Electronic Supporting Information for: Diffusive dynamics of contact formation in disordered polypeptides

Gül Zerze and Jeetain Mittal*
Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem PA 18034

Robert B. Best†
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda MD 20892
(Dated: December 30, 2015)

DETAILS OF MOLECULAR SIMULATIONS

In order to obtain initial conditions of serial simulations of n=1,2 peptides, we ran 50 ns/replica parallel tempering in well-tempered ensemble (PTWTE) [1, 2] simulations for both n=1 and n=2 peptide, using 1.2 kJ/mol and 500 kJ/mol as the Gaussian height and width, respectively. Eight replicas were used in a temperature range from 300 K to 503 K, using a bias factor of 8. We keep the 300 K replica neutral, i.e. no energy bias is applied, while allowing it to exchange with the adjacent replica (a similar application as in [3]). The last 24 frames of the 300 K replica were picked, which were saved in every 2 ns, to serve as the initial conditions of serial runs.

	ff03*		ff03w		ff03ws	
Peptide	runs	time/μs	runs	time/μs	runs	time/μs
AGQ_1	24	2.4	24	2.4	24	2.4
AGQ_2	24	4.8	24	4.8	24	4.8
AGQ_3	8	7.9	8	7.6	8	7.6
AGQ_4	8	7.8	8	7.5	8	7.4
AGQ_5	8	7.7	8	7.5	8	7.5
AGQ_6	5	4.4	5	4.2	10	9.6

TABLE I. Number of runs and total length of runs for each force field and peptide

Initial conditions for peptides with $n = 3 - 6$ were obtained by running a 200 ps constant pressure simulation at 293 K to equilibrate the box size, followed by a 20 ns constant volume simulation at 500 K. The last frame of the 500 K run was used to initiate long runs at constant pressure for computing the quenching rates. Simulations were run in one of three force field combinations: (i) Amber ff03*[4] with TIP3P water[5], (ii) Amber ff03w[6] with TIP4P/2005 water [7], and (iii) Amber ff0ws[8], also with TIP4P/2005 water. Molecular dynamics simulations were run at constant pressure of 1 bar using a Parinello-Rahman barostat with a coupling time of 5 ps and a compressibility of 4.5×10^{-5} bar^{-1}, and constant temperature of 293 K using a velocity-rescaling thermostat [9] with coupling time of 1 ps. This thermostat was chosen because it should have a minimal effect on the dynamics (compared to e.g. Langevin dynamics at high friction). Lennard-Jones interactions were evaluated every step for atoms separated by up to 0.9 nm and every 10 steps for atoms separated by between 0.9 and 1.4 nm. Mean-field corrections to energy and pressure were included for Lennard-Jones interactions beyond 1.4 nm. Long-range electrostatics were included via the particle-mesh Ewald scheme [10] with a grid spacing of ~ 0.12 nm and a real-space cut-off of 0.9 nm. A summary of the number of runs and cumulative simulation time for each peptide and force field are listed in Table I.

DISTRIBUTION OF CONTACT CONFIGURATIONS

To characterize the modes of association of Trp and Cys, we determine the distribution of the position of the Cys sulfur (SG) in a frame of reference defined by the Trp. This frame has the origin at the Trp CG carbon and the x-axis running through the CZ3 carbon. The z-axis is parallel to the vector $x \times t$, where t is the vector from the Trp CG to NE1, and $y = z \times x$ (Fig. 1 A). The corresponding distributions of spherical polar coordinates suggest there is no strongly preferred contact orientation (Fig. 1 B), and that Trp and Cys interact in similar orientations in all peptides.

