the polypeptide chains in our simulations are not confined to a lattice, and because of finite-size effects. Instead, the excluded-volume MC $P(r)$ distributions were reasonably well-described by normal distributions with mean and standard deviation increasing with sequence separation (Figure S1a). The narrower distributions observed in our MC simulations and the 3-dimensional lattice models (6, 7) compared to those from the Gaussian chain model result from excluded-volume effects and the fact that the distribution of inter-residue separations is measured between two internal points (rather than the endpoints) of the protein.

ET_{eff} was calculated using Equation S1 with an R_0 value of 5.4 nm (for the Alexa 488-Alexa 594 dye pair) and the $P(r)$ distribution observed in the excluded-volume MC ensembles for a given sequence separation. This provides the empirical relationship between ET_{eff} and inter-residue distance for excluded-volume random coils, and is intermediate between the unmodified Förster equation and the relationship predicted for Gaussian chains (3) (Figure S1b). By inverting this relationship, we can estimate the equivalent mean and standard deviation of the $P(r)$ distribution related to any ET_{eff} value measured for an excluded-volume random coil (Figure S1c). Mean inter-residue distances and the respective standard deviations were calculated in this way for each of the experimental smFRET measurements (Tables S1, S2). Constraints were then implemented in the ECMC simulations by adding the following term to the excluded-volume Rosetta (repulsive LJ) potential:

$$U_{constr} = -k_b T' \sum_{i=1}^{M} \ln\left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(r-\mu_i)^2}{2\sigma_i^2} \right) \right) \qquad \text{(S4)}$$

Here, M is the total number of constraints, r is the inter-residue distance, μ_i and σ_i are the mean and standard deviation of the $P(r)$ distribution calculated from the ET_{eff} value of constraint i, and $k_b T'$ is an effective temperature (fixed to 1) that controls the weight of the smFRET-derived constraints relative to the excluded-volume Rosetta term. For small deviations from the constraints, the resulting potential effectively includes a harmonic spring for each constraint, with a specified mean inter-residue distance and spring constant that is inversely proportional to the variance in the inter-residue separation.

Error Estimation

Major sources of potential error in the conversion of smFRET data to ECMC constraints include experimental uncertainty in the measurement of ET_{eff}, and variations in the effective Förster radius (R_0) due to anisotropic tumbling of the donor and acceptor fluorophores, or changes in the donor quantum yield. A resampling procedure was used to estimate the sensitivity of ECMC results to each of these sources of uncertainty, and to ensure that the errors were correctly and rigorously propagated throughout our results. Uncertainty due to experimental variability in smFRET is relatively straightforward to account for, since it is approximately normally distributed. The experimental uncertainty in ET_{eff} was estimated to be 0.012 based on 23 measurements of a standard sample (a 10-bp dsDNA construct double-labeled with Alexa 488 and Alexa 594), and was incorporated into the error estimation procedure as described below.

This uncertainty in ET_{eff} measurements is compounded, however, by potential variations in R_0 due to changes in fluorophore quantum yield or anisotropy induced by the local protein environment. R_0 is proportional to the sixth roots of the orientation factor κ^2 and the donor quantum yield Q_D.

Changes in quantum yield due to dynamic quenching of Alexa 488 and Alexa 594 by the local protein environment upon dye labeling were assessed by fluorescence lifetime measurements, collected on a Horiba-Jobin-Yvon Fluorolog-3 instrument equipped with time-correlated single-photon counting (TCSPC) detection and pulsed LED illumination. In each case, measured lifetimes were minimally perturbed from the values for the free dye. Conservatively, the uncertainty in quantum yield was therefore estimated to be 5%.

