UNC-CH PROGRAM IN CELLULAR AND MOLECULAR BIOPHYSICS

DMD simulation of critical size for nucleation kinetics in Superoxide Dismutase Aggregation

Dokholyan Lab

Li, Zimeng

8/24/2011

This report contains two parts. First part is a summary of my work in the project in the title; the second part outlines details of my work and understanding of the project, as well as proposals to future directions.

Content

Summary

My work in the project involves the following (in time order)

- Write the code to count number of atoms in different types in PDB file, which is used to calculate concentration
- Do the simulation of basic chemical chain reactions described by Master equations.
- Write the code to analyze the concentration vs. time graph, namely fit the data to master equation to obtain rate constant and equilibrium constant, and compare them with DMD parameters. Numerical integration is carried out in establishing the theoretical curves in concentration vs. time graph.
- Write the code to generate any length of polymers at any proportions of concentration, including rigorous reaction links within n-imers
- Write the code to separate oligomers & aggregates into their individual subunits and determine the concentration for each of them. For example, the code can distinguish oligomers M_2, M_3, M_4, M_5… even though they are assigned the same atom type

Planned simulations but hadn't done –

- Run the simulation with preset of n*-imer, draw the distribution of different i-imers, and from the distribution the minimum determines the critical value of n*
- DMD simulation determination of Aggregate pathway – trimer as nucleus for aggregates or trimer as competing species of aggregates
- Modification of DMD program (ydmd.linux) to make more realistic models – aggregate with polymer's 'spring' geometry and distinguish oligomers of same atom type within the DMD program itself.

The research experience here gives me taste for computation for the first time. Computation is very different from doing experiments in that computation needs to make hypothesis, build models, test models, and modify models if not work. Besides, computation doesn't have the disadvantage of doing experiment which may be effected by outside environment (e.g. earthquake, joke), and most importantly it is controllable with very large freedom. It is really an enjoyment when the proposed model is working and working day and night as if I discover a mysterious world. Programming is fun to me and has accompanied some of my best time.

I originally didn't place the project in the highest priority, but rather put more effort in reading papers and books. My initial motives in priority order for the four months long summer rotation is to read all papers in the lab (roughly ~150) to search for the topics I'm interested in (The lab does a wide range of topics), to systematically read a molecular biology book, and to finish the project with expected result. But it turned out the project is more time-consuming and yet more time-rewarding that I hesitated on my strategies from time to time, eventually leading to satisfactory results in none of my three goals. In the end, I only finished reading the book, but only read ~20 papers and finished ~70% of the project.

The lab group, in my opinion, is very rigorous and yet with humor. High attendance is observed for group members each working day, Prodeep and Onur wouldn't bother working at weekends. Group meetings are always spurs on my back, reminding me the distance between me and the discussion participants every time. I have to take video recordings during group meeting for replays back home to understand better. Challenge is the theme during my first rotation; I spend significant time to understand protein folding, which is not my project. The lack of connections between my project and group meeting topics leaves me dilemma on which one should I focus. However, even though the result is not what I expected, the process is rewarding and I'm sure I'll continue finishing the remaining if I can be as efficient as Poincare. It's not easy for someone like me who is too emulous to give up anything worth fighting for.

DMD simulation of critical size for nucleation kinetics in Superoxide Dismutase Aggregation

Introduction

The familiar form of amyotrophic lateral sclerosis (FALS) is linked to the unconditional aggregation of mutated Cu, Zn Superoxide Dismutase (SOD1). Nucleation kinetics is believed to control the amyloid fibril formation, a possible channel to aggregate formation in many deceases, especially the Huntingtin Decease[1]. We postulate that the aggregation of SOD1 is also controlled by nucleation kinetics, characterized by a critical nucleus that represents the free-energy barrier maxima. We use DMD to simulate the nucleation process in SOD1 aggregation and measured the critical size of the nucleus.

