Biophysical Journal, Volume *115*

Supplemental Information

Confidence Analysis of DEER Data and Its Structural Interpretation with Ensemble-Biased Metadynamics

Eric J. Hustedt, Fabrizio Marinelli, Richard A. Stein, José D. Faraldo-Gómez, and Hassane S. Mchaourab

SUPPLEMENTAL TABLES

Table S1. Model selection for fit in Fig. 4.

n	χ_v^2	ΔBIC
1	1.074	0.0
2	1.050	7.0

Table S2. Best-fit parameters for the simulated DEER signal in Fig 4.

	Δ	λ	scale	r_0	σ_r	χ_v^2
Best-fit values	0.308	4.83	1.0025	32.75	2.60	1.074
Uncertainties [a]	±0.039	±0.88	±0.0024	±0.57	±1.15	
CI lower limit [b]	-0.016	ND[g]	-0.0018	-0.25	-0.49	
CI upper limit [b]	+0.018	+0.26	+0.0018	+0.27	+0.51	
Average of 10000 best-fit values [c]	0.300	4.92	1.0001	32.50	2.49	1.000
2× std. dev. of 10000 best-fit values [d]	0.023	0.67	0.0019	0.35	0.71	0.101
Average of 10000 uncertainties [e]	±0.031	±1.09	±0.0021	±0.47	±0.93	
True values [f]	**0.3**	**5.0**	**1.0**	**32.5**	**2.5**	

[a] 2σ uncertainties estimated using Eqs. 15-17.
[b] Values given correspond to the 2σ confidence level.
[c] Average best-fit parameters for 10000 replicate simulated data sets.
[d] Twice the standard deviations of the best-fit parameters for 10000 replicate simulated data sets.
[e] Average of the 2σ uncertainties estimated from fits to 10000 replicate simulated data sets.
[f] The true parameters used to generate the simulated data.
[g] χ_v^2 does not reach the 2σ level for any λ less than the best-fit value

Table S3. Model selection for fits in Fig. 5.

	n	χ_v^2	ΔBIC
equal-width components	1	2.144	130.2
	2	1.706	72.0
	3	1.300	0.0
	4	1.311	16.9
variable-width components	1	1.542	89.0
	2	1.275	42.9
	3	1.065	0.0
	4	1.060	12.7

Table S4. Best-fit parameters for the simulated multi-component signals in Fig. 5.

equal-width components	Δ	λ	scale	r_{0_1}	σ_{r_1}	r_{0_2}	σ_{r_2}	a_2 (f_2) [g]	r_{0_3}	σ_{r_3}	a_3 (f_3) [g]	χ^2_ν
Best-fit values	0.3059	4.981	1.0035	25.00	2.82	34.61	1.95	0.650 (0.274)	44.67	3.51	0.579 (0.376)	1.300
Uncertainties [a]	±0.0038	±0.013	±0.0036	±0.41	±0.53	±0.31	±0.43	±0.026	±0.38	±0.70	±0.058	
CI lower limit [b]	-0.0037	-0.014	-0.0035	-0.41	-0.50	-0.32	-0.40	-0.029	-0.43	-0.65	-0.055	
CI upper limit [b]	0.0039	0.013	0.0037	0.43	0.60	0.31	0.48	0.026	0.35	0.75	0.063	
Average of 10000 best-fit values [c]	0.3001	4.999	1.0000	25.01	2.50	34.99	2.49	0.666 (0.331)	44.98	2.53	0.503 (0.335)	1.000
2× std. dev. of 10000 best-fit values [d]	0.0040	0.014	0.0039	0.48	0.59	0.40	0.63	0.032	0.33	0.76	0.063	0.104
Average of 10000 uncertainties [e]	±0.0040	±0.014	±0.0039	±0.47	±0.58	±0.41	±0.62	±0.032	±0.34	±0.76	±0.062	
True values [f]	**0.3**	**5.0**	**1.0**	**25.0**	**2.5**	**35.0**	**2.5**	**0.667**	**45.0**	**2.5**	**0.5**	
variable-width components	Δ	λ	scale	r_{0_1}	σ_{r_1}	r_{0_2}	σ_{r_2}	a_2 (f_2) [g]	r_{0_3}	σ_{r_3}	a_3 (f_3) [g]	χ^2_ν
Best-fit values	0.3013	4.985	1.0041	45.07	7.70	32.74	2.31	0.479 (0.224)	22.87	1.70	0.533 (0.255)	1.064
Uncertainties[a]	±0.0117	±0.044	±0.0039	±2.06	±2.93	±0.54	±0.77	±0.134	±0.38	±0.46	±0.115	
CI lower limit [b]	-0.0094	-0.043	-0.0038	-2.21	-2.17	-0.55	-0.63	-0.133	-0.39	-0.39	-0.087	
CI upper limit [b]	0.0117	0.035	0.0039	1.37	2.71	0.53	0.73	0.093	0.38	0.48	0.099	
Average of 10000 best-fit values [c]	0.3002	5.000	1.0002	44.83	7.55	32.57	2.44	0.493 (0.245)	22.49	1.47	0.504 (0.255)	1.000
2× std. dev. of 10000 best-fit values [d]	0.0119	0.042	0.0041	2.36	3.06	0.62	0.88	0.145	0.38	0.45	0.112	0.104
Average of 10000 uncertainties [e]	±0.0116	±0.041	±0.0041	±2.17	±2.96	±0.62	±0.82	±0.136	±0.37	±0.45	±0.107	
True values [f]	**0.3**	**5.0**	**1.0**	**45.0**	**7.5**	**32.5**	**2.5**	**0.500**	**22.5**	**1.5**	**0.5**	