Steady-state anisotropy measurements were performed to estimate the degree to which the isotropic tumbling of fluorophores is hindered by the protein environment. The orientation factor κ^2, which is commonly assigned a value of 0.67 based on an assumption of perfect isotropic tumbling, can in fact sample a wide range of values based on measured anisotropies of three angles: the angle between the donor emission dipole and the inter-dye vector (θ_D); the angle between the acceptor excitation dipole and the inter-dye vector (θ_A); and the azimuthal angle between the relevant donor and acceptor dipoles, ϕ. The following relationship applies (8):

$$\begin{aligned}\kappa^2 = &\left(\sin\theta_D \sin\theta_G \cos\phi - 2\cos\theta_D \cos\theta_A\right)^2 \left\langle d_D^x \right\rangle \left\langle d_A^x \right\rangle + \tfrac{1}{3}\left(1 - \left\langle d_D^x \right\rangle\right) + \tfrac{1}{3}\left(1 - \left\langle d_A^x \right\rangle\right) \\ &+ \cos^2\theta_D \left\langle d_D^x \right\rangle \left(1 - \left\langle d_A^x \right\rangle\right) + \cos^2\theta_A \left\langle d_A^x \right\rangle \left(1 - \left\langle d_D^x \right\rangle\right)\end{aligned} \tag{S5}$$

where $\left\langle d_{D,A}^x \right\rangle \approx \sqrt{r_{D,A} / 0.4}$ is determined by the mean measured anisotropy r of each fluorophore.

Since the three angles θ_D, θ_D and ϕ are unknown, they were each sampled at random from a uniform distribution over the range [0, 2π] to provide values for κ^2. This sampling procedure was repeated 1000 times to produce a realistic distribution of accessible κ^2 values (Figure S3). This distribution was converted to a distribution of perturbations in R_0 using the following relationship:

$$\frac{R_0'}{R_0} = \sqrt[6]{\frac{\kappa_i^2}{0.67}} \tag{S6}$$

where ${\kappa_i}^2$ is a random sample from the full set of possible κ^2 values. It is evident that the uncertainties in both κ^2 and R_0 are not normally distributed. Resampling error estimation is thus an appropriate method to accurately propagate and quantify the resulting uncertainties in ECMC ensembles.

Briefly, 10 decoy sets of constraints were generated incorporating estimated uncertainties due to anisotropic dye tumbling, changes in donor quantum yield, and experimental error in ET_{eff} measurement. For each decoy set, the expected value of R_0 (5.4 nm) was multiplied by a random sample from $P(R_0'/R_0)$ (*i.e.*, the blue histogram in Figure S3, to account for anisotropic tumbling), and then by the 6th root of a random sample from a Gaussian distribution with mean 1 and standard deviation 0.05 (to account for donor quantum yield effects). This modified value of R_0 was then used to estimate the empirical relationships between ET_{eff} and mean and standard deviation of the $P(r)$ distribution for an excluded-volume random coil using equation S1 as described above. Values drawn at random from a Gaussian distribution with zero mean and a standard deviation of 0.012 were added to experimental ET_{eff} values to model ET_{eff} measurement errors. These modified ET_{eff} values were then used to determine the mean (μ_i) and standard deviation (σ_i) for ECMC constraints using equation S4. ECMC simulations were performed on the 10 independently resampled decoy constraint sets to generate a set of 10 ensembles for statistical analysis. Based on this analysis, the standard error of the mean for values reported here is ~1.5% (e.g., 0.045 nm for the mean R_g value of 3.3 nm obtained for αS at pH 7.4).

Comparison between Ensembles from AAMD, UMC and ECMC

Mean R_g and inter-residue distance values of αS derived from AAMD, UMC and ECMC were in good agreement. However, the properties of the ensembles generated by the three computational approaches differed significantly in other ways. AAMD and UMC displayed greater conformational variation than ECMC, with 33% and 17%, respectively, of structures in each ensemble having end-to-end distances < 5 nm, as compared to 11% for ECMC. This difference is also reflected in broader R_g and inter-residue distance distributions for the two unconstrained methods (AAMD and UMC) than ECMC; in particular, the AAMD distributions appeared to be bimodal. Furthermore, the mean R_g scaling exponent ν was 0.51 and 0.52 for AAMD and UMC respectively, compared to the ECMC value of 0.39 (Figure S4a). To examine results from the three approaches in more detail, we constructed distance difference maps comparing AAMD and UMC to ECMC (Figure S4b). Relative to the two unconstrained methods, ECMC-derived ensembles display expansion of the span from residues 60 to 120, coupled with compaction between residues ~40 and ~60. In general, the AAMD and UMC ensembles appeared to be more similar to each other than to ECMC-derived ensembles. Studies are currently underway to investigate which features of the computational force fields are responsible for the similarities and differences between the AAMD, UMC and ECMC approaches.