Procedure

Master equation simulation

We first test our DMD simulation on simple chemical kinetics, namely permutations and second order reactions, *in vitro*. We use bond and non-bond interactions to simulate free energy barrier in chemical kinetics. Rate constant k+ and k- (corresponding to activation energy), equilibrium constant Kd (corresponding to free energy difference ΔG) are measured after simulation. We further test the model on chain reactions, especially the SOD1 aggregate pathway described by the following equation [2]–

$$\mathrm{D} \leftrightarrows M_H \leftrightarrows M_A \leftrightarrows Aggregates \quad (1)$$

Where D stands for SOD1 dimer, M_H is holo-monomer (with metal), and M_A is apo-monomer (without metal).

Without specifying the rate constant obtained in the experiment[2], which consumes significant computation time due to the slow reaction process, we instead sets a series sets of rate constant and number of SOD1s that would fall within an acceptable computation time without losing a good statistics of Arrhenius equation[3] and Boltzmann distribution[4] in equilibrium. All

the above simulation runs flawlessly and produces precise result of dynamics and equilibrium by direct interpreting and solving master equations. Even though direct simulation of experimental rate constant is not affordable, we can use the exponential shift method (see Methods) to model the much slower experimental rate constant indirectly, should the *in vitro* massive environment be captured by Master equations.

WT SOD is stable in dimers at physiological level; however equilibrium won't be affected even if we start our simulation at any other states of SOD, e.g. monomers, should the reaction equilibrium be guaranteed by the Boltzmann distribution. We start our simulation with initial concentration fully composite of monomer SODs. Our result shows reaction kinetics of Eq.(1) is fully satisfied. Later, to mimic the reaction with starting components at different proportions, e.g. 84% dimer and 16% monomer as is achieved by incubating dimers in pH 3.5 after 24 hours[2], or starting with all dimers, or the later trimer simulation, we wrote the code to generate polymers at any length and any proportions of concentrations.

One thing to note that $M_H \leftrightarrows M_A$ is actually a second order reaction$M_H \leftrightarrows M_A + metal$, SOD monomer usually accompanies with one metal molecule. The decease related metal Zn loss destabilizes SOD monomer and is thought to trigger the aggregation [2]. We didn't do the cumbersome second reaction of metal loss, but instead consider it is Zn abundant and the metal loss reaction can be approximated by first order reaction. But this also means kinetics of metal loss is not captured as what it should be.

Nucleation Simulation

Drawbacks of the above procedure are – first it takes broad simplification that aggregates are dimers of two Apo-monomers; second it ignores the nucleation process during the process of$M_A \leftrightarrows Aggregates$. Reports show that the complete pathway is –

$$M_A \leftrightarrows M^* \leftrightarrows Nucleus \leftrightarrows Aggregates \tag{2}$$

Where M* is the misfolded form of Apo-monomers. There is a conformational change within the Apo-monomer during$M_A \leftrightarrows M^*$. Nucleus is a polymer of M* with a critical size n*, written as$(\mathrm{M}^*)_{n^*}$. Therefore the complete nucleation pathway is

$$M^* \leftrightarrows (\mathrm{M}^*)_{n^*} \leftrightarrows Aggregates \tag{3}$$

Or

$$M^* \leftrightarrows (M^*)_2 \leftrightarrows \cdots \leftrightarrows (\mathrm{M}^*)_{n^*} \leftrightarrows (M^*)_{n^*+1} \leftrightarrows \cdots \tag{4}$$

Where all size larger than n* will be considered aggregates due to the energy favorable downhill in nucleation kinetics after the critical value.

We thereby introduce the nucleus model in our aggregate simulation. The free energy barrier is shown in Fig.1 and Fig.2