[a] 2σ uncertainties estimated using Eqs. 15-17.
[b] Upper and lower parameter bounds determined from confidence intervals. Values given correspond to the 2σ confidence level.
[c] Average best-fit parameters for 10000 replicate simulated data sets.
[d] Twice the standard deviations of the best-fit parameters for 10000 replicate simulated data sets.
[e] Average of the 2σ uncertainties estimated from fits to 10000 replicate simulated data sets.
[f] The true parameters used to simulate data.
[g] The *f* values are calculated for the best-fit *a* values from Eqs 9-10.

Table S5. Effect of r_0 on uncertainty in $P(R)$ parameters. 500 replicate data sets were generated for a unimodal Gaussian distance distribution with the given true r_0 and σ_r values. Data were calculated for t= -128 to 2400 ns with a time increment of 8 ns. Δ = 0.3 and λ = 5. The standard deviations were determined from the 500 best-fit r_0 and σ_r values. Modulation corresponding to r_0= 50 Å will have a period of 2400 ns.

True [a]		Standard Deviation [b]	
r_0	σ_r	r_0	σ_r
25	2.5	0.05	0.07
25	5.0	0.09	0.11
35	2.5	0.04	0.05
35	5.0	0.06	0.09
45	2.5	0.03	0.07
45	5.0	0.06	0.12
55	2.5	0.20	0.62
55	5.0	0.49	0.94
65	2.5	0.42	1.51
65	5.0	0.58	1.26

[a] The true parameters used to simulate the data.
[b] The standard deviations of the best-fit parameters from fits to 500 replicate simulated data sets.

Table S6. Best-fit parameters for the T4L DEER signals in Fig. 8.

label positions	Δ	λ	scale	r_{0_1}	σ_{r_1}	r_{0_2}	σ_{r_2}	a_2 (f_2) [a]	r_{0_3}	σ_{r_3}	a_3 (f_3) [a]
62/109	0.2589	5.0640	1.0017	29.04	1.58	25.7	1.2	0.19 (0.06)	34.95	1.05	0.66 (0.13)
62/134	0.1343	4.8429	1.0026	40.74	0.88	44.48	0.51	0.125 (0.125)	n/a	n/a	n/a
109/134	0.1928	4.8569	1.0030	30.6	5.8	30.76	1.41	0.787 (0.787)	n/a	n/a	n/a

[a] The f values are calculated for the best-fit a values from Eqs 9-10.

