JCTC Journal of Chemical Theory and Computation

Graph Measures Reveal Fine Structure of Complexes Forming in Multiparticle Simulations

Florian Lauck,‡ Volkhard Helms,* and Tihamér Geyer*

Zentrum für Bioinformatik, Universität des Saarlandes, D-66041 Saarbrücken, Germany

Received September 23, 2008

Abstract: Modern simulation techniques are beginning to be used to study the dynamic assembly and disassembly of multiprotein systems. In these many-particle simulations it can be very tedious to monitor the formation of specific structures such as fully assembled protein complexes or virus capsids above a background of monomers and partial complexes. However, such analyses can be performed conveniently when the spatial configuration is mapped onto a dynamically updated interaction graph. On the example of Monte Carlo simulations of spherical particles with either isotropic or directed mutual attractions, we demonstrate that this combined strategy allows for an efficient and also detailed analysis of complex formation in many-particle systems.

Introduction

Proteins are the main building blocks and functional units of biological cells and are involved in signal transduction, energy metabolism, immune response, and the processing of DNA and RNA. Interestingly, many individual proteins do not work independent of each other. About half of them form protein complexes that consist of two or more proteins, which can be either identical or different in size, shape, and function.[1,2] Thus, a chain of metabolic reactions catalyzed by the components of a multienzyme complex can be executed in a shorter time, because intermediate substrates need not to diffuse through the cell. They will simply be handed over from one enzyme to the next one.[3,4] Another functional advantage of organizing single proteins into complexes is their modularity. A certain subunit can be used within more than one complex to accomplish the same task.[2,5]

In the past, the formation of protein complexes was first studied on a microscopic level, where the details of the binding interfaces of the proteins were investigated experimentally by X-ray crystallography, chemical shift mapping, or site-directed mutagenesis or theoretically with docking algorithms at atomic resolution.[6,7] For the modeling of larger protein assemblies, combinatorial methods have been developed, which can efficiently find the best configuration built from a number of building blocks.[8] Parallel to these static docking approaches, explicitly time-dependent simulations were used to investigate complex formation. These range from atomistic Brownian dynamics simulations of the association of two small proteins [9−12] up to coarse-grained simulations of protein pairs with reaction patches[13] or of virus capsids.[14,15] The largest system assembled sofar by simulation techniques is the nuclear pore complex consisting of 456 protein units in yeast.[16] While it is not a problem to interpret the results of simulations with only two or three particles, it can become a tedious and computationally demanding task to monitor, whether and when complexes with tens to hundreds of constituents are formed correctly. This problem becomes even more pronounced when more realistic simulations of large numbers of different proteins are considered, where more than one complete complex can be formed or where the complete complex is in a dynamic equilibrium with its components. In such a simulation, partial complexes of various sizes may be found together with complete ones and even with complexes which are assembled incorrectly.[17]

Recently, large databases have become available that provide convenient access to experimental and computationally generated data sets on protein−protein interactions

* Corresponding author e-mail: volkhard.helms@bioinformatik.uni-saarland.de; tihamer.geyer@bioinformatik.uni-saarland.de.

‡ Current address: Department of Biopharmaceutical Sciences, University of California, San Francisco, CA 94143.

10.1021/ct800396v CCC: $40.75
Published on Web 02/04/2009

(PPI).[18,19] By neglecting the atomistic details of the components, protein interaction networks are being generated from this data which then can be analyzed with graph algorithms and statistical methods developed for networks.[20,21] In these networks, each protein type typically denotes one node (or vertex), and a known interaction between two proteins is a link. Thus, it is relatively easy to identify a specific pattern corresponding to a certain multiprotein complex. For situations such as the one described above, where one wants to identify a few target complexes in a sea of monomers and partially assembled intermediates, we will show that this task can be conveniently performed by mapping the spatial simulation onto a protein interaction graph and subsequent analysis by well-known efficient tools. This protein interaction network built from the simulation differs from the usual PPI networks, however, because each of its nodes now denotes an individual copy of a protein type. Additionally, the graph is generated dynamically. This means that links appear and disappear over time, as the proteins bind and unbind from each other during the spatial simulation.

Similar considerations arise when, for example, the formation of van der Waals clusters is studied. Here, the objective may be to identify specific "magic" configurations or deviations thereof.[22,23] These magic configurations are characterized by their high degree of symmetry, which in turn leads to easily identifiable patterns in the interaction network.

Here we aim to demonstrate the advantages of combining these two views on particle interactions: the spatial view from the simulation and the network view via the associated graph. For this, a toy system of spherical particles with either isotropic or directed interactions was simulated with a standard Monte-Carlo algorithm and concurrently mapped onto a graph which reflects the current interactions between the particles. To avoid that complexes are essentially fixed in space, the graph was also used to identify the next neighbors of a displaced particle, which are then also displaced by a fraction of this move. This idea can even be taken further so that each complex is treated as one rigid pseudoparticle. This in turn would allow reduction of the number of particles in the simulation and thus speed it up significantly.