FITTING OF DIFFUSION MODELS USING BAYESIAN METHOD

Fitting of the 1D diffusion model followed a similar method to previous work [11–13]. The all-atom simulation trajectories were projected onto the Trp-Cys distance r_{cw}, which was then discretized into regularly spaced bins spanning the populated range of that coordinate. The number of bins used was 30 for $n = 1 - 3$ and 60 for $n = 4 - 6$. The increased number of bins is required in order to described accurately the shorter distances where contact formation occurs for the longer peptides. Count matrices of the number of observations N_{ij} from bin j to bin i after a given lag time Δt were constructed. Best-fit diffusion coefficients and free energies were fitted to these data using the previously described

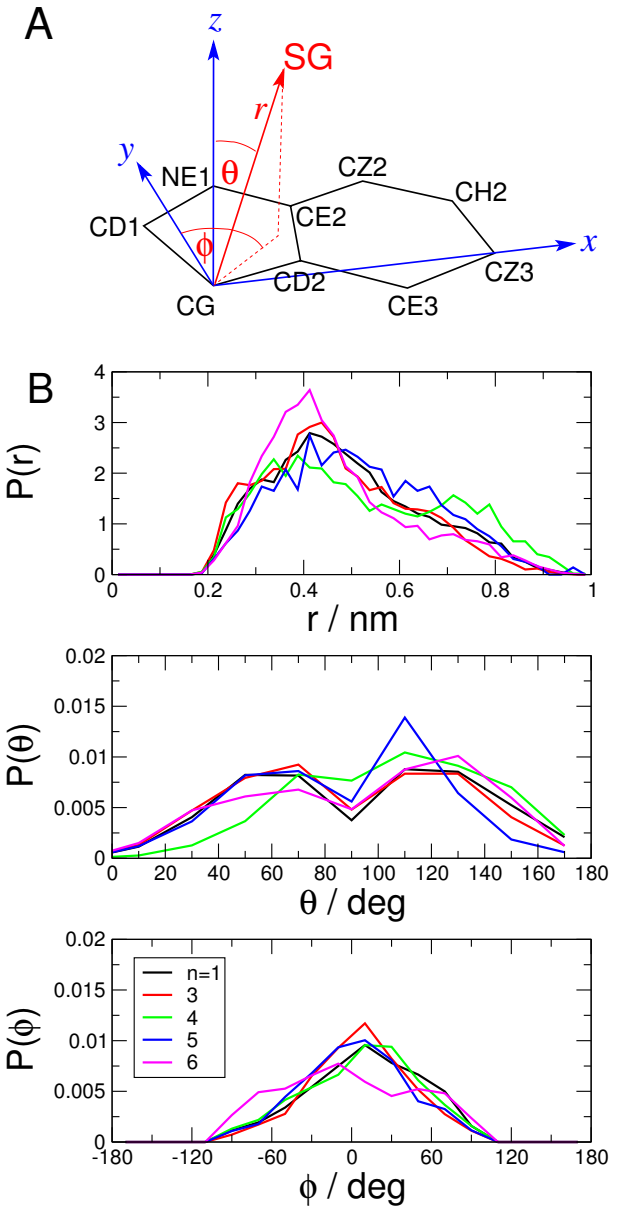

FIG. 1. Distribution of Trp-Cys contact configurations. (A) The position of the Cys SG is defined in a frame of reference relative to the Trp. (B) Distributions of spherical polar coordinates for SG in this axis frame.

Bayesian procedure. A lag time of $\Delta t = 0.1$ ns was used for $n = 1 - 2$ and a lag time $\Delta t = 0.5$ ns for $n = 3 - 6$, in order that the resulting dynamics was approximately Markovian. In Figure 2 we show the convergence of the slowest relaxation time of the model with lag time. A smoothing prior on the similarity of diffusion coefficients D_i, D_{i+1} for adjacent bins i, $i+1$ of the form

$$P(D_i, D_{i+1}) = \exp\Big[-\frac{(D_i - D_{i+1})^2}{2\gamma^2(\min(D_i, D_{i+1}))^2}\Big] \qquad (1)$$

was used. A stiffness coefficient of $\gamma = 0.1$ was chosen.