Supporting References

1. Trexler, A.J., and E. Rhoades. 2010. Single Molecule Characterization of α-Synuclein in Aggregation-Prone States. *Biophys. J.* 99: 3048-3055.

2. Elbaum-Garfinkle, S., and E. Rhoades. 2012. Identification of an aggregation-prone structure of tau. *J. Am. Chem. Soc.* 134: 16607-16613.

3. O'Brien, E.P., G. Morrison, B.R. Brooks, and D. Thirumalai. 2009. How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *J. Chem. Phys.* 130: 124903.

4. Lee, J.C., H.B. Gray, and J.R. Winkler. 2005. Tertiary contact formation in alpha-synuclein probed by electron transfer. *J. Am. Chem. Soc.* 127: 16388-16389.

5. Lee, J.C., B.T. Lai, J.J. Kozak, H.B. Gray, and J.R. Winkler. 2007. Alpha-synuclein tertiary contact dynamics. *J. Phys. Chem. B.* 111: 2107-2112.

6. Wittkop, M., S. Kreitmeier, and D. Göritz. 1996. The distribution function of internal distances of a single polymer chain with excluded volume in two and three dimensions: A Monte Carlo study. *J. Chem. Phys.* 104: 351-358.

7. Caracciolo, S., M.S. Causo, and A. Pelissetto. 2000. End-to-end distribution function for dilute polymers. *J. Chem. Phys.* 112: 7693-7710.

8. Dale, R., J. Eisinger, and W. Blumberg. 1979. The orientational freedom of molecular probes: the orientation factor in intermolecular energy transfer. *Biophys. J.* 26: 161-193.