Methods

Spatial Simulation. The spatial simulation uses a standard Metropolis Monte-Carlo (MC) approach to model the diffusion of the particles in a stochastic manner. In the canonical *NVT* ensemble the probability p_j for a specific state j of the system is

$$p_j = \frac{e^{-\beta E_j}}{Z} \tag{1}$$

with the energy E_j of this state, the thermal energy $\beta^{-1} = k_B T$ with the Boltzmann constant k_B and the temperature T, and the partition function $Z = \sum e^{-\beta E_j}$. We used a rejection sampling algorithm which creates a new state $j+1$ of the system with the new energy E_{j+1} by displacing one of the particles at a time. In the Metropolis MC scheme, the transition probability into $j+1$ is defined as $a = p_{j+1}/p_j$ which can also be written as

$$a = e^{-\beta(E_{j+1}-E_j)} \tag{2}$$

by inserting eq 1. In the following sections, energies are given in units of the thermal energy $k_B T$, and lengths are given in dimensionless arbitrary units (a.u.).

Interactions and Bonds. Without external forces, the total energy of the simulated system is the sum of all particle−particle interactions. The interaction energy V_W (W for well) between two isotropic spheres of radius σ and center−center distance r was modeled by a repulsive core and an attractive well potential of width $r_C - \sigma$ and depth ε:

$$V_W(r_{ij}) = \begin{cases} \left(100 + 500\dfrac{\sigma}{r_{ij}}\right) k_B T & \text{if } r_{ij} \le \sigma \\ -\varepsilon & \text{if } \sigma < r_{ij} \le r_C \\ 0 & \text{if } r_{ij} > r_C \end{cases} \tag{3}$$

In the examples presented here, the particles had a diameter of $\sigma = 6$ a.u. and $r_C = 7$ a.u. In simulations performed with this interaction, a link between particles i and j was added when $r_{ij} < r_C$.

To model attractive ring patches, the well potential V_W was multiplied with an angle-dependent term as in Wilber et al.:[24]

$$G_{ij}(r_{ij}, \theta_{ij}) = \exp\left[\frac{(\theta_{ij} - \nu)}{2\sigma_{PW}^2}\right] \tag{4}$$

Here, θ_{ij} denotes the angle between the interparticle vector r_{ij} and the vector Ω_i describing the orientation of particle i. ν is the opening angle of the ring, i.e., the angle between Ω_i and the ring, while σ_{PW} describes the angular patch width. In this case, the pairwise potential $V_R(r_{ij})$ between two particles i and j was defined as

$$V_R(r_{ij}) = V_W(r_{ij}) \times G_{ij}(r_{ij}, \theta_{ij}) \times G_{ji}(r_{ji}, \theta_{ji}) \tag{5}$$

In simulations using this ring potential, a link was added in the graph for $V_R(r_{ij}) < 0.4 k_B T$.

Network Measures. The number of individual interactions shorter than r_C equals the number of links in the graph and can be derived from both the spatial and the graph domain. For point particles, the theoretical maximal number of interactions between N particles is $N(N-1)/2$, because interactions are bidirectional and self-interactions are excluded. For particles with nonzero radius that may not overlap, as were used here, the maximal number of interactions is smaller.

A connected component of an undirected graph is a set of vertices that are all reachable from each other. This means that two vertices are in the same connected component if and only if there exists a path between them. The connected components distribution or complex size distribution shows how many complexes of a certain size exist in a graph. The degree distribution $P(k)$ gives the probability that a selected vertex has exactly k links. It is defined as:

$$P(k) = \frac{N(k)}{N} \tag{6}$$

where $N(k)$ is the number of vertices having degree k, and N is the total number of vertices in the graph. The clustering coefficient C_{k_i} of a node i is a measure for the connectivity in the neighborhood of i. It is calculated from the number of links that interconnect the neighbors of i, n_i, and the degree of i, k_i:

$$C_{k_i} = \frac{2n_i}{k_i(k_i - 1)} \tag{7}$$

The function $C(k)$ is defined as the average clustering coefficient of all nodes with degree k. The distance distribution $D(l)$ gives the number of shortest paths[25] of a certain length l.

An overview over commonly used network measures can be found, for example, in a review by Costa et al.[26] The network analysis of the spatial simulations was implemented using the boost graph library.[27]

Simulation Setup. The simulation runs modeled the particle diffusion in a cubic box with periodic boundary conditions. During initialization, each particle was assigned a random position and a random orientation. During one MC iteration, a trial move was generated for one particle after the other, and it was checked whether the new state of the system should be accepted. If the new move was accepted and interactions changed relative to the previous configuration, the interaction graph was updated correspondingly. After all particles had the chance to move once, the interaction graph was analyzed with respect to the properties of the complete graph such as the total number of links or its average connectivity. These measures were used to monitor the convergence of the simulation. Moreover, the connected components were determined, which were individually analyzed for their size and connectivity in a second step. At that time, we also searched for specific patterns which allow identification of, for example, completely assembled icosahedra. Additionally, conventional properties of the simulation were determined as its total energy or the radius of gyration.