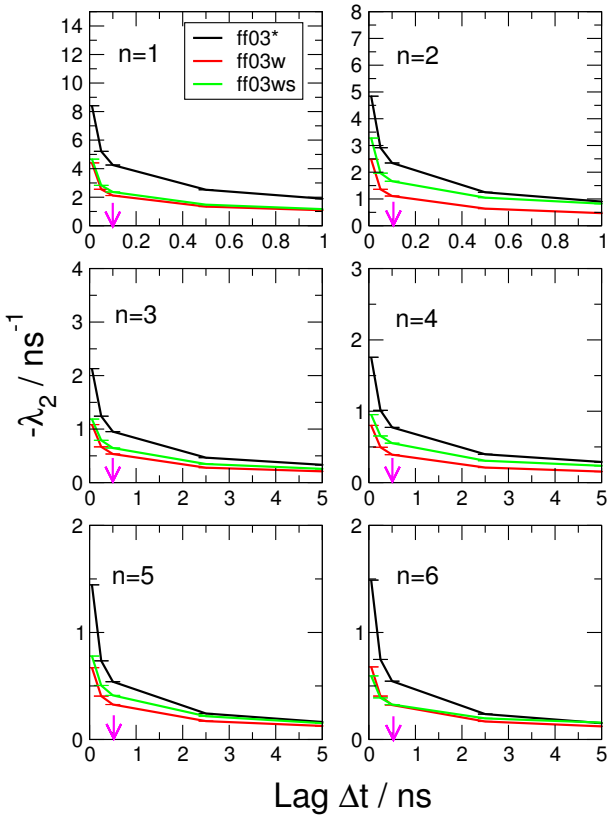

FIG. 2. Convergence of relaxation times with lag time Δt for the diffusion model.

CALCULATION OF RATES FROM DIFFUSION MODEL USING SSS THEORY

Given the one-dimensional $D(r_{\rm cw})$ and equilibrium distribution $p_{\rm eq}(r_{\rm cw})$, the Szabo-Schulten-Schulten theory[14] allows the rate of diffusion-limited contact formation k_{D+} to be calculated as [15] (for instantaneous quenching at r_c):

$$k_{D+}^{-1} = \int_{r_c}^{r_{\rm max}} \frac{\big[\int_r^{r_{\rm max}} p_{\rm eq}(s)ds\big]^2}{D(r)p_{\rm eq}(r)}dr \qquad (2)$$

For a distance-dependent reaction rate $q(r_{\rm cw})$, the diffusion-limited rate can be estimated via:

$$k_{D+}^{-1} = k_R^{-2}\int_0^{r_{\rm max}} \frac{\big[\int_r^{r_{\rm max}} (q(s) - k_R)p_{\rm eq}(s)ds\big]^2}{D(r)p_{\rm eq}(r))}dr \qquad (3)$$

where k_R is the reaction-limited rate:

$$k_R = \int_0^{r_{\rm max}} p_{\rm eq}(r)q(r)dr \qquad (4)$$

The above integrals were evaluated over the discretized bins of the diffusion model, counting for the first bin only the range between r_c and the outer edge of the bin.

Lastly, we note that experimental data are usually fit, for simplicity, to a model with a constant, position-independent, diffusion coefficient $D_{\rm const}$, i.e. $D(r) \equiv$