Table S7. Modulation depth and background parameters for the calculated DEER signals reported in Fig. 10. These values were calculated by fitting the theoretical DEER signal (Eqs. 2-6) to the experimental data using a gradient-minimization approach.

label positions	Δ_{MD}	λ_{MD}	$\Delta_{EBMetaD}$	$\lambda_{EBMetaD}$
62/109	0.256	5.0716	0.257	5.0673
62/134	0.131	4.8529	0.133	4.8417

109/134	0.178	4.9230	0.190	4.8596

Table S8. Best-fits to theT4L DEER signals in Fig. 8 using alternate basis functions with non-Gaussian shapes.

label positions	n	Shape	Modes[a]	q	χ_ν^2	ΔBIC
62/109	2	SND	2	10	0.657	0.0
62/109	3	Gaussian	2	11	0.662	6.2
62/109	2	GSND	2	12	0.654	7.7
62/109	2	GND	2	10	0.691	11.8
62/134	1	SND	1	6	1.216	0.0
62/134	1	GSND	1	7	1.219	4.5
62/134	2	Gaussian	2	8	1.219	8.6
62/134	1	GND	1	6	1.343	15.9
109/134	1	GND	1	6	1.107	0.0
109/134	2	Gaussian	1	8	1.049	0.7
109/134	1	GSND	1	7	1.099	2.9
109/134	1	SND	1	6	1.387	29.1

[a] The number of distinct local maxima in $P(R)$.

SUPPLEMENTAL FIGURES

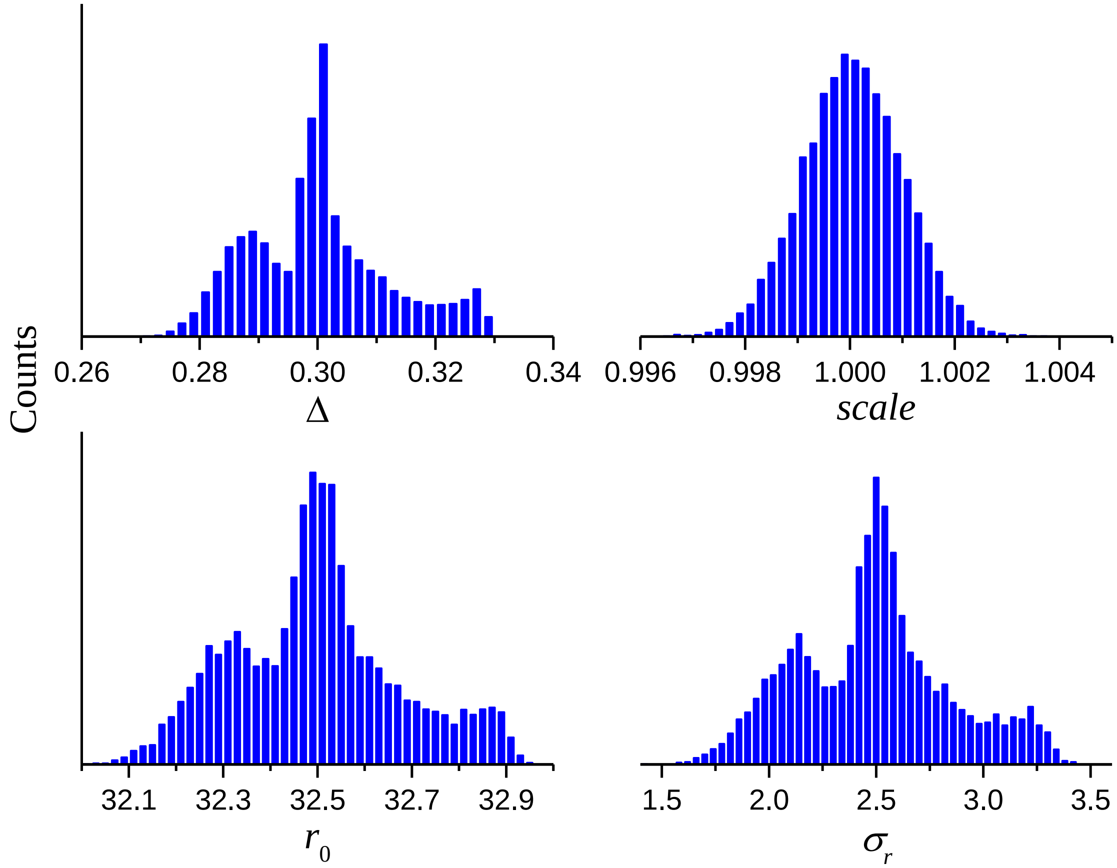

Figure S1. Histograms for each of the fit parameters from analysis of 10000 replicates of the simulated DEER signal in Fig. 4A. The corresponding histogram for the background parameter λ is shown in Fig. 4C.