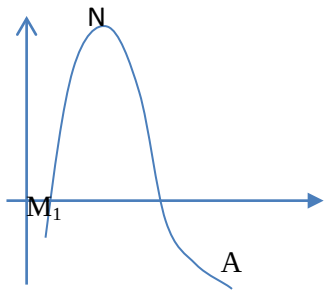

Figure 1

Simplified free energy barrier, N is nucleus

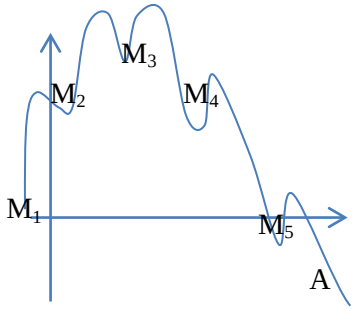

Figure 2

Free energy barrier of Nucleus as trimer

The choice of the individual barriers in Fig. 2 for each nucleation step has to be distinguished before and after the critical nucleus. Before aggregate to nucleus, the dissociation of oligomers is energetically favorable. When reaching the critical size, the intermediate acts like the top of the activation energy barrier in Fig.1. Because the nucleus intermediate is metastable, we expect the dissociation and association of one monomer during nucleus state should be the same, making it a very rare case to find any nucleus during reaction. This can be used as criteria to evaluate critical nucleus size. We expect the distribution of oligomers in terms of number of subunits as Fig.3, where the minimum population appears in the location of critical value n*.

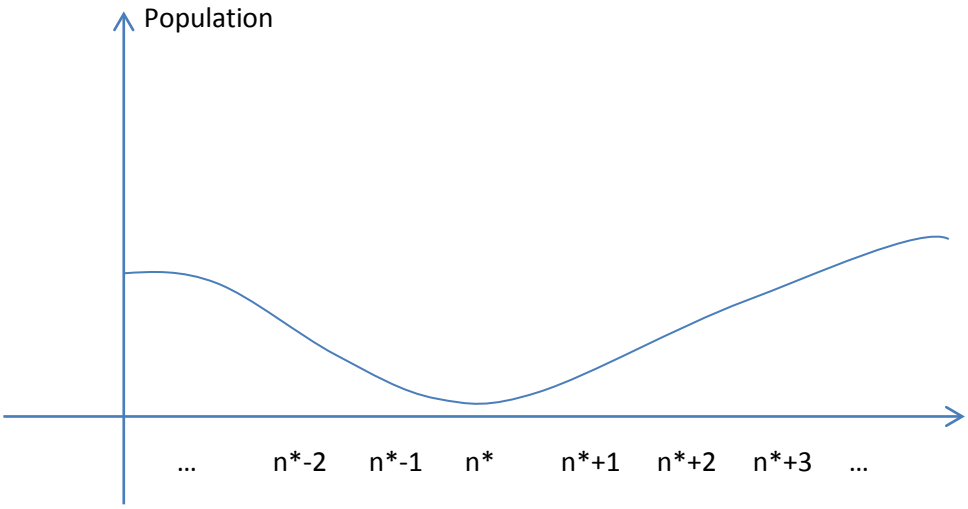

Figure 3

Distribution of oligomers

Since the rate constant for nucleus elongation is assumed to be identical to the elongation of aggregate[5], we can use the same rate constant from experiment and assume all individual barriers in Fig. 2 before nucleus share the same association rate constant (i.e. aggregation rate constant). A different approach is to apply the same pairwise contact potential to individual atoms within the same oligomer along the entire course towards aggregation. This model can be seen as a microscopic explanation to phase transition (see methods).

Pathway Determination

SOD aggregation is suspected to be undergoing one of the following two pathways, (5) and (6)

$$\mathrm{D} \leftrightarrows M_H \leftrightarrows M_A \leftrightarrows M^* \leftrightarrows N \leftrightarrows A \tag{5}$$

$$\mathrm{D} \leftrightarrows M_H \leftrightarrows M_A \leftrightarrows M^* \leftrightarrows N$$

$$\mathrm{D} \leftrightarrows M_H \leftrightarrows M_A \leftrightarrows M^* \leftrightarrows A \tag{6}$$

The difference between the two pathways is, one is aggregate prone and the other is aggregate competing. N (Nucleus) can be any oligomer with the format of $(M^*)_{n^*}$. Once n* is determined by the procedure in nucleation simulation, we run the simulation of the two pathways. The only measurable counterparts in experiment are the rate constant and equilibrium constant. We cannot detect the $M_A \leftrightarrows M^*$ in both pathways and $M^* \leftrightarrows N$ in pathway(6). These two undetectable or hypothesized reactions are manipulated to a level that is larger than

aggregation rate and yet comparable with dimer dissociate rate. As simulation goes, we record the concentration of different SOD components versus time, and we will observe a difference in the shape of the data plot between the pathways. A comparison with experiment curve (which is missing the undetectable parts) will show which pathway is taken in SOD aggregation.