Results and Discussion

As introduced before, this study aims at introducing a new way of studying particle interactions in a spatial simulation. The particle simulations themselves are standard MC simulations, and the graph analysis is done by well-established mathematical graph algorithms. The novel aspect introduced in this study is the dynamic updating and subsequent analysis of the interaction graph.

Spherical Particles. A first impression about the monitoring of association processes by graph measures will be given on the example of simulations with 200 particles that used a spherically symmetric attractive well potential of different depths $\varepsilon = 0...-4k_BT$. The simulation box had a volume of $(70 \text{ a.u.})^3$ with periodic boundary conditions.

Commonly, the convergence of simulations is monitored by the time course of the total energy of the system. Here the energy was found to converge exponentially, as complexes formed from the initially randomly distributed particles. Figure 1a shows the equilibrium value of this real space measure vs the well depth ε (filled triangles). As expected, for $\varepsilon < 1k_BT$, the total energy was only slightly smaller than zero, meaning that only a few particles were bound by the weak attractive potential. Around $\varepsilon = 1k_BT$, the total energy strongly decreased, indicating a transition from independent particles to a large cluster. This transition into a single large cluster is also clearly visible in the behavior of the radius of gyration. For small ε, its value corresponds to independent particles distributed over the volume of the simulation box. For $\varepsilon > 2k_BT$ its value of $\approx 5\sigma$ indicates a densely packed cluster which essentially does not get any denser once ε is large enough.

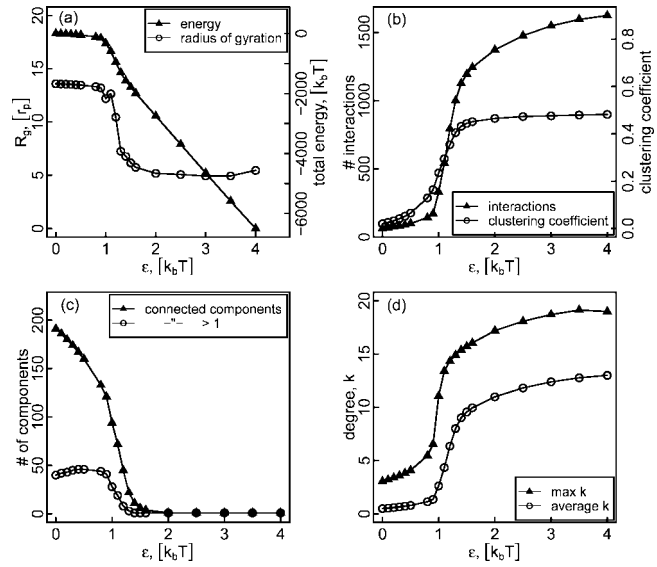

Figure 1. Complex formation monitored via real-space and graph measures. Results from simulations of 200 particles with a spherical attractive well potential for different well depths ε. All graphs show a transition around $\varepsilon = 1k_BT$, above which the particles form a large complex. The data points were connected by lines as guides to the eye. (a) Real space measures total energy (filled triangles) and radius of gyration (circles). (b) Number of interactions, i.e., links in the network derived from the cluster (triangles) and average clustering coefficient (circles) of the complete graph. (c) Total number of connected components (triangles) and number of connected components containing more than one particle (circles). (d) Maximum (triangles) and average (circles) degrees of the nodes of the cluster's graph.

Of course, the graph measures derived from the interaction graph describe the same behavior; it is the very same set of simulations. The measures shown in Figure 1b, the number of interactions and the average clustering coefficient, are the analogues to the total energy and the radius of gyration. The number of interactions again is small for small ε, increases strongly around $\varepsilon = 1k_BT$, and then saturates because the cluster cannot be compacted any further once the most dense packing is reached. The increase of the number of interactions for $\varepsilon > 1.5k_BT$ can be attributed to the behavior of the particles on the surface of the cluster, which try to penetrate deeper into the cluster, thus making the surface smoother. The clustering coefficient also increases around $\varepsilon = 1k_BT$ before it saturates, indicating the dense packing inside the cluster.

Whereas the measures of panels a and b put the emphasis on the central large cluster at $\varepsilon > 1k_BT$, Figures 1c and 1d highlight the behavior of the particles for weak attractions. According to panel c, the total number of connected components decreases monotonically with increasing ε. For $\varepsilon \leq 1k_BT$, the number of connected components of size larger than one stays roughly constant. In this regime the simulation box contains many small complexes with average sizes from two to four particles. At increasing ε they start to collapse into large clusters until for $\varepsilon \geq 2k_BT$, there is only one large component (complex) left including all 200 particles.

This behavior is also reflected by the average and maximal degrees of the particles shown in Figure 1d: a maximal degree of three for small ε indicates a complex of size four. In a densely packed hexagonal three-dimensional cluster, a particle has 12 direct neighbors. With the width of the attractive potential chosen for these simulations, also the second neighbors may be counted as bonded if the cluster is packed densely enough. This leads to the maximal degree of 18 for large ε. Again, the clustering transition can be seen clearly.

