Supplementary Materials: Simulation of Biochemical Reactions with ANN-Dependent Kinetic Parameter Extraction Method

Fei Tan and Jin Xu

S1. Derivation of $a(\mathrm{r}_0, \sigma)$

The expression of $a(\mathrm{r}_0, \sigma)$ is as follows:

$$a(\mathrm{r}_0, \sigma) = \int_0^{r_0} \int_0^{t_0} \frac{\sqrt{2}}{\sqrt{\pi t}\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dt\, dr \approx \frac{\sqrt{8\pi}}{2\sigma}\left(\mathrm{erf}\left(\frac{r_0}{\sqrt{2}\sigma}\right) - \mathrm{erf}(0)\right) \quad \text{(S1)}$$

with $t_0 = 1$, r_0 and σ representing separation distance between pairwise diffusion species and diffusion coefficient respectively

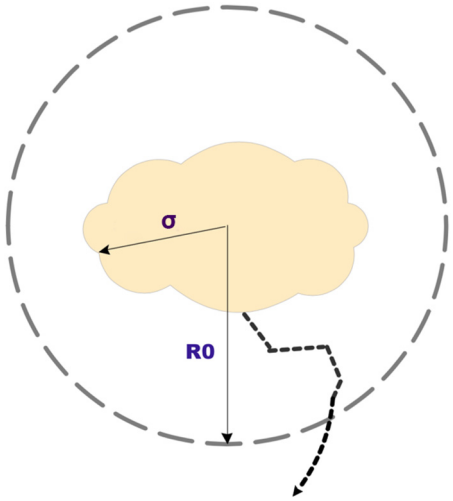

Figure S1. Schematic representation of the stochastic association model of monomeric M^{pro}-C. Each monomeric M^{pro}-C protein is represented in an irregular shape in whitish yellow, it has a protein radius on the same scale to the diffusion constant sigma. r_0 is the radius of reaction for the active monomeric proteins. reaction occurs when two activated monomeric species come within a distance of r_0 with respect to each other.

$a(\mathrm{r}_0, \sigma)$ is derived in an idealized model system of bimolecular protein-protein association, the scheme of association is depicted in Figure S1. For freely diffusion proteins undergoing Brownian dynamic motion in three-dimensional solution environment, each protein is treated as a sphere of known radius $R = 10$Å, as measured from monomeric M^{pro}-C [1]. Diffusing species begun at a random orientation in solution, collisions may occur when diffusing protein encounter each other, in other words, when their interparticle distance are within a separation of r_0. If at any time, two protein particles have come within each other with a distance smaller than r_0, then we assume an interaction had taken place. Once a protein moving under Brownian motion has visited its reaction counterparts, it remains trapped in the vicinity of each other, undergoing multiple collisions and substantial rotational reorientation before reaction. therefore, it is justified to simply consider the first time the two protein particles come across each other[2]. Isotropic diffusion coefficient

σ were given by the Stokes-Einstein relations $D = k_B T/6\pi\eta R$, a prefactor can be introduced to accounts for the orientational alignment that marks the probability of a valid interaction (Although we did not use one in our case)[3]. First, we derive the probability of interaction [4]. Since the trajectory of protein diffusion is completely random, we used a random Gaussian process to denote the positions of the protein particles.

$$B_i = \left[\sigma W_{ix}(t), \sigma W_{iy}(t), \sigma W_{iz}(t)\right] \text{ (S2)}$$

Which represents three-dimensional random Brownian motion of protein particle i, in which $W_i(t)$ is one dimensional standard Brownian motion represented by a Gaussian distribution with zero mean and a variance equal to σ^2. Calculation of association rate constant in three-dimensional Cartesian space poses great difficulty, thus in order to reduce the three-dimensional system to one dimensional, we turn to deal with the difference of the Brownian motions. At any given time, t, the difference of two Gaussian processes, say $W_i(t)$ and $W_j(t)$ is also a Gaussian process, with $X_{ij}(t)$ denoting this process, we shall have:

$$X_{ij}(t) = B_i - B_j = \sqrt{{\sigma_i}^2 + {\sigma_j}^2}W(t) \quad \text{(S3)}$$

Hence, we have arrived at a simpler expression that is also a random Gaussian process.