Again, to test the nucleus size independently of the method mentioned in nucleation simulation, we apply the noted theory of nucleation kinetics (see methods) to analytically calculate the nucleus critical size. We further compare the two critical values to verify our hypothesis of the phase transition model at microscopic level.

Results

Master Equation Simulation

Equation(1) is rewritten in differential form[2] and fit to the data simulated in DMD. Different starting concentrations of different proportions of SODs are tested. The simulation runs flawless and reproduces the Master equation with adequate large ensembles. In the population vs. time step figure below, starting concentrations are full monomers; we fit the curve using the differential form of master equations, and obtained precise result with the parameters set in the DMD simulation.

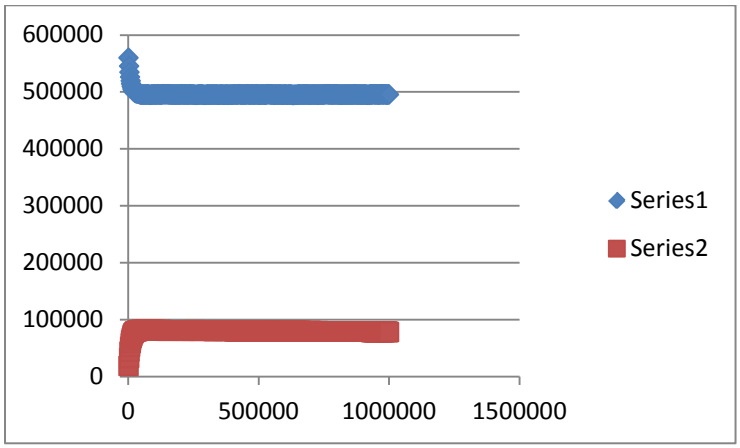

Figure 4

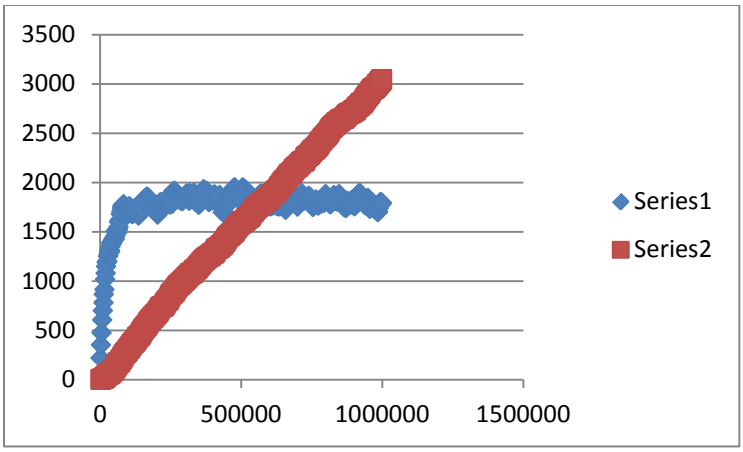

Figure 5

Fig. 4 Blue – Holo-monomers; Red – Apo-monomers

Fig.5 Blue – dimers; Red – aggregates (aggregate dimers)

A summary of the result is put in the table below –

Parameter	Direction	DMD Parameter	Master Equation Curve Fit	Error	Unit
Permutations	Holo to Apo	1e-5	9.95670819916208e-06	0.4%	time step*
	Apo to Holo	1.5e-5	1.44835590851932e-05	3.5%	
Non-bond Interaction	Holo	1	1.20652445587088	20%	Kcal/mol
	Apo	1.5	1.78409615019418	18.6%	

Bond Interaction	Dimer	-2	1.99978117233202	0.01%
	Aggregate	-6.5	6.50716026064839	1%

Table 1 Simulation result running at T=1.2 k_B**

*1 time step = 4.89e-14s

**k_B is Boltzmann constant.

The relative large discrepancies of non-bond interactions are due to initial large concentration of monomers and can be corrected when running on starting concentration of full dimers, which has a big discrepancy on bond interactions and little discrepancy on non-bond potentials.