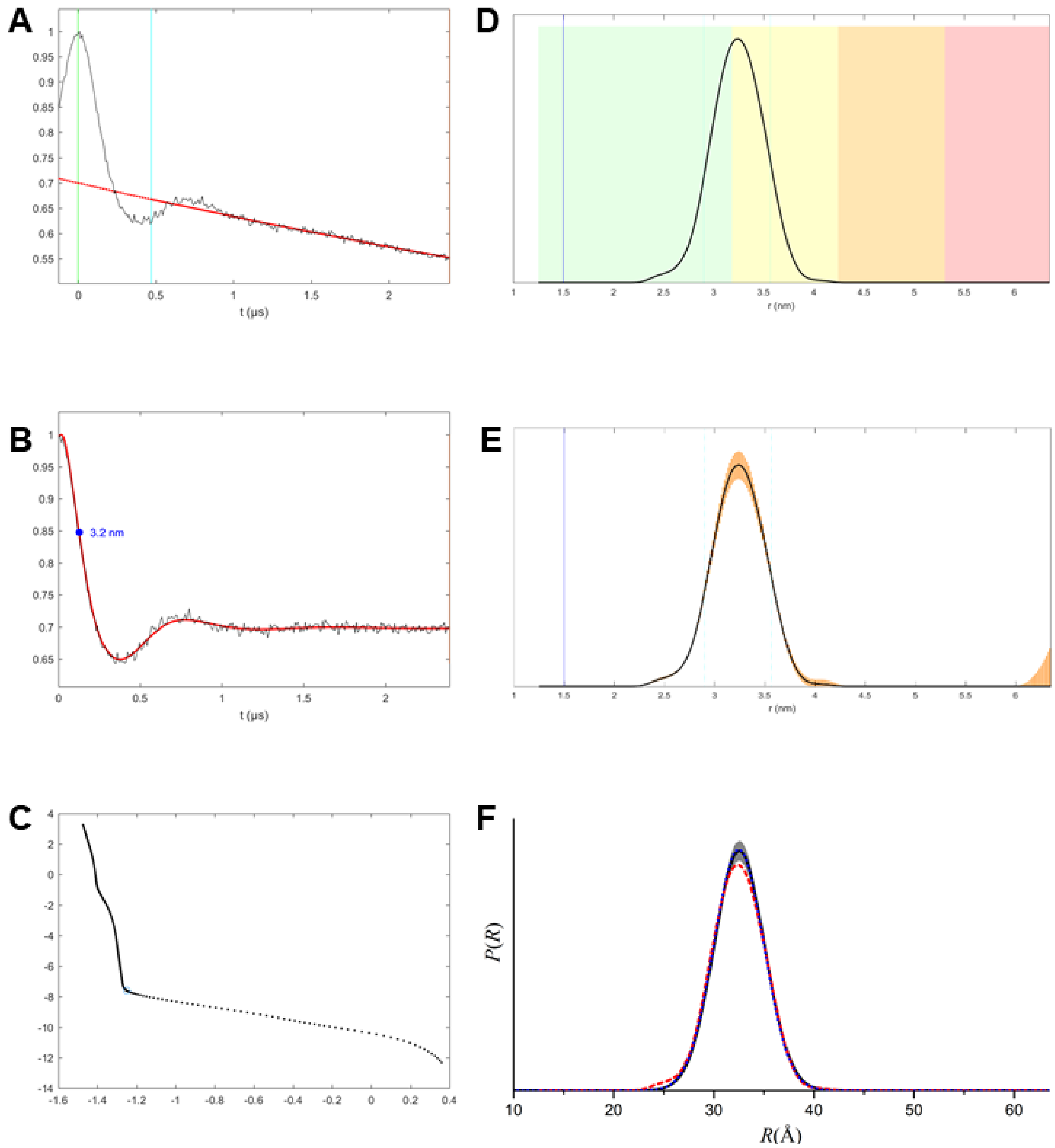

Figure S2. Fitting of the simulated signal in Fig. 1 (*noise* = 0.005) with DeerAnalysis2016. A) Scaled data (black line) with estimate background correction (red line). B) Background corrected data (black line) with initial fit (red line). C) L curve for determining regularization parameter. D) Initial distance distribution (black line) with color-coded reliability regions. E) Mean distance distribution (black line) and error bounds (orange lines) after using validation tool to vary the range of data used to estimate background correction. F) Comparison of mean distance distribution in E (dashed red line) to results from DD in main text (black line with shaded grey region indicating 2σ confidence band).

