ASSORTATIVE MIXING IN CLOSE-PACKED SPATIAL NETWORKS

Deniz Turgut, Ali Rana Atilgan, Canan Atilgan

Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey

Corresponding author: Canan ATILGAN; *e-mail*: canan@sabanciuniv.edu; *telephone*: +90 (216) 4839523; *telefax*: +90 (216) 4839550

*All computer programs used in the analyses are available upon request.

Abstract

Background

In recent years, there is aroused interest in expressing complex systems as networks of interacting nodes. Using descriptors from graph theory, it has been possible to classify many diverse systems derived from social and physical sciences alike. In particular, folded proteins as examples of self-assembled complex molecules have also been investigated intensely using these tools. However, we need to develop additional measures to classify different systems, in order to dissect the underlying hierarchy.

Methodology and Principal Findings

In this study, a general analytical relation for the dependence of nearest neighbor degree correlations on degree is derived. Dependence of local clustering on degree is shown to be the sole determining factor of assortative versus disassortative mixing in networks. The characteristics of networks constructed from spatial atomic/molecular systems exemplified by self-organized residue networks built from folded protein structures and block copolymers, atomic clusters and well-compressed polymeric melts are studied. Distributions of statistical properties of the networks are presented. For these densely-packed systems, assortative mixing in the network construction is found to apply, and conditions are derived for a simple linear dependence.

Conclusions

Our analyses (i) reveal patterns that are common to close-packed clusters of atoms/molecules, (ii) identify the type of surface effects prominent in different close-packed systems, and (iii) associate fingerprints that may be used to classify networks with varying types of correlations.

Introduction

The study of real life networks, such as the world-wide web [1], internet [2], power-grids [3] and math co-authorship [4], has put forth properties that distinguish them from classical Erdös-Rényi random networks [5]. The variety of degree distributions and other statistical measures that emerge has heightened the interest in complex networks. With the proposition of algorithms by Watts-Strogatz [3] and Barabási-Albert [6] to generate real life-like networks, this area has been investigated extensively [7,8]. The classification of networks is mostly based on measures such as degree distributions, average clustering, and average path length [9,10]. Recently, spectral properties of networks gained attention since the distribution of eigenvalues characterize several aspects of the network such as algebraic connectivity and bipartiteness [11,12,13]. Although there may be different graphs structures with identical Laplacian spectra that define the network, they often show similar characteristics in terms of network parameters [14]. Several heuristic algorithms are proposed to generate networks from their spectra [15].

In recent years, proteins were investigated as networks, by taking the amino-acids as nodes. Termed as residue networks (RN), edges between neighboring nodes are represented by their bonded and non-bonded interactions [16,17,18,19]. Several studies have shown that residue networks have small-world topology [16,20,21,22], characterized by their logarithmically scaling average path lengths with network size, despite displaying high clustering. Further studies also utilized network models for protein structures to predict hot spots [23,24,25,26], conserved sites [23,24,25,26,27,28,29], domain motions [23,24,25,26,30,31], functional residues [32,33,34,35] and protein-protein interactions [36]. The small-world topology of residue networks is established, and various network properties such as the clustering coefficient, path length, and degree distribution are used to account for, e.g. the different fold-types in proteins [27], interfacial recognition sites of RNA [28], and bridging interactions along the interface of interacting proteins [17]. In light of these studies, we expect other self-organized molecular systems of synthetic origin to display similar topology.

In fact, a hierarchical arrangement of the nodes is expected to occur in self-organization of atoms and molecules under the influence of free energetic driving forces. In graph theory, hierarchies have been quantified by the presence of (dis)assortative mixing of their degrees, defined as nodes with high degrees having a tendency to interact with other nodes of

(low)high degrees [37]. Analytical and computational models for generating assortatively mixed networks were proposed [38,39]. Newman has shown that assortatively mixed networks percolate more easily and they are more robust towards vertex removal [38,40]; most social networks are examples of these. In this work, we find RN of proteins to also have assortative mixing, although many biological networks such as protein-protein interactions and food webs were found to display disassortative behavior.

It is expected that in networks displaying any degree of correlations, local properties of the constructed graphs will have an effect on the global features. However, a connection between the local and global network properties and the underlying structure of molecular systems has yet to be established. In this study, we derive a relationship relating the nearest neighbor degree correlation of nodes, their degree, and clustering coefficient. We next show that a linear relationship is valid for two types of self-organized molecular systems: (i) Folded proteins and (ii) block co-oligomers in a solvent that encourages micelle formation. Furthermore, simulated configurations of Lennard-Jones clusters also approximate the findings as well as a simple polymeric system forced into a close-packed structure under extremely high pressure. We also show that model hexagonal close packed (HCP) structures may be used to reproduce many of the graph properties of the above-mentioned systems. A brief description of the model systems are summarized under the Methods section. This study is a first step towards using statistical characterization in determining the design principles underlying organization of complex molecular networks.