Nucleation Simulation and Pathway determination

This part will be finished if problems relating to the incapability of the current DMD program (not DMD itself) are solved (see discussion)

Methods

MD

Classical MD (Molecular Dynamics) theory employs classical mechanics and classic statistical mechanics to simulate system of large number of particles in a limited volume. Specifically, Newton's law to solve equations of motion and canonical ensemble to solve parameters for the bulk system are generally used[6]. Other computational methods, including Monte Carlo, Markov chain random walk, finite element and finite difference method, implement the tools available for powerful computer simulation.

DMD Simulation

DMD (Discrete Molecular Dynamics) is a simplification of MD (Molecular Dynamics)[7]. The first simplification is the interaction potential – while MD uses continuous interaction potential[8], such as the Lenard Jones potential, DMD uses pairwise contact potential that is discretized as step potentials (or square well potential). Another simplification is the search algorithm, DMD employs an optimized search algorithm called collision table, a similar but more efficient type than its counterpart - event tree used in MD. DMD also uses random walk in establishing event order.

The specific strength of DMD after such broad simplification of MD is that it captures the essence of bulk system without compromising much computational power. Master equation will not work in relatively small system with notable nonlinearity (huge statistic fluctuation). Therefore any master equation compatible with *in vitro* system won't work *in vivo* conditions, where number of particles of interest is significantly smaller. For example, it is thought that *in vivo* we only have a few thousand SOD1 in human motor neuron cell; while *in vitro* we could

produce numerous SOD1 for experimental detection. Therefore, DMD is especially useful here in illustrating the dynamics in vivo, as an independent tool besides master equation.

Nucleation Kinetics

Wetzel R., et al. first introduced nucleation study in polyglutamine aggregation, a suspect pathway to Huntingtin Decease (HD)[1]. The overall nucleation is a complicated process, but essence can be captured in the initial phase of Nucleation kinetics. The integrated rate equation is obtained from the following simplifications:

Concentration of nucleus:

$$\frac{dc^*}{dt} = J^* c^* \Rightarrow \Delta c^* = J^* c^* \Delta t \quad (7)$$

Concentration of monomers that have gone to polymers:

$$\frac{d\Delta c^*}{d\Delta t} = J\Delta c^* = JJ^* c^* \Delta t \Rightarrow \Delta c^* = \frac{1}{2} JJ^* c^* t^2 \quad (8)$$

The symbol above stands for: c*-concentration of nucleus, J* - elongation rate of nucleus, t – time, J – elongation rate of polymers

Substitute $c^* = K_{n^*} c^{n^*}$ and $J = k_+ c$ into the nucleation master equation above, we obtain the exclusive form of the equation with measurable quantities - $\Delta = \frac{1}{2} k_+^{\,2} K_{n^*} c^{(n^*+2)} t^2$. Note that K_{n^*} is the equilibrium constant describing the monomer and nucleus equilibrium and k+ is the forward elongation rate constant. n* is the critical nucleus size that distinguish the energy favorable states of monomer and aggregates.

The simplified nucleation master equation produced a slope of $k_+^{\,2} K_{n^*} c^{(n^*+2)}$ when plotting Δ vs. t. Notable, the rate constant changes with concentration c, so another plot of log (slope) vs. log(c) will obtain the critical nucleus size. Precise experiment has been carried out to measure n* in Poly(Gly) aggregate formation[9], and was found to be Polymer repeat length dependent[5]. The measured critical nucleus size produces reasonable age-of-onset estimate, which verifies the success of nucleation kinetics in some aggregate pathways.