These simple graph measures can be computed from the simulations with little additional overhead. When compared to an analysis in coordinate space alone, they readily provide a wealth of insight into the behavior of the particles at weak interactions, about the clustering transition, and about the internal structure of the single complex for strong attraction. Certainly, particles with a spherically isotropic attraction are an extremely simple toy system, which can also be fully understood without the graph analysis. When more structured interactions lead to the formation of specific spatial structures as is the case in the following two examples, the advantages of including the graph analysis will become more obvious.

Particles with Equatorial Ring Patches. When the attraction between the particles was restricted to their equatorial region, membrane-like two-dimensional structures developed as expected. In the following example, an additional attractive well potential similar to eq 3 with $r_C = 4$ a.u. and a depth of $\varepsilon_W = 4k_BT$ was applied between the particles and the bottom wall of the simulation box. The simulation box had a volume of $(100 \text{ a.u.})^3$. Periodic boundary conditions were applied in the other two dimensions, and the top wall of the box opposite of the attractive surface simply reflected the particles. This setup mimics the formation of ordered layers of, for example, proteins on solid surfaces.[28]

Figure 2 shows four snapshots from a spatial simulation with 200 particles. Initially, the particles formed small clusters or adsorbed to the bottom wall (top panels). However, because of construction, the clusters in solution were less stable than those adsorbed to the wall. Consequently, these small clusters eventually fell apart because of thermal fluctuations, and all particles were finally attached to the bottom wall. There, their mutual interaction led to the formation of a hexagonally structured layer. The apparent gap between the two seed clusters seen in the third snapshot at 10^5 iterations was finally filled by the remaining free particles. Although the last snapshot still shows a clear grain boundary between the two domains originating from the seed clusters, this boundary finally disappeared because of reorganization of the particles, and a single, hexagonally ordered layer of adsorbed particles was formed (snapshot not shown).

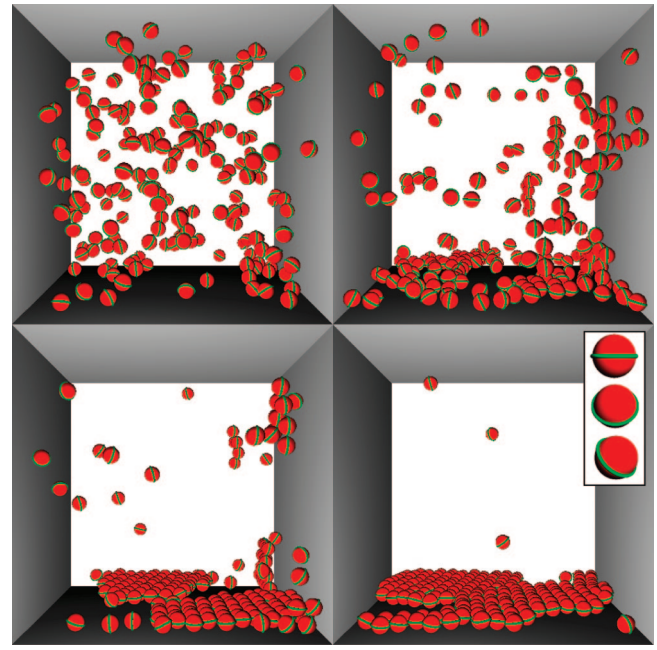

Figure 2. Formation of an ordered layer, visualized in the spatial domain: Snapshots after 10^3, 10^4, 10^5, and 2×10^5 iterations (top left to bottom right) from a simulation of 200 particles with attractive equatorial patches and an adhesive bottom wall of the simulation box. The inset in the last snapshot shows a close-up view of one of the particles in three different orientations. The ring indicates the region where two particles attract each other. According to the given interactions, the particles adsorb to the surface as an ordered layer with a hexagonal structure.

A much richer understanding of the dynamic processes during the formation of the surface layer is obtained when the measures derived from the interaction graph are considered, too. Figure 3a shows the sizes of the connected components that occur during the equilibration phase of the simulation. A dot denotes that at least one cluster of the corresponding size was present at a given iteration. This statistical representation illustrates several trends: (i) from the many small clusters that form in the beginning, a number of clusters of sizes around 30 particles emerged. (ii) Many of these intermediate-size clusters slowly shrank again. (iii) Several clusters constantly grew in size, finally forming one large adsorbate. The stepwise increase of these clusters during the first 10^4 iterations suggests that several seeds quickly associated into a single cluster that then grew by attracting individual particles from the bulk. (iii) The fraction of the small few-particle clusters in the bulk constantly decreased both in cluster size and number.

