An Agent-Based Model for Molecular Self-Organization

Sara Fortuna and Alessandro Troisi

CSC Lunchtime Seminar

February 2, 2009
Outline

1. Molecular Self-Organization
   - Molecular Dynamics
   - Monte Carlo
   - Genetic Algorithms

2. Molecular ensembles are Complex Systems
   - Cellular Automata and Ising Model
   - Agent-based

3. The Agent Based Model
   - Definitions
   - Rules
   - Overall Algorithm
   - Results

4. Conclusion
Self-Organizing systems
.. in everyday life..

soap-bubbles!
Self-Organizing systems
.. in biological systems..

amyloids
www.shef.ac.uk/mbb/staff/staniforth

cell membranes
tobacco mosaic virus

Sara Fortuna
An Agent-Based Model for Molecular Self-Organization
Non-covalent interactions are the self-organization driving forces:

- electrostatics (ion-ion: $100 - 350 \text{kJmol}^{-1}$)

- hydrogen bonding ($4 - 120 \text{kJmol}^{-1}$)

- $\pi - \pi$ stacking interactions ($< 50 \text{kJmol}^{-1}$)

- hydrophobic effects ($< 40 \text{kJmol}^{-1}$)

- dispersion forces (van der Waals: $< 5 \text{kJmol}^{-1}$)
Molecular Self-Organisation

Free Energy Surface

- many local minima
- many kinetic traps
- difficult to model

Wales, 2006
Molecular Dynamics
Application: cytosine self-assembly in 2D

90 13-atoms molecules, all atom representation, Amber force-field: 0.5ns in 5 hours on COW
Improving Molecular Dynamics

MD is limited to very small time scales.

MD implementations:
- Hyperdynamics (Voter, 1997)
- Activation-Relaxation Techniques (Mosseau, 2000)
- Metadynamics (Laio, 2002)
- temperature accelerated methods (Maragliano, 2006)
- good to study dimers, trimers, ... 

DISADVANTAGES: small time scales, small system sizes.
Simple systems can be studied with the Metropolis algorithm.

*simulation*

*experiment*
## Improving Monte Carlo

<table>
<thead>
<tr>
<th>MC implementations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- biased MC simulations (<strong>Wu, 1992</strong>; ...; <strong>Liu, 2004</strong>)</td>
</tr>
<tr>
<td>- Basin hopping (<strong>Wales, 1997</strong>)</td>
</tr>
<tr>
<td>- mixed MC-stochastic dynamics methods (<strong>Guarnieri, 1995</strong>)</td>
</tr>
<tr>
<td>- data augmentation (<strong>Troisi, 2005</strong>)</td>
</tr>
</tbody>
</table>

**ADVANTAGE**: multiscale moves  
**DISADVANTAGE**: cannot escape all the kinetic traps!
Data Augmentation
Application: Micelle formation or Fibres
Bhattacharyay, 2008
Genetic Algorithms
this is an example of Artificial Intelligence

**GA characteristics:**
- search technique based on natural evolution
- rule based
- good for cluster optimization

**ADVANTAGE:** the system evolves in time
**DISADVANTAGE:** the individuals does not communicate!
At every step the *individuals* through a *crossover operations* generates a new *population*.

---

*Johnston, 2003*
Self-Assembling Systems are Complex Systems

Sara Fortuna An Agent-Based Model for Molecular Self-Organization
Self-Assembling Systems are Complex Systems
Self-Assembling Systems
are Complex Systems

Hierarchical structures
Self-Assembling Systems are Complex Systems

- Hierarchical structures

Sara Fortuna
An Agent-Based Model for Molecular Self-Organization
Self-Assembling Systems
are Complex Systems

- Hierarchical structures
Self-Assembling Systems
are Complex Systems

- Hierarchical structures
- Multiple scales
Self-Assembling Systems are Complex Systems

- Hierarchical structures
- Multiple scales
- Cooperative effects
Self-Assembling Systems are Complex Systems

- Hierarchical structures
- Multiple scales
- Cooperative effects
Self-Assembling Systems are Complex Systems

- Hierarchical structures
- Multiple scales
- Cooperative effects

COMPLEX SYSTEMS
Self-Assembling Systems are Complex Systems

- Hierarchical structures
- Multiple scales
- Cooperative effects

RULE BASED MODELS
Cellular Automata

Discrete rule based models.