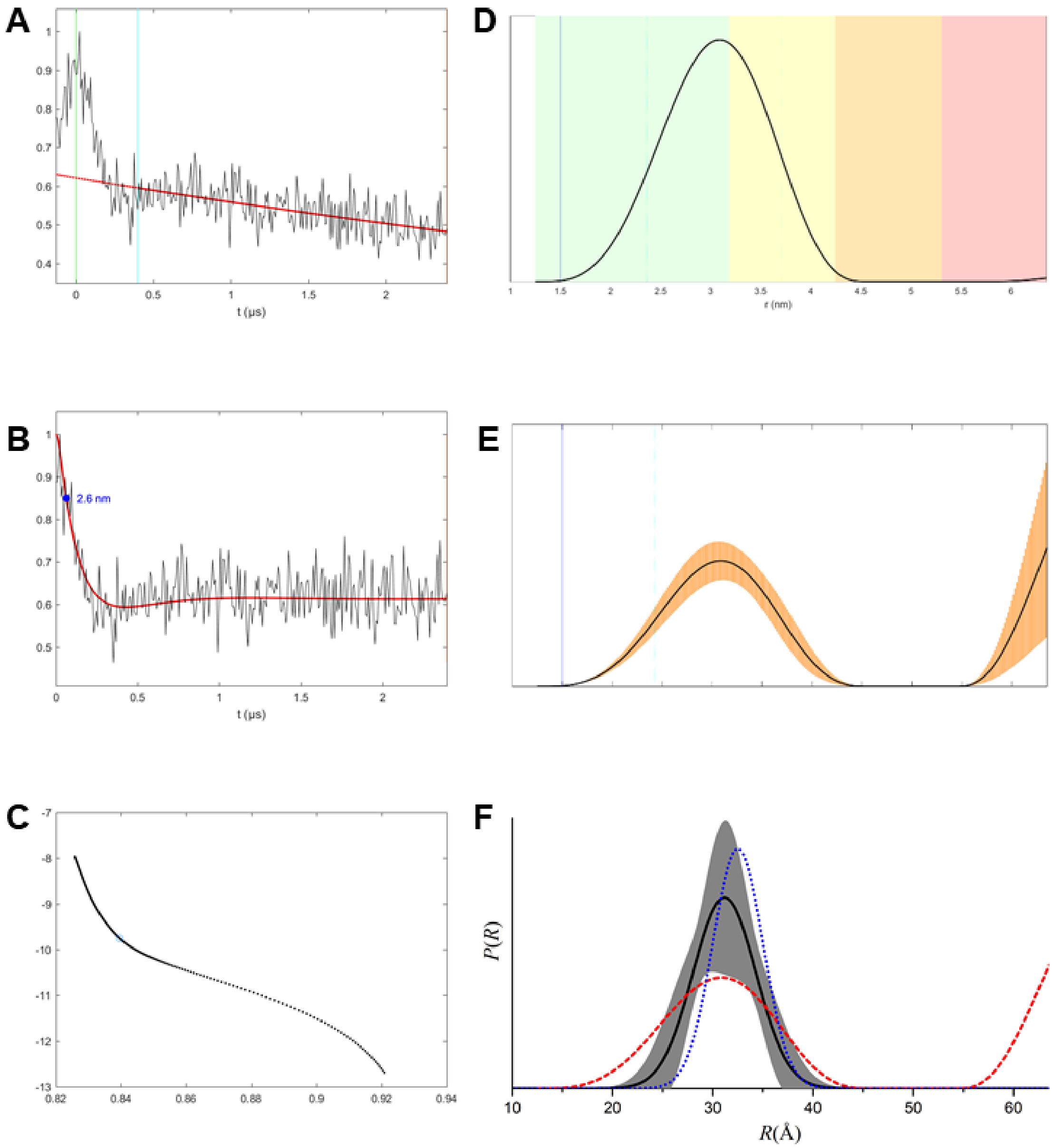

Figure S3. Fitting of the simulated signal in Fig. 1 (*noise* = 0.05) with DeerAnalysis2016. Panel descriptions as in Fig. S2.