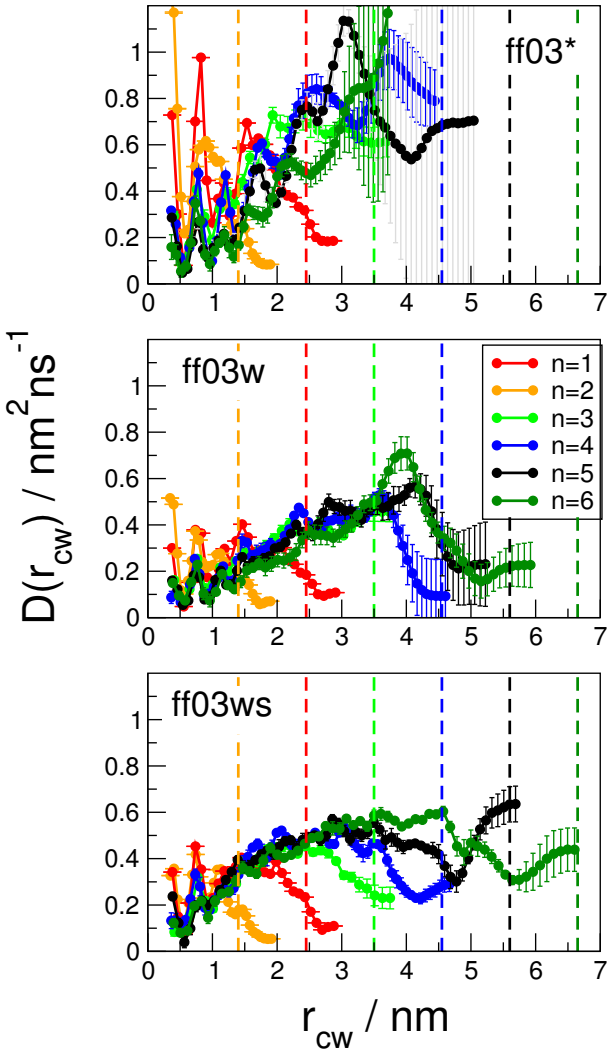

FIG. 3. Position-dependent diffusion coefficients from Bayesian fit. Vertical broken lines give the approximate maximum extension for each length n_b, given by 0.35 n_b nm.

D_{const}. In order to facilitate comparison with experiment, we can determine the effective D_{const} that would result in the same diffusion-limited quenching rate as we obtain with our distance-dependent $D(r)$, by defining D_{const} via:

$$\frac{1}{D_{\text{const}}}\int_0^{r_{\max}} \frac{\left[\int_r^{r_{\max}}(q(s)-k_R)p_{\text{eq}}(s)ds\right]^2}{p_{\text{eq}}(r))}dr \quad (5)$$
$$= \int_0^{r_{\max}} \frac{\left[\int_r^{r_{\max}}(q(s)-k_R)p_{\text{eq}}(s)ds\right]^2}{D(r)p_{\text{eq}}(r))}dr$$

Essentially, we apply the experimental analysis to our simulated data: we treat our rates calculated from the full distance-dependent diffusion model as experimental data, and use our simulated $p_{\text{eq}}(r)$ as the Trp-Cys distribution, allowing an effective constant diffusion coefficient to be obtained.

FITTING OF DISTANCE-DEPENDENT QUENCHING RATES TO EXPERIMENTAL DATA

Our initial model for the distance-dependent quenching rate was a step function with parameters taken from the literature [16]. Since this is a fairly strong assumption about the distance dependence, we also tried a more realistic alternative. Assuming that the quenching mechanism can be described as electron transfer, we modelled the distance-dependent rate $q(r_{\text{cw}})$ with an exponential function:

$$q(r_{\text{cw}}) = k_0 \exp[\beta(r_{\text{cw}} - r_c)] \quad (6)$$

We chose a contact distance of $r_c = 0.4$ nm as before; the choice is quite arbitrary as an equivalent expression could be obtained with a different choice, by appropriately scaling k_0. We used the data reported by Lapidus *et al.* [17] to determine the optimal parameters. The main difference from their work is that we use a pair distance distribution function determined from molecular simulation rather than a uniform distribution. We ran a 1 μs of a single N-acetyl tryptophanamide (NATA) and zwitterionic cysteine in water. We did not attempt to include trehalose due to the requirement to have a suitable force field for it, as well as the difficulty of sampling such a system; we instead assume that the distribution in water is sufficiently similar. The probability density is shown in Fig. 4. Note that we have not attempted to normalize by the available volume at each distance as for a regular pair distribution function, because that operation is not easily defined for the minimum Trp-Cys distance we use as a coordinate. Instead, we use observed density of distances $p(r_{\text{cw}})$ directly; the drop-off at larger r_{cw} due to the finite system size is unimportant since the quenching is so short ranged.