- 1D: Wolfram models (1983)

- 2D: Ising models, Potts models

Those are local models able to describe the global behaviour of a system (e.g.: phase transitions).
Agents
An agent is an autonomous system

An agent is capable of sensing his environment, taking decisions and performing autonomous actions.

**Example**

A thermostat!

It switches autonomously on or off basing his decision on the temperature of its environment.
The model is based on the collective behaviour of a set of *agents*. Agents have properties (e.g., position) and actions (e.g., move).
Self-assembly agents

An agent $A_i$ is identified with a collection of particles $n_i$ (a cluster)

- the starting dimension of $A_i$ is 1 (i.e. each cluster initially contains only one particle)
- $A_i$ then evolves following a set of rules
Why Agents?

We can combine all the advantages of the other techniques:

- MC: multiscale moves
- GA: evolution

We call *agent* a stable portion of the system,

and let each portion of the system decides when doing what.
The system

- $N$ particles
- 2 hierarchical levels

### (1) Particles

**Properties:** position and orientation in space;

**Actions:** MC moves (i.e. translations and rotations);

### (2) Agents

**Properties:** composition, position and orientation in space;

**Actions:** MC moves, Merge with other Agents, Split, Disaggregate;
Rules

An Agent can choose among a set of actions.

Every action is performed if a condition is satisfied.

\[
\text{RULE} = \text{CONDITION} + \text{ACTION}
\]

The conditions depend on the current configuration of the system:
- interaction energy between two agents
- agent internal energy
The actions an agent can perform are:

- move
- merge
- split
- disaggregate
Move

Both single particle moves and agent moves can be performed

\[ P(s \rightarrow s') = \min[1, e^{-\beta(E_{s'}-E_s)}] \]  \text{ with } \beta = \frac{1}{k_B T} \]  \text{ (1)}

Sara Fortuna
An Agent-Based Model for Molecular Self-Organization
If the interaction energy between two agents is smaller than a certain threshold, the agents will merge.

\[ E_{ij} \leq E_M \quad \Rightarrow \quad \text{MERGE} \]  

(2)
If the internal energy of an agent agent is greater than a certain threshold, the agents will lose a particle

\[ E_k > E_S \Rightarrow \text{SPLIT} \]

the ejected particle will form the new agent.

This action will be repeated iteratively until \( E_k \leq E_S \).
Two agents $k$ and $i$ are considered part of the same aggregate $\mathcal{A}$ if their interaction energy is lower than a certain threshold energy $E_D$:

$$E_{ki} < E_D \Rightarrow A_k, A_i \in \mathcal{A}$$ (3)

If an agent forms an aggregate with more than 3 agents of his surrounding, the aggregate should “disaggregate”.
Disaggregate Spectral bisection

Spectral bisection

1. connectivity matrix definition
2. diagonalization of the connectivity matrix
3. partition of $\mathcal{A}$ in two connected blocks $\mathcal{A}^-$ and $\mathcal{A}^+$
Disaggregate
Spectral bisection

**Spectral bisection**

1. connectivity matrix definition
2. diagonalization of the connectivity matrix
3. partition of $\mathcal{A}$ in two connected blocks $\mathcal{A}^-$ and $\mathcal{A}^+$
Overall Algorithm

1. Choose one agent
2. Check if to disaggregate?
   - Yes: Disaggregate
   - No: Agent move
3. Check if to merge?
   - Yes: Merge
   - No: Split
4. Choose a particle
5. Particle move
6. End simulation

Equilibration steps:
- Start loop
- MC or AB step?
- Choose one agent
- Disaggregate?
- Yes: Disaggregate
- No: Agent move
- Merge?
- Yes: Merge
- No: Split
- Split?
- Yes: Split
- No: End simulation
- End loop
Molecular Self-Organization

Molecular ensembles are Complex Systems

The Agent Based Model

Conclusion

Definitions

Rules

Overall Algorithm

Results

choose one agent
disaggregate?

yes

no

no

yes

merge?

merge

no

split?