Results

Relationship between k_{nn} and k. We expand on the treatment in ref. [41] to derive a general relationship for the nearest-neighbor degree correlation, k_{nn}, for graphs with non-negligible clustering coefficients, C, defined below.

An un-weighted simple network can be identified fully via the adjacency matrix (**A**), constructed as

$$A_{ij} = \begin{cases} 1 & \text{if nodes } i \text{ and } j \text{ are connected} \\ 0 & \text{if nodes } i \text{ and } j \text{ are not connected} \end{cases} \tag{1}$$

Several parameters are defined to classify networks; each can be computed from the adjacency matrix and are considered as either a local or a global parameter. The simplest parameter is the connectivity, k_i, of node i, also known as the degree;

$$k_i = \sum_{j=1}^{N} A_{ij} \tag{2}$$

Poisson, Gaussian or Power law degree distributions are frequently observed in many real life networks.

Higher order degree correlations are also of importance and may be utilized to identify more distinguishing features of the network. For instance, second degree correlation of a node i, denoted by $k_{nn,i}$, is the average connections of its neighbors and may be written in terms of the adjacency matrix.

$$k_{nn,i} = \sum_{j=1}^{N} \sum_{m=1}^{N} A_{ij} A_{jm} = \sum_{j=1}^{N} A_{ij} k_j \tag{3}$$

$k_{nn,i}$ is also referred to as nearest-neighbor degree correlation. Normalized third degree correlations (C_i), also termed clustering coefficient, is widely used to characterize the distinctness of networks [3,6]. It is defined as the ratio of the number of interconnections between a node's neighbors to the number of all its possible connections, i.e.;

$$C_i = \frac{\frac{1}{2} \sum_{j=1}^{N} \sum_{m=1}^{N} A_{ij} A_{jm} A_{mi}}{\frac{k_i(k_i - 1)}{2}} \tag{4}$$

While k_i, $k_{nn,i}$, and C_i are descriptors of local structure, another common parameter used to classify the global structure of networks is the average shortest path length, L_i of a node. Given that the shortest number of steps to reach node i from node j along the network is L_{ij}, it is the average number of steps that are traversed from all other nodes to node i.

The generating function, $G_0(x)$, for the probability distribution of vertex degrees k is given by,

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k \tag{5}$$

where $|x| \leq 1$, p_k is the probability that a randomly chosen vertex on the graph has degree k, and its distribution is normalized with $G_0(1)=1$. The $G_0(x)$ function generates the probability distribution, capturing all the discrete probability values through the derivatives property,

$$p_k = \frac{1}{k!} \frac{d^k G_0}{dx^k}\bigg|_{x=0} \tag{6}$$

The nth moment of the distribution can thus be calculated from

$$\langle k^n \rangle = \sum_k k^n p_k = \left[\left(x\frac{d}{dx} \right)^n G_0(x) \right]_{x=1} \tag{7}$$

In particular, the average degree of a vertex is $\langle k \rangle = z = \sum_k k p_k = G_0'(1)$. Here the superscript prime denotes differentiation with respect to x.

If one randomly chooses m vertices from a graph, than the powers property of the generating function provides a route to generating the distribution of the sum of the degrees of those vertices by $[G_0(x)]^m$.

We define *outgoing edges* from the first neighbors of a randomly chosen vertex as those connecting to vertices different from the first neighbors of the originally chosen vertex. It is first necessary to define the generating function for the distribution of the degree of the vertices one arrives at, along a randomly chosen edge. That vertex will be reached with probability proportional to its degree, kp_k, so that the normalized distribution is generated by

$$\frac{\sum_k k p_k x^k}{\sum_k k p_k} = x\frac{G_0'(x)}{G_0'(1)} \tag{8}$$

Starting from a randomly chosen vertex and following each of its edges to arrive at the k nearest neighbors, each of the vertices arrived at will have outgoing edges that is given by the degree of that vertex less the edge that one arrives along and the backlinks, b. The latter are defined as the edges that interconnect the nearest neighbors of the original vertex. Thus, the generating function for the outgoing edges from *each* vertex is,

$$G_1(x) = \frac{\sum_k k p_k x^{k-1-b}}{\sum_k k p_k} \tag{9}$$

Note that b itself depends on k.