The hexagonal structure shown in Figure 2d results in a special signature in the degree distribution $P(k)$, as illustrated in Figure 3. Here, $P(k)$ was averaged over the simulation after the equilibration phase, i.e., between iterations 2×10^5 and 10^6. Most of the graph nodes had $k = 6$ as in a two-dimensional hexagonal lattice. The nodes with $k < 6$ are the ones with missing neighbors found on the rims of the adsorbed layer. The few nodes with $k = 7$ are particles in

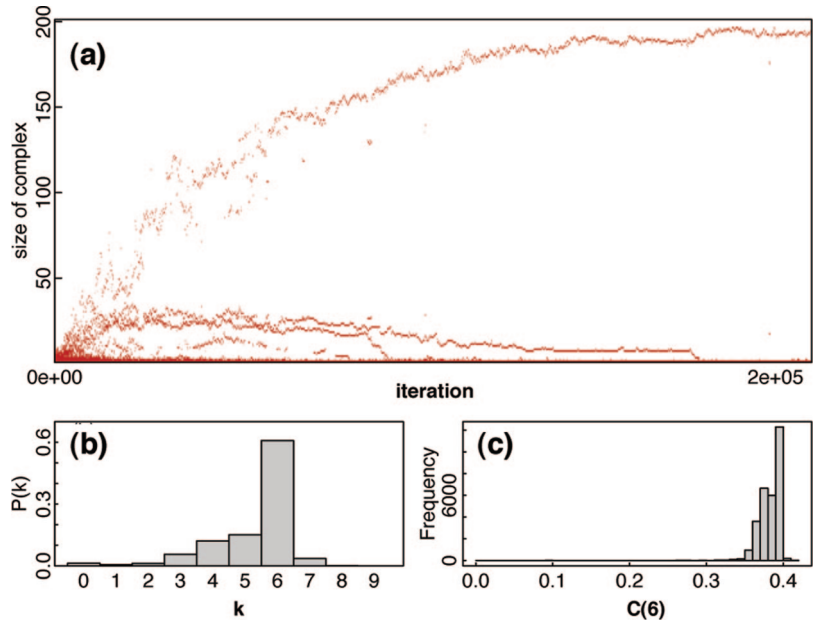

Figure 3. Formation of an ordered layer, visualized in the network domain. (a) Sizes of the connected components that formed during the equilibration phase of the simulation shown in Figure 2. (b) Averaged degree distribution and (c) average clustering coefficient distribution of the particles with $k = 6$ during the steady state after 2×10^5 iterations.

the region of the grain boundary where the hexagonal lattice was disturbed.

The graph analysis also allows focusing on a certain group of particles, as shown for the clustering coefficient of the nodes with $k = 6$ in Figure 3c. In a perfect hexagonal lattice, all nodes have $C(6) = 0.4$. This value also occurred most frequently in the equilibrated particle layer. However, because of the rearrangement dynamics around the grain boundary, some of the six neighbors of a certain particle sometimes separated too far from each other so that no bond was established and the connectivity in this region was reduced. One could now use this information to identify these less interconnected and thus less stable regions of the cluster and examine their dynamics in more detail.

Compared to the first example, the graph analysis in this slightly more complex scenario provides additional insights into the dynamics of the particle adsorption. These would be hard to obtain and even harder to visualize from the spatial view alone. Together with the few snapshots of Figure 2, the cluster size evolution of Figure 3 provides an intuitive visualization of the dynamics of the growth and shrinking stages of the clusters. For example, around iteration 8×10^4, the large cluster, consisting of about 160 particles at that time, temporarily fell apart into a larger and a smaller part of 130 and 30 particles, respectively. Such events had occurred already earlier, too, around iteration 5×10^4. However, here the smaller fragment is not so well visible in Figure 3a among the other smaller clusters.

Icosahedral Complexes. The previous example showed how graph measures can be used to collect and visualize detailed information about the dynamics during structure formation and how regular structures (and irregularities therein) can be easily identified and interpreted. The last application presented here is the formation of icosahedral complexes. In this case, the objective is to identify these highly regular structures among all the particles in the simulation, to monitor their formation, and to assess their quality.

With the opening angle of the attractive ring patch set to 58°, one particle can bind to five other particles, which leads to the formation of icosahedral complexes of twelve particles.[24] A spatial representation of a completely assembled icosahedron and its graph representation are shown in the second row of Figure 4. Each of the particles has exactly five neighbors. Consequently, the degree distribution is zero everywhere except for $P(k = 5) = 1$ (see third column of Figure 4). Among the five neighbors, five of the ten possible links are established, leading to a clustering coefficient of $C(5) = 1/2$ for each of the particles (fourth column). A third signature of an icosahedron is its distribution of path lengths, which consists of 30 paths of lengths one and two, respectively, and six paths of length three that connect the pairs of the diametrically opposite particles (last column).