In stochastic differential equations, by the Ito formula, the derivative of process $dX_{ij}(t) = \frac{dt}{X_{ij}(t)} + dW(t)$

In which W(t) is a standard Gaussian process. It is clear that $X_{ij}(t)$ is a Bessel process. The transition probability $p(r, s, t)$ of the Bessel process $X_{ij}(t)$ is as follows:

$$p(r, s, t) = \frac{r}{s\sigma^2}\sqrt{\frac{2}{\pi t}}e^{-\frac{r^2+s^2}{2t\sigma^2}}sinh\frac{rs}{t\sigma^2} \text{ (S4)}$$

The transition probability $p(r, s, t)$ refers to how likely protein particles comes from an interparticle distance of s to a separation of r within one another. Disregarding the degradation of M^{pro}-C since no significant degradation within experimental timescale is observed, the expected rate of interaction can be given by equation (5), the graph of $a(r_0, \sigma)$ is plotted in Figure S2. as a function of r_0 and σ in three-dimensional space.

$$a(r_0, \sigma) = \int_0^{r_0}\int_0^{\infty}\int_{t=0}^{t_0} p(r, s, t)\, dsdrdt$$

$$= \int_0^{r_0}\int_0^{\infty}\int_{t=0}^{t}\frac{r}{s\sigma^2}\sqrt{\frac{2}{\pi t}}e^{-\frac{r^2+s^2}{2t\sigma^2}}sinh\frac{rs}{t\sigma^2}dsdrdt = \frac{\sqrt{8t_0\pi}}{\sigma^2}\left(\text{erf}\left(\frac{r_0}{\sqrt{2t_0}\sigma}\right) - \text{erf}(0)\right)$$

(S5)

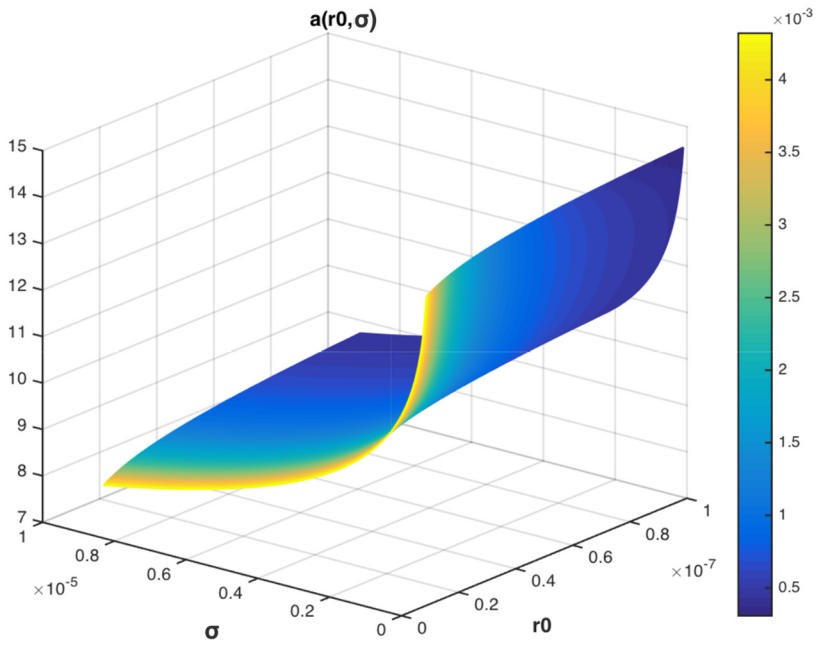

Figure S2. The 3D graph of a(r_0, sigma(σ)) as a function of r_0(reaction radius) and sigma(σ) (diffusion coefficient) (plotted by Matlab).

S2. Two-State Model of the Interconversion Process

In application of AKPE to the models of monomeric M^{pro}-C association. We obtain monomeric M^{pro}-C and dimeric M^{pro}-C concentration as a function of time in different experimental conditions, the two-state interconversion curve can be derived as follows:

$$2M \underset{k_a}{\overset{k_d}{\leftrightarrow}} D \quad \text{(S6)}$$

$$x = r_1 \frac{1 - \exp(-r_2 t)}{1 + {r_1}^2 \exp(-r_2 t)} \quad \text{(S7)}$$

$$k_d = r_2 \frac{1 - r_1}{1 + r_1} \quad \text{(S8)}$$

$$k_a = \frac{r_2 r_1}{2c_0(1 - {r_1}^2)} \quad \text{(S9)}$$

k_a and k_d are forward and reverse kinetic rate constants and x is the concentration of dimeric M^{pro}-C, r_1, r_2 are propensity parameters of reactions.