The number of backlinks, b, is given in terms of the clustering coefficient, C, around a given node with degree k. Using the definition of C, with the number of interconnections, I,

between its first neighbors, $C = I / [k(k-1)/2]$, the average number of backlinks for each of the k neighboring nodes is, $b = 2I / k = C(k-1)$. This will lead to the generating function for outgoing edges as:

$$G_1(x) = \frac{\sum_k k p_k x^{(k-1)(1-C)}}{z} \tag{10}$$

The generating function for the distribution of all outgoing links from the k neighbors of the original node is then obtained from the powers property:

$$G_k(x) = G_1(x)^k = \left[\frac{\sum_k k p_k x^{(k-1)(1-C)}}{z} \right]^k \tag{11}$$

The average number of outgoing links is computed from the first moment of the generating function evaluated at $x = 1$. In general, this leads to

$$G_k^{'}(1) = k\left(\frac{\langle k^2 \rangle}{z} - 1 \right) - \frac{k}{z} \sum_k C k p_k (k-1) \tag{12}$$

k_{nn} is the nearest neighbor correlations, defined as the total number of neighbors of a given node which emanates from a selected node of k neighbors. Thus, it is the sum of the number of outgoing links per neighbor, the backlinks per neighbor and the link that connects the original node to the first neighbor:

$$k_{nn} = \frac{G_k^{'}(1)}{k} + b + 1 = \left\{ \frac{\langle k^2 \rangle}{z} - \frac{1}{z} \sum_k C k p_k (k-1) \right\} + C(k-1) \tag{13}$$

The first term in curly brackets is constant, carrying information on the moments of the distributions, depending on how C is related to k. The second term determines the assortative versus disassortative behavior of the network. For example, if C decreases with k as a single exponential, $C \propto \exp(-ak)$, we may get assortative or disassortative mixing depending on the strength of the decay. For the cases of $C \to 0$, one gets uncorrelated networks. On the other hand, for the particular case of a system where C is finite, yet independent of k, equation 13 reduces to the simple linear expression:

$$k_{nn} = Ck + \frac{\langle k^2 \rangle (1-C)}{z} \tag{14}$$

with slope C and the intercept depending on the degree distribution. For example, for a Poisson distributed network, e.g. approximated by RN constructed from folded protein structures as was shown in [16,17], $p_k = z^k e^{-z}/k!$, the relation takes the form

$$k_{nn} = Ck + (1+z)(1-C) \tag{15}$$

In this work, we study concentrated atomic/molecular systems which have a weak dependence of clustering coefficient on degree. We shall see that the linear dependence of equation 14 suffices to describe their nearest neighbor degree correlations.

In passing, we note that an algorithm for generating networks with given clustering dependence on degree has been proposed [42]. However, the algorithm fixes the average clustering coefficient and has no control over the distribution of clustering for a given degree, while this distribution is crucial in our derivation. Moreover, to construct the desired network, the constraint for networks to be assortatively mixed is imposed.

Statistical properties of close-packed atomic/molecular systems. The nearest neighbor degree correlations are displayed in figure 1 for the five systems studied. We find that all of them display assortative mixing. Furthermore, they are well-approximated by a linear relationship. In fact, one may use equation 14, which was obtained assuming that clustering is independent of degree, to predict the clustering coefficient (from the slope) and the ratio $<k^2>/z$ (from the intercept), to assess the range of validity of this assumption. In Table 1 is a comparative list of the predictions and the actual values calculated for the systems at hand. We find that the predictions overlap with the actual network values for all systems. Since the linear dependence, as well as the match between the predicted values of C and $<k^2>/z$ depend on C being independent of k (see the reduction of eq. 13 to obtain eq. 14), we further examine this property in conjunction with degree distributions (figure 2). For all the systems studied, there is a decreasing trend of C with k, although it is quite weak for RN, micellar networks (MN) and Lennard-Jones clusters (LJC). Taken together with the degree distributions, also displayed in Figure 2 with the gray shaded curves, the variation of C with k is even less significant in the regions within one-standard deviation of the average degree for these three systems. Below we discuss in detail the implication of these observations for the individual systems studied.

Self-organized molecular structures: Residue networks and micellar networks. Previous studies on RN showed that these networks have high clustering as opposed to their random counterparts and have comparable shortest path lengths as the random networks; therefore, they can be considered as having small-world topology [16,20,21,22]. In these studies, comparisons were performed for the average properties throughout the network between the RNs and their randomly rewired counterparts. Although average values do confirm that RNs have small-world properties, detailed analyses of the individual parameters are needed to assess similarity with artificially generated networks.