split

yes

Sara Fortuna

An Agent-Based Model for Molecular Self-Organization
Molecular ensembles are Complex Systems

The Agent Based Model

Conclusion

Definitions

Rules

Overall Algorithm

Results

Sara Fortuna  An Agent-Based Model for Molecular Self-Organization
### Parameters

<table>
<thead>
<tr>
<th>Description</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>fraction of single particle moves</td>
<td>$\chi_{MC}$</td>
</tr>
<tr>
<td>merge energy</td>
<td>$E_M$</td>
</tr>
<tr>
<td>split energy</td>
<td>$E_S$</td>
</tr>
<tr>
<td>disaggregate energy</td>
<td>$E_D$</td>
</tr>
<tr>
<td>split temperature</td>
<td>$k_B T_S$</td>
</tr>
<tr>
<td>disaggregate temperature</td>
<td>$k_B T_D$</td>
</tr>
<tr>
<td>max agent size</td>
<td>$n_{max}$</td>
</tr>
</tbody>
</table>
Adaptation

$E_M, E_S, E_D$ depend on the energy of the most stable agent of the same size formed during the simulation.

There is an array $E_{\text{min}}[n]$ that keep track of the minimum energy of an agent of each size.

The reference energies are updated as the simulation evolves.
Model systems

Interaction Energy

\[ E_{kl} = \sum_a \sum_b [E_{ab}^{LJ} + E_{ab}^C] \quad \text{where} \quad a \in k, b \in l. \quad (4) \]
Agent growth

(a) 110,000 steps
(b) 290,000 steps
(c) 480,000 steps
(d) 2,000,000 steps
Agent growth

$n = 1$
$n = 2$
$n = 3$
$n = 4$
$n = 5$

Agent growth over 3 million steps.
Lower Energy after the same number of steps with respect to a Monte Carlo simulation

\[ \langle R \rangle = \frac{\langle E_{AB} \rangle - \langle E_{MC} \rangle}{\langle E_{MC} \rangle} \]

- **\( \chi_{MC} = 0.00 \)**
  \[ \langle R \rangle = 0.14 \]

- **\( \chi_{MC} = 0.30 \)**
  \[ \langle R \rangle = 0.52 \]