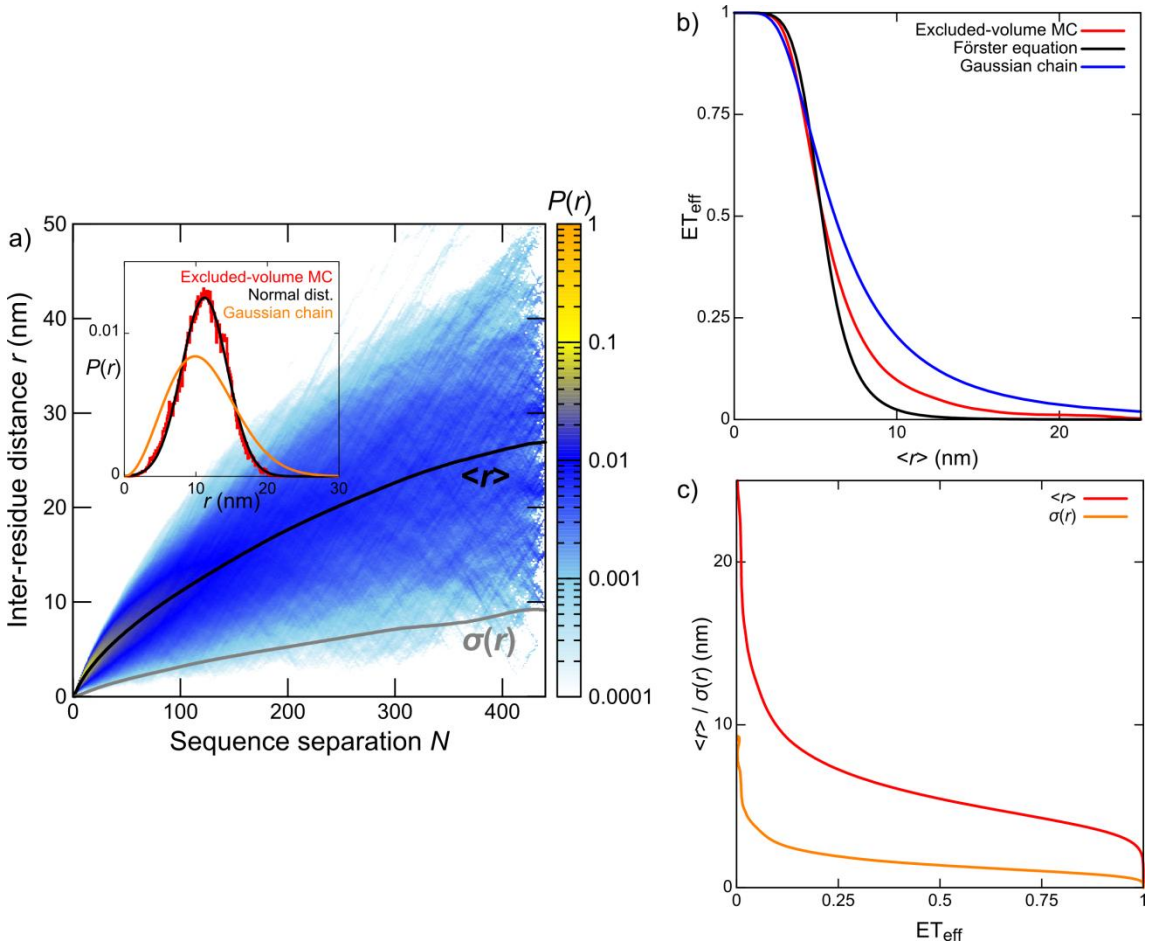

Figure S1. (a) Inter-residue distance distribution $P(r)$ of an unconstrained excluded-volume MC simulation as a function of sequence separation. Mean and standard deviation of $P(r)$ are displayed as black and gray lines, respectively. Inset: plot of $P(r)$ distribution at a sequence separation of 100 residues (red), demonstrating that the distributions from excluded-volume MC calculations are much better described by normal distributions (black) than by the Gaussian chain model (orange). (b) Relationship between inter-dye distance and ET_{eff}, calculated using the Förster equation alone (black), the excluded-volume approach described here (red), and the Gaussian chain model (blue). (c) The expected means and standard deviations of $P(r)$ distributions corresponding to different values of ET_{eff}. These mean and standard deviation values were used to parametrize harmonic constraints for ECMC calculations based on measured ET_{eff} values.

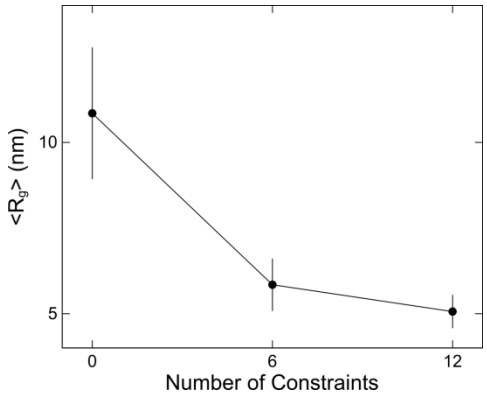

Figure S2. Mean R_g values obtained by ECMC for tau in solution using 0, 6 or 12 smFRET constraints, demonstrating similar asymptotic behavior to αS (cf. Figure 2a): most of the compaction relative to the random-coil limit is observed even with a partial set of 6 constraints.

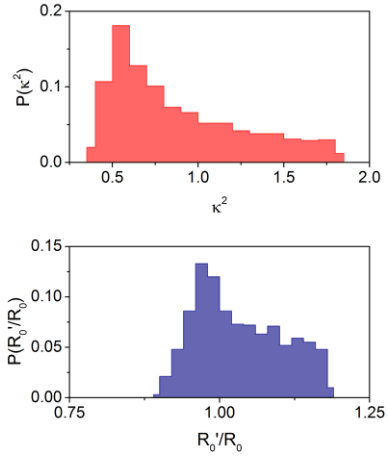

Figure S3. (Top) Distribution of the orientation factor κ^2 based on 1000 random samples for the three angles describing the dye geometry θ_D, θ_D and ϕ on the interval [0, 2π] and typical experimental anisotropy values for labeled proteins. (Bottom) Distribution of perturbations in Förster radius R_0 due to uncertainty in κ^2.