Any perturbation in the assembly of an icosahedron either changes the number of particles or their connectivity. This in turn is reflected in perturbations of $P(k)$, $C(k)$, and the distribution of path lengths. A correct icosahedron can thus be identified by the correct size and the correct network measures. In contrast to the correctly assembled icosahedron from the second row, the third row of 4 illustrates the signatures of a slightly misfolded icosahedral complex. Here, $P(k)$ has an additional nonzero entry for $k = 4$ and $P(5) < 1$. The $C(k)$ and $D(l)$ distributions are perturbed, too. Because of the partly opened structure, some shortest paths have the length four. These different behaviors allow easy and unambiguous assess to the quality of an icosahedron, even if the spatial representation seems correct. Figure 4 also shows two examples of incorrectly assembled complexes of $N = 8$ and $N = 16$ particles, respectively.

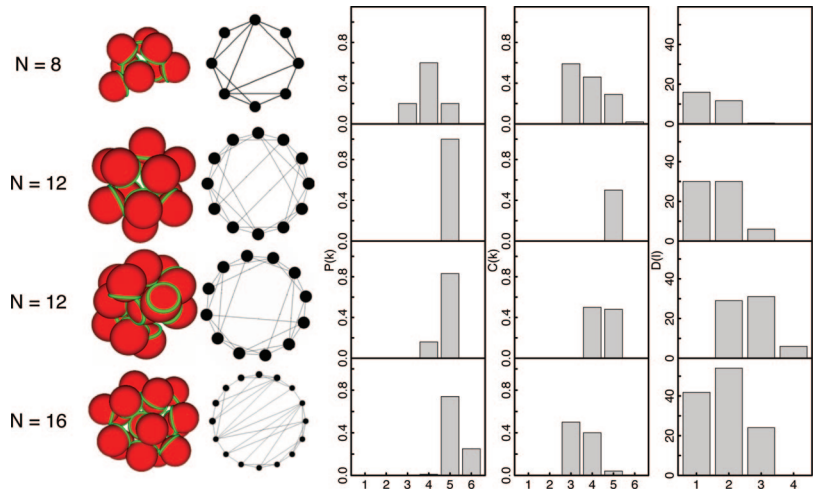

Figure 4. Correctly assembled clusters can be identified via their network signatures. Examples of complexes found in the simulations of icosahedron formation. The spatial representation of the chosen complex is shown in the first column, the corresponding network graph in the second column. The following columns give the degree distribution $P(k)$, the degree-averaged clustering coefficient distribution $C(k)$, and the distribution of path lengths $D(l)$, respectively. The second row belongs to a correctly assembled icosahedron with $N = 12$ particles, while the third row shows a slightly misfolded icosahedron. The first and last rows belong to complexes of $N = 8$ and $N = 16$ particles, respectively.

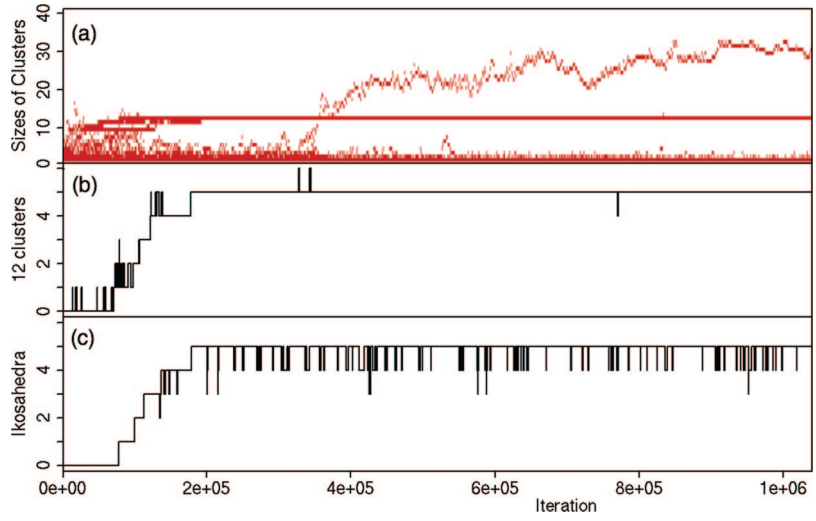

Figure 5. Formation of clusters of size 12 and of icosahedra during a simulation of 100 particles with attractive ring patches with a 58° opening angle. (a) Sizes of the formed clusters, (b) number of clusters of size 12, and (c) number of correctly assembled icosahedra.

With respect to the dynamics of complex formation, the first signature of icosahedron formation during a simulation is the occurrence of connected components of size 12, see Figure 5a. This simulation was performed with 100 particles in a box of size $(70 \text{ a.u.})^3$ with periodic boundary conditions. Obviously, this number of particles would allow for the formation of up to eight complete icosahedra. However, it was unclear whether this maximum number was kinetically accessible on a typical simulation time scale. Here, five to six complexes of size 12 assembled during the first 10^5 iterations, see Figure 5b. From the remaining fragments, a large unstructured complex of about 30 particles formed at a later stage. For each of the complexes of size 12, we computed $P(k)$, $C(k)$, and $D(l)$ and compared them to the distributions expected for a correct icosahedron (cf. Figure 4). This filtering gave four to five correct icosahedra (see Figure 5c). Comparing panels b and c shows that there were essentially five icosahedra present in the simulation run. Because of the random displacements, one of the monomers sometimes left its correct position in the complex by breaking some of the bonds. However, these deviations were not strong enough to break the complex apart completely. Note that the nearly correct icosahedron from Figure 4 (third row) was taken from this simulation.