In reference [16] it was shown that the degree distributions of RN are Poisson; the mean is 6.2. Therein, it was also shown that the residues in the core have a mean clustering coefficient of ca. 1/3, whereas this value approaches 0.5 for the nodes that reside along the surface. Averaged over the set of 595 proteins, the clustering coefficient of RN has the value 0.38. The linearity between k_{nn} and k holds for all sizes of proteins, despite the size differences, in addition to the slight decreasing dependence of C with k. We adopt equation 15 to analyze the relationship between k_{nn} and k in RN and we find that the slope may be identified by the average clustering coefficient of the network. The values of $<C>$ and $z=<k^2>/z$ calculated directly from the network and predicted via equation 14 are listed in Table 1. Within the error bounds, the predictions of theory are valid; the only slight deviation occurs as an underestimation of $<C>$ for the smaller proteins where the surface effects (and the variance in C) are more pronounced. We shall later elaborate further on the surface effects.

We expect other self-organized molecular structures to display network properties similar to the RN obtained from proteins, provided that they are thermodynamically stable and have a given average structure around which fluctuations are observed. Similar to the proteins, these structures follow certain organization rules due to the (in)compatibility of their chemical units with the solvent. Other environmental factors, such as the temperature or the concentration, play a role on the type of organization observed. As example systems, we choose micelles of different morphologies formed by the ABC type co-oligomers, whose coordinates are obtained from dissipative particle dynamics (DPD) simulations, as described in the Methods section.

At low concentrations, these oligomers organize to form spherical micelles. As the concentration increases, adjacent spheres merge and attain a cylindrical morphology. Further increase in the concentration results in the formation of lamellae. As an inset to figure 3, we

display the spherical, cylindrical and the lamellar formations excerpted from oligomer concentrations of $v = 0.3$, 0.6, and 0.9, respectively. Note that it is the core region (i.e. the fluorinated regions shown as white spheres) that maintains the stable morphology, while the corona formed by the red and gray beads shows large fluctuations in conformation. Thus, we use the coordinates of the white blobs to generate the MN. The degree distribution and the dependence of clustering coefficient on degree of a sample network with $v = 0.6$ are shown in figure 2. It is important to note that, regardless of the type of self organization, these network parameters show the same pattern as RN. We approximate their degree by Poisson distribution.

Similar to RN, analysis of k vs. k_{nn} relationship for MN reveals a positive linear correlation regardless of morphology (Figure 1). The values of $<C>$ and $z=<k^2>/z$ calculated directly from the network and predicted via equation 15 are also listed in Table 1. Nodes with less than four and more than 15 connections are omitted due to lack of statistics of blobs with too few or too many neighbors. Theoretical predictions of $z=<k^2>/z$ from the intercept of the k vs. k_{nn} relation is in excellent agreement with the numerical results. The slope of the best-fitting line slightly overestimates the average clustering coefficient.

The linear relationship between k_{nn} and k also predicts the increase in z with size in RN and the decrease in z with concentration (and morphology change) in MN. The theory slightly underestimates the clustering coefficient of RN whereas it overestimates that of MN. This is due to surface effects: In proteins, nodes along the surface have high clustering coefficients as shown in reference [16]. Because these nodes have few links that are interconnected, they increase the average clustering coefficient then would be directly predicted by an overall fit to the data in figure 1. Conversely, in MN surface nodes along the core are connected to the solvo-phillic arms of the chains. These connections, which are omitted in the calculations, since our network construction is based on only the core of the micelles and not the corona, have the reverse effect on the average value of the clustering coefficient.

Effect of excluded volume: Lennard-Jones and HCP clusters. Atoms or groups of atoms occupy a specific volume in space, and as a result, there is an upper bound on the number of neighbors that may be within the direct interaction range of a given node. Since our nodes comprise of coarse-grained groups of atoms that are not arranged spherically symmetric, we observe number of neighbors as large as 19 for a few nodes. This is in contrast to the maximum coordination of 12 expected of regular lattices of spherical particles. All of the

networks studied here have this property of an upper bound on the degree. However, the extent to which this excluded volume effect influences the predictions of the previous subsections is unclear. To further investigate this point, we study LJC, which are clusters of atoms of minimum energy that interact purely via Lennard-Jones interactions. We confine our attention to those within the size range up to 550 particles which is compatible with the network sizes of RN and MN studied here. Although LJC conform to an icosahedral arrangement of atoms, they have incomplete cores (i.e. holes within the structure). We therefore also study hypothetical atomic clusters which have complete occupancy of HCP lattice sites.