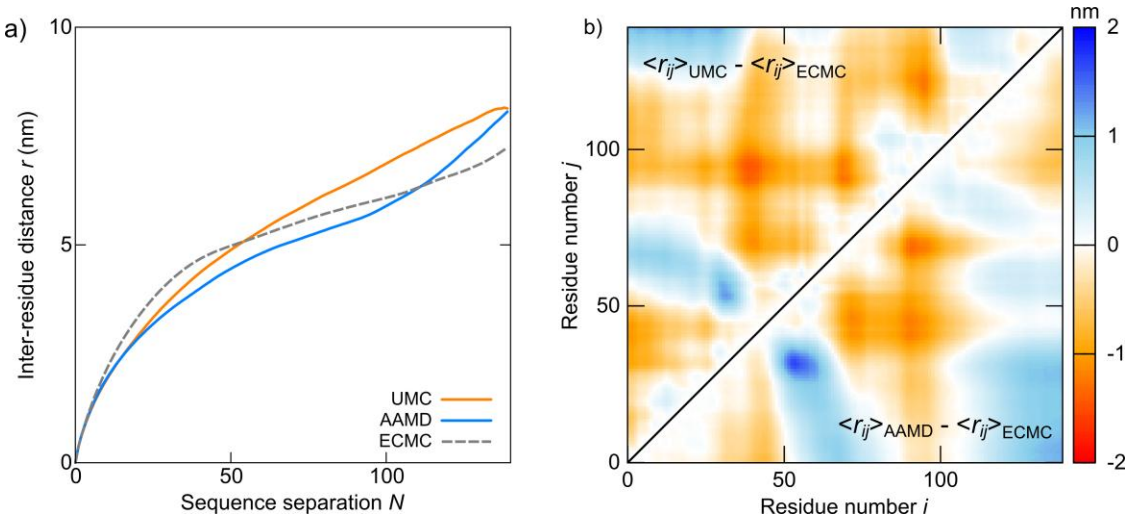

Figure S4. (a) The scaling of the inter-residue separation with N for UMC (orange) and AAMD (blue) ensembles compared to ECMC results (grey dashed line). (b) Distance difference maps comparing UMC (top) and AAMD (bottom) ensembles with ECMC results. The map illustrates local differences in structure between the three approaches. Compaction relative to ECMC is indicated by orange-red, and expansion by blue.

Table S1. ECMC constraint mean and standard deviation values corresponding to ET_{eff} values previously measured (1) for 12 residue pairs of αS at pH 7.4 and pH 3.0.

	pH 7.4			pH 3.0		
Residue pair	ET_{eff}	μ_i (nm)	σ_i (nm)	ET_{eff}	μ_i (nm)	σ_i (nm)
9-130	0.36	6.34	1.63	0.36	6.34	1.63
33-130	0.53	5.28	1.32	0.70	4.46	1.07
54-130	0.50	5.40	1.35	0.67	4.64	1.13
72-130	0.55	5.16	1.29	0.77	4.18	0.99
92-130	0.65	4.70	1.15	0.86	3.73	0.83
33-72	0.72	4.40	1.05	0.67	4.64	1.13
9-54	0.70	4.46	1.07	0.70	4.46	1.07
72-92	0.85	3.79	0.85	0.93	3.25	0.70
54-72	0.85	3.79	0.85	0.91	3.39	0.75
9-72	0.64	4.77	1.15	-	-	-
9-33	0.86	3.73	0.83	-	-	-
54-92	0.66	4.68	1.13	-	-	-

Table S2. ECMC constraint mean and standard deviation values corresponding to ET_{eff} values previously measured (2) for 12 residue pairs of tau in the absence and presence of heparin.

	- heparin			+ heparin		
Residue pair	ET_{eff}	μ_i (nm)	σ_i (nm)	ET_{eff}	μ_i (nm)	σ_i (nm)
17-433	0.22	7.61	2.02	< 0.1	12.5	50
17-291	0.42	5.92	1.50	< 0.1	12.5	50
17-103	0.17	8.30	2.24	0.17	8.30	2.24
103-184	0.70	4.50	1.08	0.35	6.39	1.65
103-291	0.33	6.53	1.69	0.14	8.85	2.42
291-433	0.42	5.92	1.50	0.30	6.78	1.77
322-433	0.51	5.40	1.37	0.35	6.39	1.65
354-433	0.62	4.86	1.19	0.39	6.10	1.57
291-322	0.81	3.99	0.92	0.81	3.99	0.92
291-354	0.61	4.91	1.21	0.68	4.59	1.11
244-354	0.37	6.23	1.60	0.52	5.35	1.34
184-291	0.36	6.31	1.62	0.47	5.62	1.42