Phase Transition

Statistically, phase transition occurs in homogeneous nucleation theory when sporadic events of high energy particles overcome the energy barrier of nucleus. Microscopically, this can be attributed to a simple model as described in Fig. 4

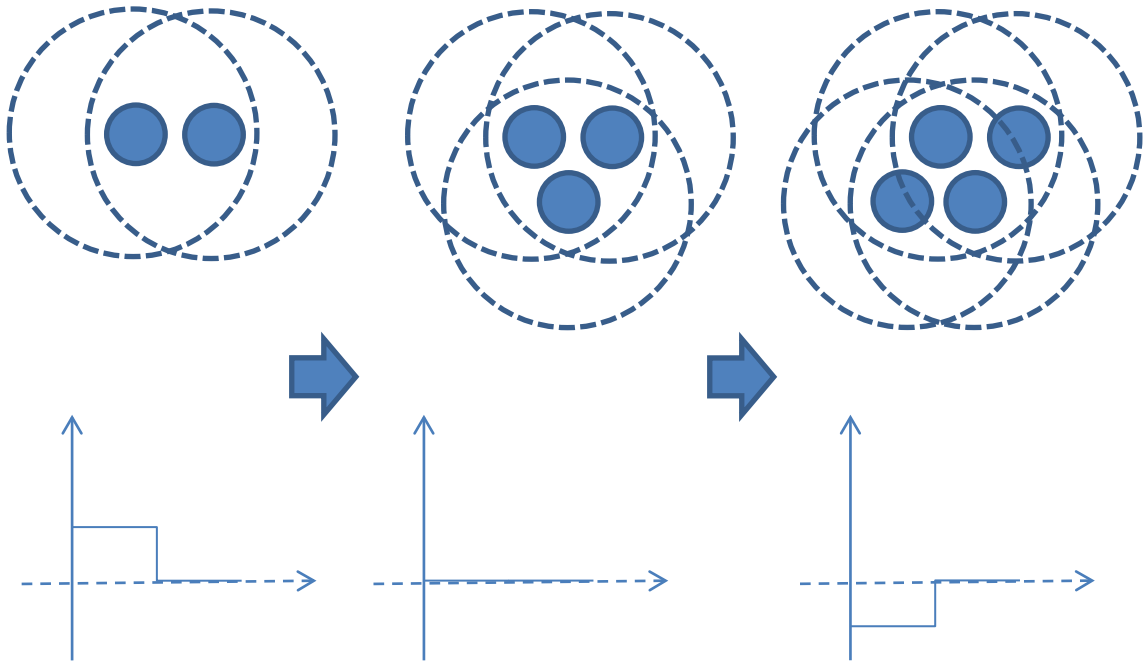

Figure 6

Microscopic explanation for phase transition, with n*=3

Figure 4 sketches the nucleation-aggregation pathway. Two Apo-monomer (M*) forms dimer $(M^*)_2$, the nucleus, in an energetically unfavorable manner (dissociation > association). The energy for Apo-dimer is larger. This pairwise interaction energy won't change in the course of aggregation, saying, the energy for dissociation is always small than the energy for association no matter what polymer state the aggregate is. However, as is shown in the middle figure, even though the pairwise interaction remains the same, monomer in trimer has to overcome two energy barriers to be dissociated, which will represent a slower dissociation rate. In the case of n*=3, the slowed dissociation rate equals to the unchanged association rate (dissociation = association). If more bonds are formed, e.g. tetramer, we can foresee that association will surpass dissociation. Writing the rate constant in terms of energy barrier, we can see a constant change from square barrier to square well.

Energy Shift Method

Even the reaction in vitro is too slow to be used in DMD simulation. The reaction has to be comprised to a manipulated fast rate by applying low energy barrier in DMD parameters. This energy barrier theory applies only to large system where general mean field theory could work and generate good statistics. Even though, we still face the free energy difference, which is about ~9kcal/mol within the dimer-monomer equilibrium and about the same magnitude within

the Apo-monomer and aggregate equilibrium. It would be impossible to reach equilibrium in a reasonable time. Therefore both simulating real life reaction rate and equilibrium distribution are not feasible.

An alternative way is to shift the free energy barrier in the same manner as shifting the activation energy barrier. The advantage of doing this is simulation can be performed without sacrificing much computer time. The disadvantage is that energy shift uses Arrhenius equation and Boltzmann distribution theory, which is only a good approximation in large scale system. It may apply to in vitro system, but it won't resolve the nonlinearity in system with small number of particles, i.e. the in vivo case. Therefore, any conclusion from master equation solvable problem would have little interpretations on the decease related phenomena.