The degree distributions of these systems are jagged and cannot be described as Poisson (figure 2). We find a linear relationship between k_{nn} and k, as in the previous self-organized systems (figure 1). For LJC, the dependence of C on k is very similar to those of MN, following a nearly linear trend with a small negative slope (-0.02). For HCP, there is a stronger dependence of C on k, yet for degrees that are observed more frequently, the average clustering remains almost constant (C is 0.36 for $k = 12$ and 0.40 for $k = 9$). In both types of systems, while the $<k^2>/z$ values are well-predicted by equation 14, we find $<C>$ to be consistently underestimated by the theory, more so for LJC than for HCP (Table 1). As discussed in the previous subsection for RN, this is again due to the surface effects, which is more prominent for the irregular surfaces of LJC.

Effect of chain connectivity: Polybutadiene (PBD) Melts. Finally, we study polymeric melts to discern the additional effect of connectivity on the statistical properties of the networks. The linear relationship between k_{nn} and k is also observed for this system which is forced into a close-packed structure by applying very high pressure. Degree distribution deviates from Poisson as for LJC and HCP, while clustering behavior is similar to those obtained for HCP. Both $<C>$ and $<k^2>/z$ are predicted via the theoretical fit (Table 1), with a slight overestimation of $<C>$. The overestimation is due to the fact that we truncate the system at the periodic boundaries of the cubic simulation box, and therefore the neighbors of some of the surface beads are artificially eliminated. Similar overestimation was also obtained for MN, where the corona neighbors of the core beads were removed. Thus, the effect of chain connectivity only plays a role in defining a correct neighborhood structure for the surface beads.

Putting together these results, we conclude that the excluded volume leads to the assortative

mixing of the local structure, described by the positive slope of between k_{nn} and k curves. Furthermore, the extrapolation of the curves to low connectivity ($k \to 0$) leads to an excellent prediction of the $<k^2>/z$ values, regardless of the type of system studied (figure 4). Additional constraints on the local organization of the beads would lead to further local structuring which is measurable by the slope of these curves converging to $<C>$. We find that chain connectivity alone does not bring about such local organization of the beads as observed for PBD system at moderate density (data not shown). However, systems attaining dense core structures do converge to this limit. Such close-packing may be attained by imposing external factors such as the high pressure on PBD; alternatively, the core regions of self-organized systems prefer to realize such an arrangement due to the free energetic requirements of arranging chains with both solvo-phobic and solvo-phillic regions in a solvent that creates the driving force for the formation of the densely packed core [43].

Discussion

This study is based on the premise that network structures are better classified by the distributions of their network parameters rather than the average values. One previous example has been with approximating residue networks derived from proteins with the regular ring lattice: Although it is relatively easy to generate a corresponding ring lattice with few random rewired links having the same average degree and clustering coefficient as the RN [16], neither the second degree correlations nor the global properties (e.g. average path length) are reproduced with this approach. However, comparison of distributions of the parameters involved is not straightforward.

To make the problem tractable, we derive a relationship between k_{nn} and k for networks with arbitrary degree distributions, but with narrowly distributed finite clustering. This subset of constraints is relevant to the study of complex systems, because the results directly apply to the study of self-organized molecular structures which are characterized by Poisson degree distributions, and narrowly distributed clustering coefficients. In randomly-packed chain systems this relationship is expected to be lost, as is observed when the corona region of the micellar networks (i.e. the disorganized parts of the chains protruding into the solvent) is also included in the calculations (data not shown). We validate the derived linear relationship between k_{nn} and k on several model networks based on three dimensional regular structures, polymeric melts forced into close-packing by external pressure as well as those constructed from proteins and micelles of self-organizing co-oligomers.

Excluded volume and close-packing together control the plateau value of the clustering coefficient reached for nodes which are located in the core of the systems studied; i.e. those with high degree. Moreover, they impose a decreasing trend on C with increasing k, as well as providing restrictions on degree distributions. These constraints lead to assortative mixing in the graph structure. The presence of a single chain (as in RN), many chains (as in MN and PBD) or no chains (as in LJC and HCP) does not have an effect on these trends.

The close packed structures emerge as model systems that approximate the network properties of self-organized molecular structures: They yield the local statistical averages and distributions similar to that of the self-assembled systems. Using these model networks as the basis, one may generate novel networks by introducing a few random links whereby the local properties are preserved while the desired global properties are approximated. The ultimate goal is to use both statistical and spectral characterization to design networks with desired properties and to determine the principles underlying organization of complex networks.

Methods: Model Systems and Network Construction

Self-organized molecular structures. In this subsection we describe how the networks are constructed for the two self-organized molecular structures studied in this work.

Residue Networks: These networks are formed from experimentally determined protein structures obtained from the Protein Data Bank (PDB) [44]. For the RN calculations we utilize a set of 595 single-chain proteins with sizes between 54-1021 and having a sequence homology less than %25 [45]. This protein set is identical to the set we used in our previous studies [16,17] and is listed as a supplementary file in [17].