We fit the absorbance decays from Lapidus *et al.* [17] to the function

$$A(t) = A_0 L(t) \exp\Big[-\frac{Q}{Q_0}\int_0^{r_{\max}} p_{\text{eq}}(r)[1-S(t|r)]dr\Big] \quad (7)$$

where $L(t)$ is the experimentally determined triplet lifetime in the absence of Cys, A_0 is a scaling coefficient for fitting, Q and Q_0 are the concentration in the experiment and the NATA-Cys simulation respectively, and $S(t|r) = \exp[-q(r)t]$.

We systematically varied the two free parameters β and k_0 on a grid, obtaining an overall χ^2 surface shown in Fig. 5. While there is clearly a lot of correlation between the parameters, we picked the values at the center of the optimal region as optimal values. Using these values ($k_0 = 1.0 \times 10^8$ s^{-1}, $\beta = 33.333$ nm^{-1}), we indeed find a very good fit to the experimental data, as shown directly in Fig. 6. There is some discrepancy at long times for the lower concentration, which has been explained in terms

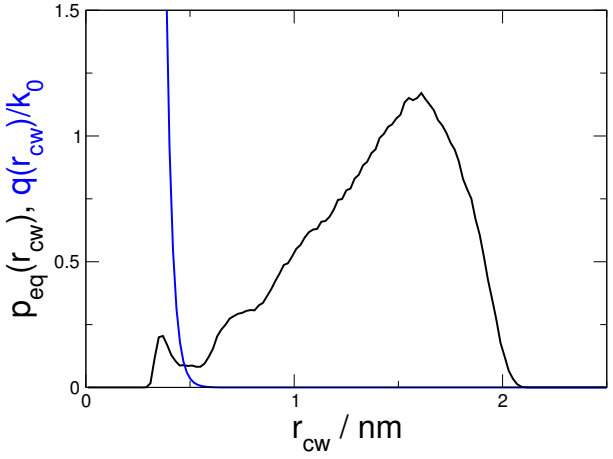

FIG. 4. Distribution function $p(r_{\rm cw})$ for the minimum distance between N-acetyltryptophanamide (NATA) and zwitterionic cysteine from simulation in explicit water (black) and distance-dependence of fitted quenching rate (blue).

of very slow diffusion within the glass [17], however this effect is less important at the higher concentration.

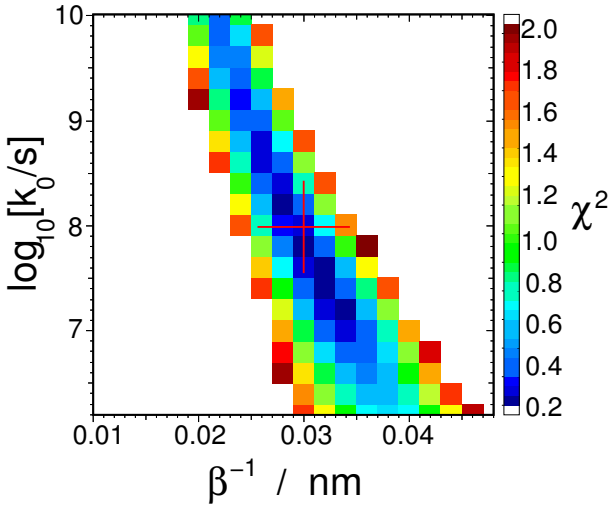

FIG. 5. χ^2 surface for fit to Trp-Cys quenching data in a room temperature trehalose glass. Red cross indicates chosen "optimal" parameters.