S3. List of Mutants and the Method of its Generation

We Analyze the interconversion kinetics of truncation mutant M^{pro}-C^{295}, M^{pro}-C^{296}, M^{pro}-C^{298} and M^{pro}-C^{301} which have deleted C terminal 11 residues, 10 residues, 8 residues, 5 residues respectively, and has decreasing overall association rates. We also took point mutations M^{pro}-C^{P107A}(with residue 107 Proline mutated to alanine), M^{pro}-C^{F108A}(with residue 108 Phenylalanine mutated to alanine), M^{pro}-C^{D109A}(with residue 109 Aspartic Acid mutated to alanine), M^{pro}-C^{V110A}(with residue 110 Valine mutated to alanine), M^{pro}-C^{R112A}(with residue 112 Arginine mutated to alanine), M^{pro}-C^{C114A}(with residue 114 Cysteine mutated to alanine), double mutation M^{pro}-$C^{F108AD109A}$ (with residue 108 Phenylalanine and Aspartic Acid 109 both

mutated to alanine), into consideration, for instance, mutating residues that are structurally crucial to the interconversion process to alanine, reduces the effects long sidechains might place on protein–protein interaction. Applying our proposed method AKPE to these mutants might be able to give us a more specific grasp on which reaction process this change in residue composition might affect the most. The experimental data used in our methods are acquired under the same experimental condition.

S4. Full Table of Reaction Coefficients Produced by AKPE

Due to the limited space in the main text, the full Table 2. is presented here, this table includes the predicted value of every reaction coefficient and its errors in a total of 10 repetitive runs of AKPE for M^{pro}-C-WT and its mutants at different conditions.

Table S1. The resulting kinetic parameters for M^{pro}-C proteins at different temperature as an outcome of AKPE.