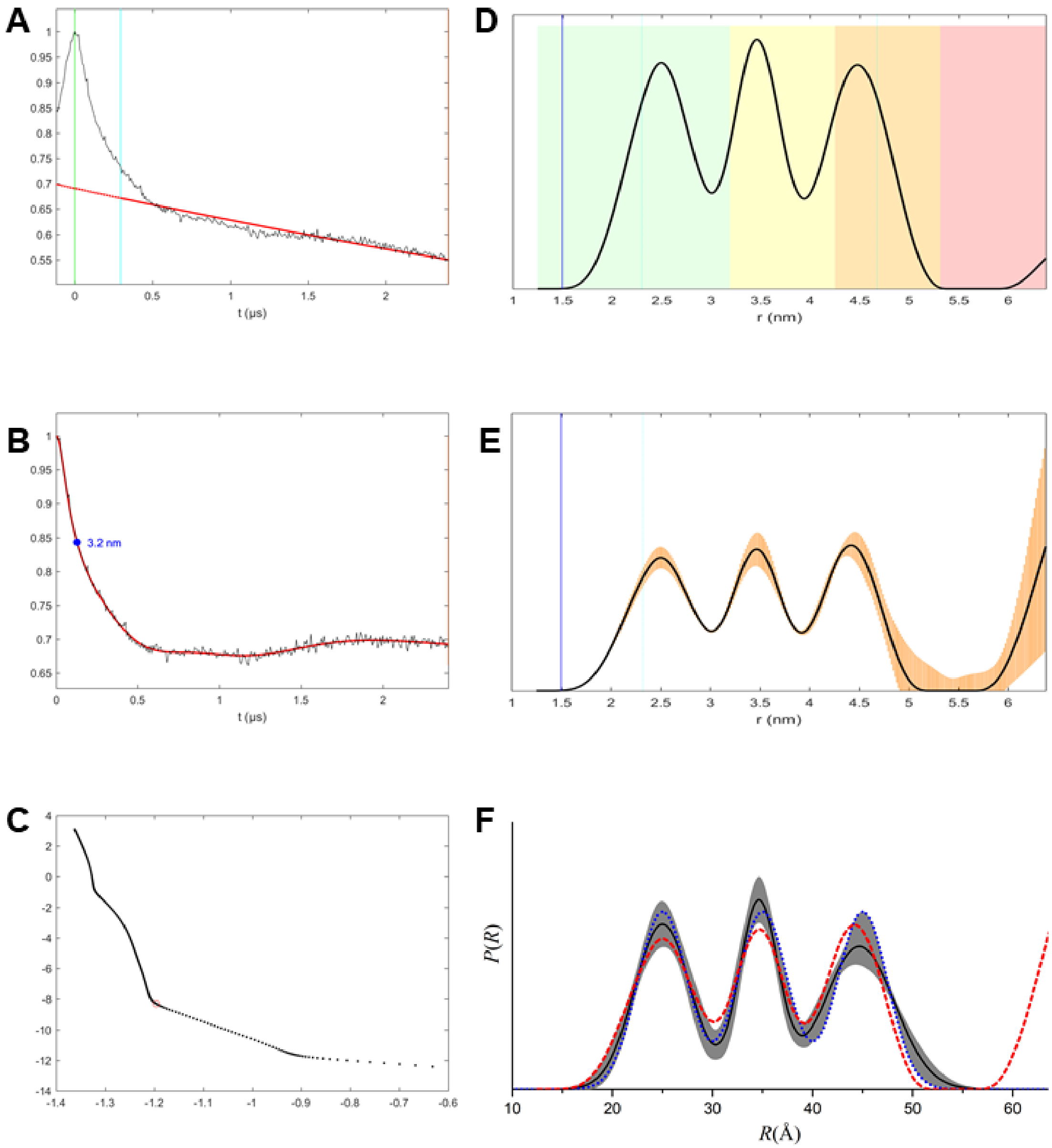

Figure S4. Fitting of the simulated multi-component signal in Fig. 5 (equal-width components) with DeerAnalysis2016. Panel descriptions as in Fig. S2.