Using Information from the Graph for Cooperative Moves in the Simulation. A shortcoming of the Monte-Carlo approach used so far is that the particles are moved independently. With the interaction potential used, moving a nonbonded particle does not change the total energy, and this new state is always accepted. Thus, the particles diffusively move through the simulation volume as long as they are not bonded to any other particle. But as soon as they are bonded to one or more of the other particles, any

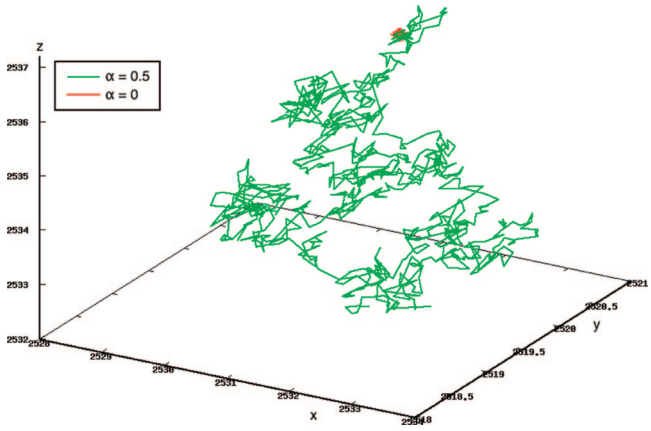

Figure 6. Diffusion of complexes is restored when the Monte-Carlo trial moves are propagated through the network. Trajectory of an icosahedron with (green, $\alpha = 0.5$) and without (red, $\alpha = 0$) coupling of the Monte-Carlo moves to the next neighbors. When only one particle is moved at a time, the cluster itself remains fixed in space, whereas a diffusive movement is observed when a fraction α of the random displacements of the Monte-Carlo algorithm are propagated through the interaction network and the connected next neighbors are displaced, too. The trajectories shown here are representative parts of simulations of twelve particles, initialized to form a correctly assembled icosahedral complex.

displacement of the chosen particle is strongly suppressed, because this would involve the breaking of bonds and the corresponding increase of the energy. Hence, complexes become essentially fixed in space with such a one-particle-at-a-time updating scheme.

This artificial behavior can be elegantly overcome by making use of the information from the graph about the connected components. Our idea was to treat each formed complex as one pseudoparticle that is propagated as a whole. Keeping reality in mind, we suggest that when one particle in a complex is displaced, its neighbors should also be dragged along via the bonds. With the information from the graph, it is now straightforward to not only displace a single particle in the spatial Monte-Carlo algorithm, but to also propagate this displacement along the bonds. Here we implemented a simple variant, where the next graph neighbors of the chosen particle are also displaced by a fraction $\alpha < 1$ of the test move.

Figure 6 shows how such a propagation of the trial moves to the next neighbors with $\alpha = 0.5$ makes a complete icosahedral complex diffuse through the simulation volume (green trajectory). With $\alpha = 0$, i.e., when only a single particle is displaced at a time, the center of mass of the complex only exhibits small fluctuations around its otherwise fixed position (red trajectory). This implementation could be extended by considering not only nearest neighbors but also second-nearest neighbors up to an arbitrary distance.

When cluster formation is considered, the coupling effectively reduces the relative displacements of neighboring particles. This has the same effect as slightly extending the range of the interaction potential. Consequently, with increasing α, clusters form already at smaller potential depths ε (data not shown).

Summary

In summary, we have shown how a dynamically updated graph can be used conveniently to analyze and observe complex formation during a spatial simulation. Since the graph analysis is independent of the simulation, the Monte Carlo approach chosen here for simplicity can be replaced by a different model, as for instance Brownian dynamics or force field simulations. The criterion for placing an interaction surely needs to be adjusted appropriately. Then, for each pair of particles, a separate meaningful definition of an interaction has to be chosen based for example on a maximal relative distance and possibly some angles measuring their relative orientation and/or on a threshold to the interaction energy.

On the graph side, it appears worthwhile to explore the usefulness of more advanced graph measures in the future. For simulations at higher concentrations, for example, the complete complexes start to interfere with each other. These interacting complexes could be distinguished by means of a biconnected or k-connected components analysis as well as a cluster search when a simple analysis of connected components would fail.[26,29] Particles and interactions that connect two clusters can then be identified via their betweenness.[29] For simulations of more advanced particles that form clusters with a specific shape, a motif search as used by Pržulj et al.,[30] Middendorf et al.,[31] or Baskerville et al.[32] could be applied.