Parameter / Mutant	Sequence	k_a/RMSD		k_b/RMSD		k_c/RMSD		k_d/RMSD		r_0/RMSD		σ/RMSD	
M^{pro}-C-WT at 303K	187-306	1.68×10^{9}	9.66×10^{8}	96.6	4.95	3.65×10^{-5}	4.09×10^{-6}	1.20×10^{-7}	4.55×10^{-8}	1.64×10^{-8}	1.06×10^{-8}	9.92×10^{-6}	2.35×10^{-7}
M^{pro}-C-WT at 306K	187-306	8.50×10^{9}	2.23×10^{9}	76.5	31.3	1.34×10^{-4}	1.47×10^{-5}	2.01×10^{-7}	1.33×10^{-7}	5.04×10^{-8}	3.35×10^{-8}	8.40×10^{-6}	1.30×10^{-6}
M^{pro}-C-WT at 308K	187-306	3.89×10^{10}	2.68×10^{10}	62.5	37	4.22×10^{-4}	5.71×10^{-5}	1.60×10^{-5}	1.40×10^{-5}	6.21×10^{-8}	2.88×10^{-8}	5.87×10^{-6}	2.85×10^{-6}
M^{pro}-C-WT at 310K	187-306	1.48×10^{11}	1.32×10^{11}	56.8	34.4	1.22×10^{-3}	8.95×10^{-5}	4.71×10^{-4}	3.94×10^{-4}	6.42×10^{-8}	3.31×10^{-8}	4.29×10^{-6}	2.52×10^{-6}
M^{pro}-C^{301} at 310K	187-301	2.13×10^{11}	1.79×10^{11}	57	27.8	2.66×10^{-3}	1.86×10^{-4}	5.56×10^{-4}	2.54×10^{-4}	4.59×10^{-8}	3.06×10^{-8}	3.15×10^{-6}	1.62×10^{-6}
M^{pro}-C^{298} at 310K	187-298	3.37×10^{11}	2.23×10^{11}	63.6	24.9	4.75×10^{-3}	2.96×10^{-4}	5.85×10^{-4}	2.92×10^{-4}	6.57×10^{-8}	2.37×10^{-8}	2.66×10^{-6}	1.08×10^{-6}
M^{pro}-C^{296} at 310K	187-296	5.59×10^{11}	1.80×10^{11}	55.2	32.3	1.16×10^{-2}	6.15×10^{-4}	4.89×10^{-4}	2.18×10^{-4}	4.99×10^{-8}	3.31×10^{-8}	2.29×10^{-6}	8.93×10^{-7}
M^{pro}-C^{295} at 310K	187-295	7.13×10^{11}	4.66×10^{11}	50.8	20.8	6.01×10^{-2}	6.27×10^{-4}	5.06×10^{-4}	2.35×10^{-4}	5.91×10^{-8}	2.13×10^{-8}	2.20×10^{-6}	8.82×10^{-7}
M^{pro}-C^{R112A} at 310K	R112A	1.63×10^{10}	2.54×10^{9}	75.3	25.9	1.21×10^{-3}	2.21×10^{-7}	1.62×10^{-6}	8.94×10^{-7}	1.35×10^{-8}	9.66×10^{-9}	4.36×10^{-6}	1.46×10^{-6}
M^{pro}-C^{P107A} at 310K	P107A	1.74×10^{11}	1.35×10^{11}	53.6	27.9	9.04×10^{-4}	3.14×10^{-4}	3.38×10^{-4}	2.89×10^{-4}	5.52×10^{-8}	2.33×10^{-8}	3.16×10^{-6}	1.44×10^{-6}
M^{pro}-C^{V110A} at 310K	V110A	1.81×10^{11}	1.63×10^{11}	63.5	29.7	5.01×10^{-3}	4.61×10^{-4}	3.64×10^{-4}	2.83×10^{-4}	5.35×10^{-8}	3.44×10^{-8}	3.09×10^{-6}	1.44×10^{-6}
M^{pro}-C^{D109A} at 310K	D109A	5.15×10^{10}	6.78×10^{09}	50.9	30.1	5.69×10^{-4}	5.10×10^{-5}	5.03×10^{-4}	2.29×10^{-4}	4.82×10^{-8}	3.22×10^{-8}	4.54×10^{-6}	1.38×10^{-6}
M^{pro}-C^{F108A} at 310K	F108A	9.84×10^{9}	2.76×10^{9}	64	34.7	2.69×10^{-4}	2.96×10^{-5}	1.33×10^{-6}	6.75×10^{-7}	4.90×10^{-8}	4.08×10^{-8}	7.92×10^{-6}	1.33×10^{-6}
M^{pro}-C^{C114A} at 310K	C114A	8.92×10^{10}	3.72×10^{9}	63.7	35.4	2.36×10^{-3}	2.47×10^{-4}	5.29×10^{-4}	3.18×10^{-4}	7.65×10^{-8}	2.58×10^{-8}	4.41×10^{-6}	1.50×10^{-6}
M^{pro}-$C^{F108AD109A}$ at 310K	F108AD109A	7.14×10^{10}	3.40×10^{10}	63	25.9	1.37×10^{-3}	1.60×10^{-4}	6.30×10^{-4}	2.53×10^{-4}	5.92×10-8	2.66×10^{-8}	4.36×10^{-6}	1.31×10^{-6}

S5. Parameter of Neural Networks and PSO in Implementing AKPE

AKPE was built with Matlab. According to the reaction model stated in the last section, we setup the neural network module of AKPE, the input layer contains m experimental data points obtained through experiments, the hidden layer contains ten nodes, the outputs of the neural network represents dimer concentration, which we will use to compare with corresponding experimental values, trainable parameters of the neural network module are all the coefficients from the reaction model, weights and offsets from neural networks. PSO is implemented in another module of AKPE, minimizing a combined loss from both the experimental data fitting and the neural network's approximation of differential equations. The final parameters of neural networks, the number of population and iteration of PSO are provided here in Table S2, S3 and S4. ×10