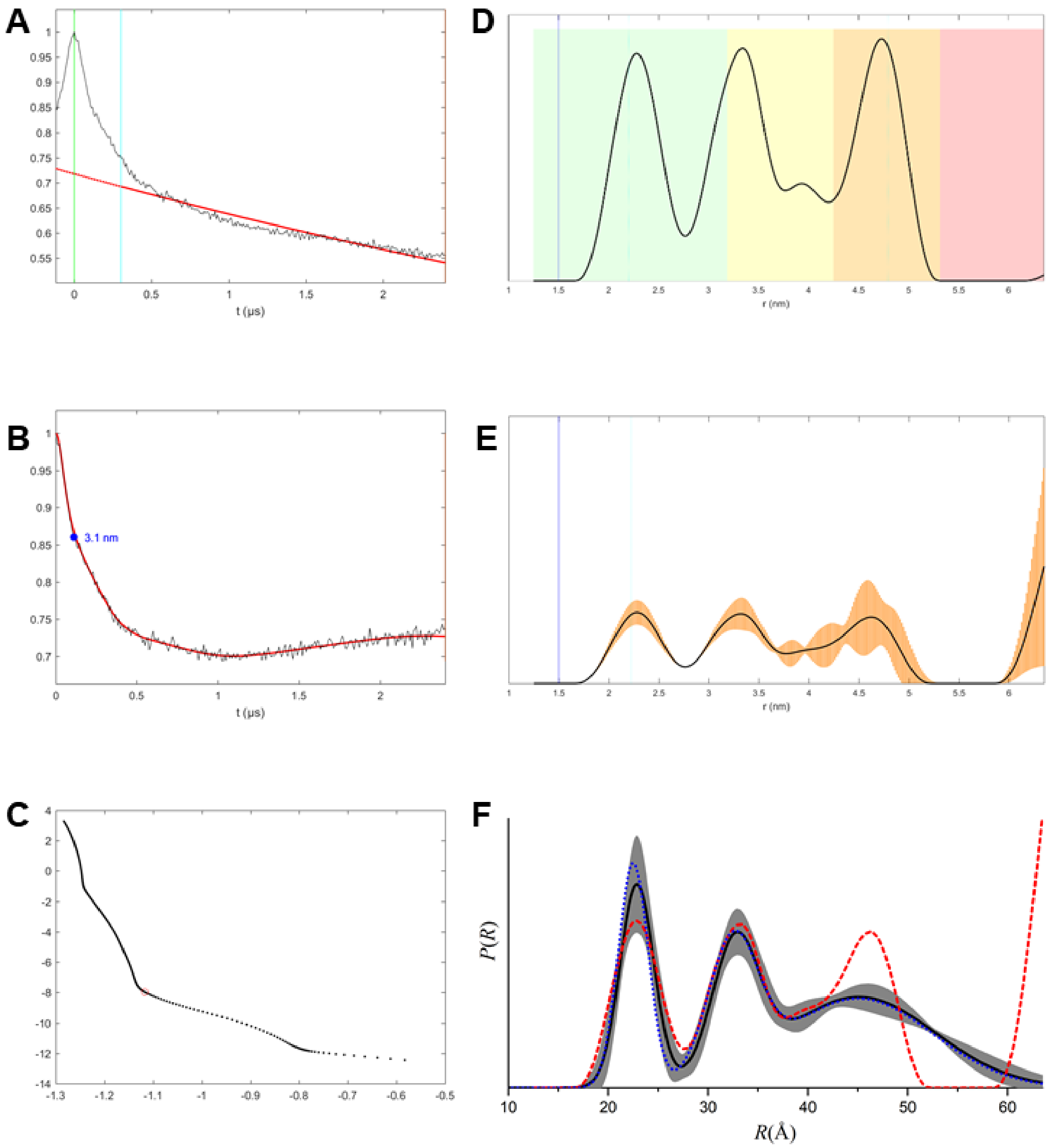

Figure S5. Fitting of the simulated multi-component signal in Fig. 5 (variable-width components) with DeerAnalysis2016. Panel descriptions as in Fig. S2.