The results presented here focused on problems with a known geometry of the correctly formed complex. Moreover, the first example showed that the combination of spatial simulation and network analysis may also be advantageous in situations where the final structure is not known a priori or when, such as in the rather amorphous focal adhesions, a single "correct" configuration does not even exist. Then, the network analysis can be used to identify and to separately monitor the convergence of the "clustered part" of the simulation. Comparing subsequent network snapshots additionally allows us to distinguish the more stable regions of such dynamic assemblies from the more dynamic growth regions.

References

(1) Aloy, P.; Böttcher, B.; Ceulemans, H.; Leutwein, C.; Mellwig, C.; Fischer, S.; Gavin, A.; Bork, P.; Superti-Furga, G.; Serrano, L.; Russell, R. B. *Science* **2004**, *303*, 2026–2029.

(2) Gavin, A. C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumpelfeld, B.; Edelmann, A.; Heurtier, M. A.; Hoffman, V.; Hoefert, C.; Klein, K.; Hudak, M.; Michon, A. M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, B.; Bork, P.; Russell, R. B.; Superti-Furga, G. *Nature* **2006**, *440*, 631–636.

(3) Milne, J. L. S.; Shi, D.; Rosenthal, P. B.; Sunshine, J. S.; Domingo, G. J.; Wu, X.; Brooks, B. R.; Perham, R. N.;

Henderson, R.; Subramaniam, S. *EMBO J.* **2002**, *21*, 5587–5598.

(4) Elcock, A. H.; McCammon, J. A. *Biochemistry* **1996**, *35*, 12652–12658.

(5) Krause, R.; von Mering, C.; Bork, P.; Dandekar, T. *BioEssays* **2004**, *26*, 1333–1343.

(6) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195–2199.

(7) Halperin, I.; Ma, B. Y.; Wolfson, H.; Nussinov, R. *Proteins* **2002**, *47*, 409–443.

(8) Inbar, Y.; Benyamini, H.; Nussinov, R.; Wolfson, H. J. *J. Mol. Biol.* **2005**, *349*, 435–447.

(9) Northrup, S. H.; Boles, J. O.; Reynolds, J. C. L. *Science* **1988**, *241*, 67–70.

(10) Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997**, *72*, 1917–1929.

(11) Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A. *J. Mol. Biol.* **1999**, *291*, 149–162.

(12) Spaar, A.; Dammer, C.; Gabdoulline, R. R.; Wade, R. C.; Helms, V. *Biophys. J.* **2006**, *90*, 1913–1924.

(13) Schluttig, V.; Alamanova, D.; Helms, V.; Schwarz, U. S. *J. Chem. Phys.* **2008**, *129*, 155106.

(14) Berger, B.; Shor, P. W.; Tucker-Kellog, L.; King, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7732–7736.

(15) Hagan, M. F.; Chandler, D. *Biophys. J.* **2006**, *91*, 42–54.

(16) Alber, F.; Dokudovskaya, S.; Veenhoff, L. V.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Rout, M. R.; Sali, A. *Nature* **2007**, *450*, 683–694.

(17) Zlotnick, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5549–15550.

(18) BOND Web Portal. http://bond.unleashedinformatics.com (accessed Dec. 15, 2008).

(19) Xenarios, I.; Salwinski, L.; Duan, X. J.; Higney, P.; Kim, S.; Eisenberg, D. *Nucleic Acids Res.* **2002**, *30*, 303–305.

(20) Albert, R.; Barabási, A. L. *Rev. Mod. Phys.* **2002**, *74*, 47–97.

(21) Barabási, A.-L.; Oltvai, Z. N. *Nature Rev. Gen.* **2004**, *5*, 101–113.

(22) Grigoryan, V. G.; Alamanova, D.; Springborg, M. *Phys. Rev. B* **2006**, *73*, 115415.

(23) Alamanova, D.; Grigoryan, V. G.; Springborg, M. *J. Phys. Chem.* **2007**, *111*, 12577–12587.

(24) Wilber, A. W.; Doye, J. P. K.; Louis, A. A.; Noya, E. G.; Miller, M. A.; Wong, P. *J. Chem. Phys.* **2007**, *127*, 085106.

(25) Dijkstra, E. W. *Numer. Math.* **1959**, *1*, 269–271.

(26) Costa, L. d. F.; Rodrigues, F. A.; Travieso, G.; Boas, P. R. V. *Adv. Phys.* **2007**, *56*, 167–242.

(27) Boost C++ libraries. http://www.boost.org (accessed Dec. 15, 2008).

(28) Quinn, A.; Mantz, H.; Jacobs, K.; Bellion, M.; Santen, L. *Eur. Phys. Lett.* **2008**, *81*, 56003.

(29) Girvan, M.; Newman, M. E. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7821–7826.

(30) Pržulj, N.; Corneil, D. G.; Jurisica, I. *Bioinformatics* **2004**, *20*, 3508–3515.

(31) Middendorf, M.; Ziv, E.; Wiggins, C. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3192–3197.

(32) Baskerville, K.; Grassberger, P.; Paczuski, M. *Phys. Rev. E* **2007**, *76*, 036107.

CT800396V