Given a protein, each amino-acid is represented by a node that is centered at the position of C_β atoms, or the C_α atom in the case of Glycine. Edges are added between two nodes (i.e. $A_{i,j} = 1$ in equation 1), if they are closer than a selected cutoff distance, r_c. We call these constructions RN. We use r_c=6.7Å as in our previous work, which is the distance where the first coordination shells ends, as computed from the radial distribution function (RDF) shown in figure 3. See references [16,17,46] for more details on the construction of residue networks and the choice of r_c .

Micellar Networks. Unlike proteins, there is no experimentally available atomistic structure data for self-organized synthetic molecules. We therefore generate such data using the coarse grained simulation methodology DPD. In DPD, the equilibrium morphology of a group of

beads is obtained by integrating out the fast motion of atoms. In addition to the random and dissipative forces, the net forces on the beads are soft and repulsive conservative forces. The simulation is carried out by integrating Newton's law of motion. DPD simulations allow reaching long time scales for macromolecular systems. Thus, morphologies of self-organized systems of large sizes can be studied. Here, we simulate the micelle formation by ABC type oligomers of styrene-co-perfluoroalkylethylacrylate in tetrahydrofuran (F beads). The co-oligomer consists of ten styrene monomers (A beads), seven perfluoroheptane monomers (C beads) and a linker monomer (B bead). The styrene monomers in the co-oligomer have a tendency to interact with the solvent, whereas the fluorinated parts prefer to segregate, thus resulting in micelle formation. The equilibrium morphology depends on the concentration of oligomer in the solution [47]. Force on bead i is given by $\mathbf{f}_i = \sum_{j \neq i} (\mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R) + \sum_k \mathbf{F}_{ij}^{conn}$, where the respective forces are due to interaction, dissipative and random forces between beads i and j, and chain connectivity between bead i, its neighbors k along the chain contour. A general overview of the DPD method and parameterization details for this particular system is given in [48].

We report results from systems where the volume fraction, v, of the oligomers is 0.3, 0.6 and 0.9, respectively. We find that at these concentrations, the triblock co-oligomers self-organize into spherical, cylindrical and lamellar morphologies respectively, as the concentration is increased. Once the organized structures are obtained, we focus on one substructure from the simulated system; e.g. the set of oligomers that form a complete sphere are taken as the structure whose network will be formed. Thus, the spherical structure is made up of 50 chains, the cylindrical structure has 100 chains, and the lamellar structure has 150 chains. In each sample structure, we concentrate on the fluorinated segments, which have self-organized due to the driving forces inherent to the system beads. By computing the RDFs around these beads, we find that the first coordination shell ends at 1.1 DPD units (see figure 3). We use this cutoff distance to form the network (equation 1) whose properties are studied. Chain connectivity of a copolymer is preserved regardless of the particle separation; i.e. (i, i+1) connections are always present. Also shown as an inset to figure 3 are sample configurations of spherical, cylindrical and lamellar formations excerpted from oligomer concentrations of v = 0.3, 0.6, and 0.9, respectively.

Other atomic/molecular structures. We also study other densely packed systems of atomic/molecular origin, to investigate the effects of excluded volume and chain connectivity

on the observed statistical properties. To this end, we focus on the structure of networks obtained from Lennard-Jones clusters and clusters imposed on HCP lattices (to test influence of excluded volume on the results) as well as polybutadiene melts (to test the combined effect of excluded volume and chain connectivity). The network data are obtained as described below.

Lennard-Jones Clusters. The structure of clusters of atoms is an area of intense scientific research, since the properties of materials become size dependent when systems are small enough. By clusters, we refer to groups of atoms from tens to thousands of atoms. LJC are a group of atoms that contain purely Lennard-Jones interactions between pairs of atoms. Geometric optimization of these clusters requires developing efficient search algorithms, since the conformational space available to a cluster of atoms increases explosively. The atomic coordinates of LJC for sizes 3-1000 are deposited on the Cambridge Cluster Database [49]. Many of them are described by icosahedral motifs with an incomplete core [50]. Here we examine clusters of sizes $350 - 550$, in intervals of 50 atoms. The cutoff distance for adjacency matrix construction is 1.6 Å [51]; see figure 3 for the RDF.