<table>
<thead>
<tr>
<th>shape</th>
<th>( \langle E_{MC} \rangle )</th>
<th>( \langle E_{AB} \rangle )</th>
<th>( \langle R \rangle )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \chi_{MC} = 0.00 )</td>
<td>( \chi_{MC} = 0.00 )</td>
<td>( \chi_{MC} = 0.30 )</td>
</tr>
<tr>
<td>(1)</td>
<td>-1039 ± 36</td>
<td>-1288 ± 33</td>
<td>0.24 ± 0.05</td>
</tr>
<tr>
<td>(2)</td>
<td>-1014 ± 26</td>
<td>-1198 ± 87</td>
<td>0.18 ± 0.09</td>
</tr>
<tr>
<td>(3)</td>
<td>-1020 ± 21</td>
<td>-1270 ± 51</td>
<td>0.25 ± 0.06</td>
</tr>
<tr>
<td>(4)</td>
<td>-1106 ± 19</td>
<td>-1321 ± 62</td>
<td>0.13 ± 0.06</td>
</tr>
<tr>
<td>(5)</td>
<td>-1202 ± 17</td>
<td>-1437 ± 85</td>
<td>0.20 ± 0.07</td>
</tr>
<tr>
<td>(6)</td>
<td>-1278 ± 48</td>
<td>-1494 ± 82</td>
<td>0.17 ± 0.08</td>
</tr>
<tr>
<td>(7)</td>
<td>-1225 ± 17</td>
<td>-1536 ± 95</td>
<td>0.25 ± 0.08</td>
</tr>
<tr>
<td>(8)</td>
<td>-1241 ± 23</td>
<td>-1462 ± 53</td>
<td>0.18 ± 0.05</td>
</tr>
<tr>
<td>(9)</td>
<td>-1141 ± 46</td>
<td>-1480 ± 35</td>
<td>0.30 ± 0.06</td>
</tr>
<tr>
<td>(10)</td>
<td>-1212 ± 29</td>
<td>-1342 ± 41</td>
<td>0.11 ± 0.04</td>
</tr>
<tr>
<td>(11)</td>
<td>-1386 ± 31</td>
<td>-1512 ± 3</td>
<td>0.09 ± 0.02</td>
</tr>
<tr>
<td>(12)</td>
<td>-1358 ± 48</td>
<td>-1414 ± 20</td>
<td>0.04 ± 0.04</td>
</tr>
<tr>
<td>(13)</td>
<td>-1364 ± 37</td>
<td>-1585 ± 20</td>
<td>0.16 ± 0.03</td>
</tr>
<tr>
<td>(14)</td>
<td>-1324 ± 29</td>
<td>-1590 ± 6</td>
<td>0.20 ± 0.03</td>
</tr>
<tr>
<td>(15)</td>
<td>-1422 ± 45</td>
<td>-1540 ± 10</td>
<td>0.08 ± 0.04</td>
</tr>
<tr>
<td>(16)</td>
<td>-1068 ± 35</td>
<td>-1256 ± 66</td>
<td>0.18 ± 0.07</td>
</tr>
<tr>
<td>(17)</td>
<td>-1188 ± 21</td>
<td>-1200 ± 106</td>
<td>0.01 ± 0.09</td>
</tr>
<tr>
<td>(18)</td>
<td>-1144 ± 23</td>
<td>-1080 ± 58</td>
<td>-0.06 ± 0.05</td>
</tr>
<tr>
<td>(19)</td>
<td>-1187 ± 52</td>
<td>-1188 ± 58</td>
<td>0.00 ± 0.07</td>
</tr>
<tr>
<td>(20)</td>
<td>-1070 ± 15</td>
<td>-1145 ± 58</td>
<td>0.07 ± 0.06</td>
</tr>
<tr>
<td>(21)</td>
<td>-1068 ± 23</td>
<td>-1451 ± 56</td>
<td>0.36 ± 0.06</td>
</tr>
<tr>
<td>(22)</td>
<td>-1256 ± 52</td>
<td>-1427 ± 52</td>
<td>0.14 ± 0.06</td>
</tr>
<tr>
<td>(23)</td>
<td>-1164 ± 42</td>
<td>-1530 ± 45</td>
<td>0.31 ± 0.06</td>
</tr>
<tr>
<td>(24)</td>
<td>-1202 ± 38</td>
<td>-1348 ± 69</td>
<td>0.12 ± 0.07</td>
</tr>
<tr>
<td>(25)</td>
<td>-1207 ± 35</td>
<td>-1298 ± 83</td>
<td>0.08 ± 0.08</td>
</tr>
<tr>
<td>(26)</td>
<td>-1208 ± 22</td>
<td>-1212 ± 52</td>
<td>0.00 ± 0.05</td>
</tr>
<tr>
<td>(27)</td>
<td>-1186 ± 31</td>
<td>-1475 ± 46</td>
<td>0.24 ± 0.05</td>
</tr>
<tr>
<td>(28)</td>
<td>-2083 ± 83</td>
<td>-1734 ± 55</td>
<td>-0.17 ± 0.04</td>
</tr>
</tbody>
</table>

average 0.14 0.52
Lower Energy after the same number of steps with respect to a Data Augmentation

![Graph showing energy vs. steps for different models MC, DA, AB]

Sara Fortuna
An Agent-Based Model for Molecular Self-Organization
The equilibrium configuration is more compact.
The equilibrium configuration is more ordered

Radial distribution function $g(d)$
interface with TINKER

apply the method to realistic systems
Conclusion

1. Models for Molecular Self-Assembly
   - MD for nucleus formation
   - MC with group moves for many particles
   - GA for cluster optimization

2. Molecular ensembles are Complex Systems

3. Agent-Based model for 3D off-lattice systems
   - new set of rules for molecular self-organisation
   - lower energy configuration is the same number of steps
   - more compact and longer range order
   - find the lowest energy for each cluster size
   - interface with TINKER (real molecule)
Acknowledgments

Thanks to:

The Leverhulme Trust

“Modelling Techniques for Molecular Self-Assembly”

- Alessandro Troisi
- Natalia Martsinovich

former members:

- Arijit Bhattacharyay (Pune, India)
- Konrad Diwold (Leipzig, Germany)