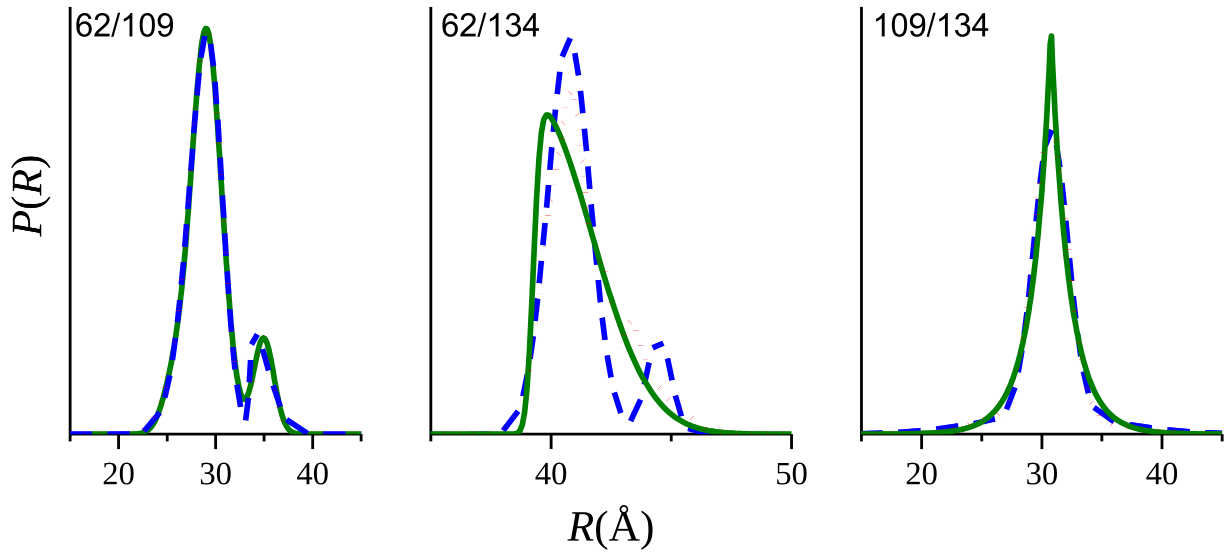

Figure S6. Best-fit distance distributions for theT4L DEER signals in Fig. 8 using alternate basis functions with non-Gaussian shapes. The $P(R)$ for the optimal non-Gaussian models ($\Delta BIC = 0$ in Table S8) are shown as solid green lines. For comparison, the optimal Gaussian models from Fig. 8 are shown as dashed blue lines and the $P(R)$ from the EBMetaD calculations including confidence bands from Fig. 9B are shown as dotted red lines.

SUPPLEMENTAL METHODS

Fitting can be performed with non-Gaussian basis functions in place of the Gaussian functions of Eq. 11. Here, we have considered the following alternative basis functions.

<u>Generalized Normal Distribution (**GND**) with non-zero excess kurtosis</u> (1)

$$p_j(R) = \frac{\beta_j}{2Q\sigma_{rj}\Gamma({}^{1}/_{\beta_j})} e^{-\left(\frac{|R-r_{0j}|}{Q\sigma_{rj}}\right)^{\beta_j}} \qquad \text{[S1]}$$

$$Q = \sqrt{\frac{\Gamma({}^{1}/_{\beta_j})}{\Gamma({}^{1}/_{\beta_j})}}$$

where Γ is the Gamma function

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$$

The shape of each component is controlled by the additional parameter β_j.

<u>Skew Normal Distribution (**SND**) with non-zero skew</u> (2)

$$p_j(R) = \frac{1}{\sqrt{2\pi\,\sigma_{rj}^2}} \mathrm{e}^{\frac{-\left(R-r_{0j}\right)^2}{2\sigma_{rj}^2}} \left[1 + \mathrm{erf}\left(\frac{Z}{\sqrt{2}}\right)\right] \qquad \text{[S2]}$$

$$Z = \zeta_j \frac{R - r_{0j}}{\sigma_{rj}}$$

where erf is the error function

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

The shape of each component is controlled by the additional parameter ζ_j.

<u>Generalized Skew Normal Distribution (**GSND**) with non-zero excess kurtosis and skew</u>

$$p_j(R) = \frac{\beta_j}{2Q\sigma_{rj}\Gamma(1/\beta_j)} e^{-\left(\frac{|R-r_{0j}|}{Q\sigma_{rj}}\right)^{\beta_j}} \left[1 + \mathrm{erf}(\frac{Z}{\sqrt{2}})\right] \qquad [S3]$$

The shape of each component is controlled by the additional parameters β_j and ζ_j.

SUPPLEMENTAL REFERENCES

1. Nadarajah, S. 2005. A Generalized Normal Distribution. J. Appl. Stat. 32:685-694.
2. Azzalini, A. 1985. A Class of Distributions which Includes the Normal Ones. Scand. J. Stat. 12:171-178.