Discussion

Incapability

The incapability of the current DMD program is technical problem, but is solvable. Notably, the current criteria to distinguish the atoms before reaction and after reaction make the simulation unworkable.

```
REACTIONS
1  1  3  3  44  0  1
1  3  3  3  1  44  0  1
```

As is shown above, if we assume all aggregate oligomers are the same atom type (we can't set up all oligomers atom types in the parameter file), e.g. type 3 in the picture, first we cannot distinguish which one of atoms in the broken link will be 1 or 3, in the second line of the REACTIONS section. For example, in the following figure,

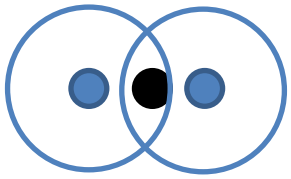

Figure 7

This is a trimer; the two blues are not bonded, but is connected with both of them bonds to the black one. If we assume the bonded ones are type 3, therefore if using the protocal in the above parameter set up, the program cannot decide whether blue or black should be type 1 or 3 after reaction, should they break the bond. It is possible that the blue remains type 3 while the black one turns to type 1 after breaking bonds, which means there are type 3 atoms which are not bonded after reactions! (Type 3 atoms are set up as bonded atoms)

This also comes to problem when dealing with aggregate-dimers, as is illustrated in the figure below

Figure 8

When the bonded blues in an aggregate-dimer on the left side break bond, there is always one monomer that remains blue (type 3). However, the reality is both the monomers should be black (type 1). We try to improve the situation by applying the first line in the REACTIONS section above, however DMD program will rewrite the first line if the second line is read into memory. Therefore you can never assign two situations to bonded reactions in practice. This is something has to be done with the current DMD program's incapability.

Again, if we try to use different types of atoms (instead of using the same type for all oligomers), more serious problem will occur in that the other atoms in the same oligomer won't change their types if more atoms are aggregated to the oligomer. This is due to DMD only deal with pairwise contact potential and related atom type changes, it won't do anything on other atoms in the same oligomer. For example, if a trimer obtains a monomer to become a tetramer, the two atoms involved in the new bond formation changed their types, but the other two old atoms in the trimer won't change their type. This becomes particular dangerous if we assume trimer is the nucleus and there will be significant interactions changes before and after the critical value, saying, the two old atoms in the trimer remains aggregate unfavorable while the newly bonded atoms become aggregate favorable and readily going downhill the free energy barrier.

I asked other people, Feng and David. We haven't thought up a way dealing with this situation except modifying the DMD program. This is eventually doable but we need to reconsider whether it is plausible to sacrifice time and energy on this project.

Significance

This project is more like a test of physics phase transition theory in a biological system. It has plausible significance in the theory and also in understanding the process of the reactions. Other than that, I haven't found any applications in practical industrial application.

Acknowledgement

Special thanks go to David and Feng with the discussion, and Prodeep, Rachel and Liza's generous help with my other lab problems.

Resources

Recourses can be downloaded on my website: http://www.physics.unc.edu/~lizimeng/

Bibliography

1. Chen, S., F.A. Ferrone, and R. Wetzel, *Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11884-9.
2. Khare, S.D., M. Caplow, and N.V. Dokholyan, *The rate and equilibrium constants for a multistep reaction sequence for the aggregation of superoxide dismutase in amyotrophic lateral sclerosis.* Proc Natl Acad Sci U S A, 2004. **101**(42): p. 15094-9.
3. Brown, T.L., *Chemistry. The Central Science*. 9 ed. 1038.
4. House, J.E., *Principles of Chemical Kinetics*: Wm. C. Brown Publishers.
5. Kar, K., et al., *Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent.* Nat Struct Mol Biol, 2011. **18**(3): p. 328-36.
6. Ma, W.K., *Computational Physics*. 1 ed. 2000, Hefei: USTC Press. 165.
7. Proctor, E.A., F. Ding, and N.V. Dokholyan, *Discrete molecular dynamics.* John Wiley & Sons, Ltd, 2011. **1**.
8. DC., R., *The Art of Molecular Dynamics Simulation*. 2 ed. 2004: Cambridge University Press.
9. Bhattacharyya, A.M., A.K. Thakur, and R. Wetzel, *polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15400-5.