Table S2. The parameters of the neural network in AKPE

w1	b1	α
2.68×10-05	0.67025473	0.9998435

2.39×10-05	0.99863649	0.41425477
-6.27×10-06	0.94169232	-0.4998373
6.17×10-06	0.01840251	0.44436241
-2.93×10-05	1	-0.4164729
-8.39×10-06	0.88240921	-0.4948223
-5.99×10-05	0.99506593	0.27270384
4.08×10-06	0.10531877	0.88100106
-1.01×10-05	0.99994034	-0.4997785
-1.40×10-05	0.98617847	0.15259489
7.26×10-05	0.00428411	-0.1510174
-7.00×10-05	0.43940955	0.0282455
-3.03×10-05	0.00177412	-0.3120699
4.99×10-06	0.0298546	0.89503142
7.49×10-05	0.99284843	-0.1778428
6.98×10-05	0.0013665	-0.2878194
9.06×10-05	0.49073122	0.26599752
2.19×10-05	0.94554073	0.79305023
-9.98×10-05	0.74292142	-0.4416691
-5.99×10-06	0.40098239	-0.4999235
-0.0001	0.00276597	0.75822455
6.00×10-05	0.97220853	-0.1793768
-6.93×10-06	0.81064372	-0.3926612
-5.88×10-06	0.8620817	-0.4725592
5.35×10-05	0.44359862	-0.4610584
7.36×10-05	0.62739504	0.00693284
-1.43×10-05	0.99776007	-0.4032593
-4.76×10-05	0.00720816	0.54666463
9.51×10-05	0.19593137	0.76335488
-9.98×10-05	0.92304722	-0.4459476

The number of iterations of PSO is 100,000, the swarm size of PSO is 29, therefore, the personal best experience and the velocity of particles are produced as two 29×99 2D matrices, the personal best fitness is a 29×1 matrix, the output of PSO are the best solution (best_x) and the least error (best_fitness-opt_f=7.00), the best solution and the personal best fitness produced by PSO are given below:

Table S3. Best solution given by PSO

b×10st_x				
2.68×10-05	-0.0001	0.00428411	0.9998435	0.75822455
2.39×10-05	6.00×10-05	0.43940955	0.41425477	-0.1793768
-6.27×10-06	-6.93×10-06	0.00177412	-0.4998373	-0.3926612

6.17×10-06	-5.88×10-06	0.0298546	0.44436241	-0.4725592
-2.93×10-05	5.35×10-05	0.99284843	-0.4164729	-0.4610584
-8.39×10-06	7.36×10-05	0.0013665	-0.4948223	0.00693284
-5.99×10-05	-1.43×10-05	0.49073122	0.27270384	-0.4032593
4.08×10-06	-4.76×10-05	0.94554073	0.88100106	0.54666463
-1.01×10-05	9.51×10-05	0.74292142	-0.4997785	0.76335488
-1.40×10-05	-9.98×10-05	0.40098239	0.15259489	-0.4459476
7.26×10-05	0.67025473	0.00276597	-0.1510174	10000000
-7.00×10-05	0.99863649	0.97220853	0.0282455	5.59×10-06
-3.03×10-05	0.94169232	0.81064372	-0.3120699	9.82×10-10
4.99×10-06	0.01840251	0.8620817	0.89503142	0.00719144
7.49×10-05	1	0.44359862	-0.1778428	3.45×10-08
6.98×10-05	0.88240921	0.62739504	-0.2878194	4.78×10-08
9.06×10-05	0.99506593	0.99776007	0.26599752	1.0003052
2.19×10-05	0.10531877	0.00720816	0.79305023	1.99993671
-9.98×10-05	0.99994034	0.19593137	-0.4416691	1
-5.99×10-06	0.98617847	0.92304722	-0.4999235	

Table S4. Personal best fitness given by PSO

P_f
7.07660066
7.00249126
7.00242878
7.00246221
7.00243235
7.00250859
7.00247682
7.00248454
7.00243616
7.00255289
7.00242918
7.00241916
7.00247937
7.0024817
7.18122807
7.00241897
7.00243916
7.00249145
7.00251989
7.00251212

7.00247941
7.00248878
7.00243025
7.04504704
7.00241983
7.00239959
7.12370954
7.00244889
7.00242239

References

1. He CH, Bayakhmetov S, Harris D, Kuang Y, Wang X. A Predictive Reaction-Diffusion Based Model of E.coliColony Growth Control. Ieee Control Systems Letters. **2021**;5:1952-7.
2. Protein Expression - E.coli - Improving Protein Stability - EMBL.pdf
3. Zhao TK, Sun YB, Zhu QH, Li X, Baghaee M, Wang YN, et al. A contraction-reaction-diffusion model for circular pattern formation in embryogenesis. Journal of the Mechanics and Physics of Solids. **2021**;157.
4. Wang SS, Qi J, Diagne M. Adaptive boundary control of reaction-diffusion PDEs with unknown input delay. Automatica. **2021**;134.