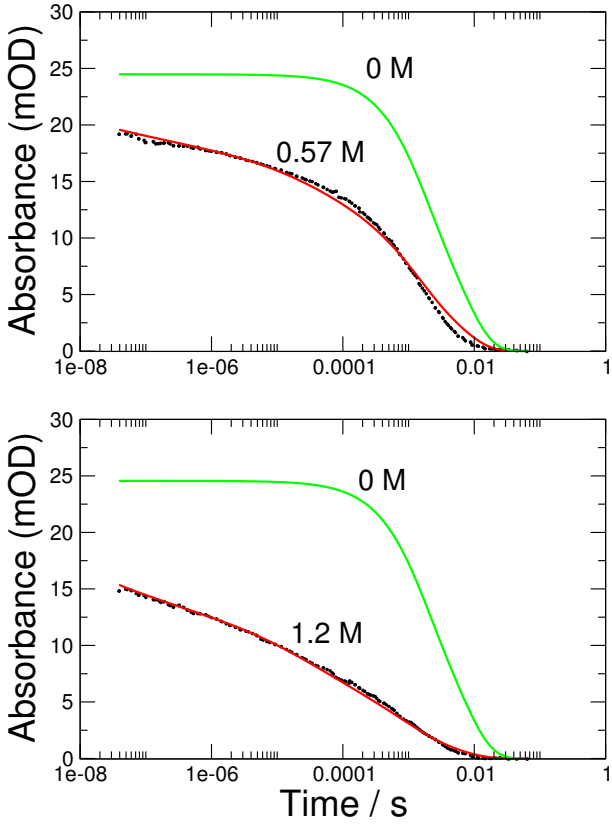

FIG. 6. Fit to raw data for Trp-Cys quenching from Lapidus *et al.*[17]. Green curve shows the survival $L(t)$ of the Tryptophan triplet in the absence of Cysteine, black curve is the experimental data in the presence of Cysteine concentration shown, red curve is the fit with the optimal parameters for the distance dependence of the quenching rate.

* jeetain@lehigh.edu
† robertbe@helix.nih.gov

[1] M. Bonomi and M. Parrinello, Phys. Rev. Lett. **104**, 190601 (2010).
[2] M. Deighan, M. Bonomi, and J. Pfaendtner, J. Chem. Theory Comput. **8**, 2189 (2012).
[3] L. Sutto and F. L. Gervasio, Proc. Natl. Acad. Sci. U. S. A. **110**, 10616 (2013).
[4] R. B. Best and G. Hummer, J. Phys. Chem. B **113**, 9004 (2009).
[5] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. **79**, 926 (1983).
[6] R. B. Best and J. Mittal, J. Phys. Chem. B **114**, 14916 (2010).
[7] J. L. F. Abascal and C. Vega, J. Chem. Phys. **123**, 234505 (2005).
[8] R. B. Best, W. Zheng, and J. Mittal, J. Chem. Theor. Comput. **10**, 5113 (2014).
[9] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[10] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[11] G. Hummer, New J. Phys. **7**, 34 (2005).
[12] R. B. Best and G. Hummer, Phys. Rev. Lett. **96**, 228104 (2006).
[13] R. B. Best and G. Hummer, Phys. Chem. Chem. Phys. **13**, 16902 (2011).
[14] A. Szabo, K. Schulten, and Z. Schulten, J. Chem. Phys. **72**, 4350 (1980).
[15] L. J. Lapidus, P. J. Steinbach, W. A. Eaton, A. Szabo, and J. Hofrichter, J. Phys. Chem. B **106**, 11628 (2002).
[16] I.-C. Yeh and G. Hummer, J. Am. Chem. Soc. **124**, 6563 (2002).
[17] L. J. Lapidus, W. A. Eaton, and J. Hofrichter, Phys. Rev. Lett. **87**, 258101 (2001).