Hexagonal Close Packed lattice based atomic clusters: We pack a set of N-atoms (nodes) on the lattice sites so that we have a finite system that has all lattice sites filled, unlike LJC that have incomplete cores. We emphasize that, we have studied the properties of simple cubic, body-centered cubic, face-centered cubic and HCP arrangements, although here we present representative data from the latter only, as all these systems lead to similar conclusions. In the HCP structure, nodes are arranged on a plane in a hexagonal formation, and planes are stacked on top of each other with alternating order. Although we display the RDF of this system in figure 3, we do not choose a cutoff distance where the first coordination shell ends, but we rather connect the first nearest neighbors to obtain the network; the fixed cutoff value is marked on the figure with the vertical dashed line. The generating function (equation 5) for $N = 500$ sites is

$$G_0(x) = 0.004x^3 + 0.032x^4 + 0.076x^5 + 0.028x^6 + 0.060x^7 + 0.080x^8 + 0.224x^9 + 0.048x^{10} + 0.064x^{11} + 0.0384x^{12}.$$

Polybutadiene Melts. We investigate networks constructed from PBD melts that have been obtained from molecular dynamics (MD) simulations. The system consists of monodisperse *cis*-1,4-PB of 32-chains, each with 32 repeat units (C_{128}). The initial coordinates of the system studied was prepared in Amorphous Construction Module of the Accelerys Material Studio

4.4 [52] at a density of 0.92 gr/cm^3, which occupies a cubic box of 47 Å on each side. Minimization, pre-equilibration and integration of the equations of motions were done with the NAMD program [53]. The interaction potentials for PBD chains reported in [54] are adopted. For all simulations, 1 fs integration time step was used. Temperature and pressure were maintained constant in the MD simulations at their prescribed values by employing the Langevin thermostat-barostat. For the non-bonding interaction cut-off distance of 10 Å was used with a switching function turned-on at 8 Å.

To obtain well-equilibrated samples of PBD chains with correct chain statistics, the initial structure which is energy minimized for 10000 steps is depressurized by placing the chains into a larger cubic box of 300 Å on each side. NVT simulations of this low-density system is carried out for 10 ns at 430 K. We then cool the system to 300 K by equilibrating for an additional 20 ns. Consequently, we compress it with NPT simulations at 1 atm at 430 K for 1 ns. We check that the conformational properties (as measured by the characteristic ratio) and the thermodynamic measurable (e.g. thermal expansion coefficient and compressibility) are compatible with the values in reference [54]. The data used in the current calculations are finally obtained from highly pressurized PBD melts via NPT simulations at 100 GPa and 430 K. We collect data for 50 ns. PBD melts are coarse grained by using the coordinates for the center of mass of carbon atoms in the butadiene repeat units. RDFs are obtained as usual, and cutoff distance for network construction is chosen at 5 Å, the ending point of the first coordination shell (figure 3).

References

1. Huberman BA, Adamic LA (1999) Internet - Growth dynamics of the World-Wide Web. Nature 401: 131-131.
2. Vazquez A, Pastor-Satorras R, Vespignani A (2002) Large-scale topological and dynamical properties of the Internet. Physical Review E 65: 066130.
3. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393: 440-442.
4. Barabasi AL, Jeong H, Neda Z, Ravasz E, Schubert A, et al. (2002) Evolution of the social network of scientific collaborations. Physica A 311: 590-614.
5. Erdös P, Rényi A (1961) On the evolution of random graphs. Bulletin of the Institute of International Statistics.
6. Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509-512.
7. Kleinberg J (2000) The small-world phenomenon: an algorithm perspective. Proceedings of the thirty-second annual ACM symposium on Theory of computing. Portland, Oregon, United States: ACM. pp. 163-170.
8. Paperin G, Green D, Leishman T (2008) Dual Phase Evolution and Self-organisation in Networks. pp. 575-584.
9. Albert R, Barabasi AL (2002) Statistical mechanics of complex networks. Reviews of Modern Physics 74: 47-97.
10. Newman MEJ (2003) The structure and function of complex networks. Siam Review 45: 167-256.
11. Bollábas B (1998) Modern graph theory: Springer.
12. Godsil C, Royle G (2001) Algebraic graph theory: Springer.
13. Mohar B (1991) The Laplacian spectrum of graphs. Graph Theory, Combinatorics, and Applications 2.
14. Chung F (1997) Spectral graph theory: AMS.
15. Ipsen M, Mikhailov AS (2002) Evolutionary reconstruction of networks. Physical Review E 66: 046109.
16. Atilgan AR, Akan P, Baysal C (2004) Small-world communication of residues and significance for protein dynamics. Biophysical Journal 86: 85-91.
17. Atilgan AR, Turgut D, Atilgan C (2007) Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication. Biophysical Journal 92: 3052-3062.
18. Brinda KV, Vishveshwara S (2005) A network representation of protein structures: Implications for protein stability. Biophysical Journal 89: 4159-4170.
19. Sathyapriya R, Vishveshwara S (2007) Structure networks of E-coli glutaminyl-tRNA synthetase: Effects of ligand binding. Proteins-Structure Function and Bioinformatics 68: 541-550.
20. Bagler G, Sinha S (2005) Network properties of protein structures. Physica a-Statistical Mechanics and Its Applications 346: 27-33.
21. Greene LH, Higman VA (2003) Uncovering network systems within protein structures. Journal of Molecular Biology 334: 781-791.
22. Vendruscolo M, Dokholyan NV, Paci E, Karplus M (2002) Small-world view of the amino acids that play a key role in protein folding. Physical Review E 65: 061910.
23. Bahar I, Atilgan AR, Demirel MC, Erman B (1998) Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability. Physical Review Letters 80: 2733-2736.

24. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Folding & Design 2: 173-181.

25. Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG (1999) Collective motions in HIV-1 reverse transcriptase: Examination of flexibility and enzyme function. Journal of Molecular Biology 285: 1023-1037.

26. Demirel MC, Atilgan AR, Jernigan RL, Erman B, Bahar I (1998) Identification of kinetically hot residues in proteins. Protein Science 7: 2522-2532.

27. Higman VA, Greene LH (2006) Elucidation of conserved long-range interaction networks in proteins and their significance in determining protein topology. Physica a-Statistical Mechanics and Its Applications 368: 595-606.

28. Lee CY, Lee JC, Gutell RR (2007) Networks of interactions in the secondary and tertiary structure of ribosomal RNA. Physica a-Statistical Mechanics and Its Applications 386: 564-572.

29. Muppirala UK, Li ZJ (2006) A simple approach for protein structure discrimination based on the network pattern of conserved hydrophobic residues. Protein Engineering Design & Selection 19: 265-275.

30. Aftabuddin M, Kundu S (2007) Hydrophobic, hydrophilic, and charged amino acid networks within protein. Biophysical Journal 93: 225-231.

31. Bode C, Kovacs IA, Szalay MS, Palotai R, Korcsmaros T, et al. (2007) Network analysis of protein dynamics. Febs Letters 581: 2776-2782.

32. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, et al. (2004) Network analysis of protein structures identifies functional residues. Journal of Molecular Biology 344: 1135-1146.

33. Baysal C, Atilgan AR (2001) Coordination topology and stability for the native and binding conformers of chymotrypsin inhibitor 2. Proteins-Structure Function and Genetics 45: 62-70.

34. del Sol A, Fujihashi H, Amoros D, Nussinov R (2006) Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families. Protein Science 15: 2120-2128.

35. del Sol A, Fujihashi H, Amoros D, Nussinov R (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. Molecular Systems Biology: 000243245400019.

36. Brinda KV, Vishveshwara S (2005) Oligomeric protein structure networks: insights into protein-protein interactions. Bmc Bioinformatics 6: 296.

37. Pastor-Satorras R, Vazquez A, Vespignani A (2001) Dynamical and correlation properties of the Internet. Physical Review Letters 87: 258701.

38. Newman MEJ (2003) Mixing patterns in networks. Physical Review E 67: 026126.

39. Xulvi-Brunet R, Sokolov IM (2004) Reshuffling scale-free networks: From random to assortative. Physical Review E 70: 066102.

40. Newman MEJ (2002) Assortative mixing in networks. Physical Review Letters 89: 208701.

41. Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. Physical Review E 6402: 026118.

42. Serrano Á, Boguñá M (2005) Tuning clustering in random networks with arbitrary degree distributions. Physical Review E 72: 036133.

43. Can H, Kacar G, Atilgan C (2009) Surfactant formation efficiency of fluorocarbon-hydrocarbon oligomers in supercritical CO2. Journal of Chemical Physics 131: 124701.

44. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, et al. (2002) The Protein Data Bank. Acta Crystallographica Section D-Biological Crystallography 58: 899-907.

45. Fariselli P, Casadio R (1999) A neural network based predictor of residue contacts in proteins. Protein Engineering 12: 15-21.

46. Atilgan C, Okan OB, Atilgan AR Orientational Order Governs Collectivity of Folded Proteins. Proteins: Structure, Function, Bioinformatics 78: 3363-3375.

47. Ozen AS, Sen U, Atilgan C (2006) Complete mapping of the morphologies of some linear and graft fluorinated co-oligomers in an aprotic solvent by dissipative particle dynamics. Journal of Chemical Physics 124: 064905.

48. Kacar G, Atilgan C, Ozen AS (2010) Mapping and Reverse-Mapping of the Morphologies for a Molecular Understanding of the Self-Assembly of Fluorinated Block Copolymers. Journal of Physical Chemistry C 114: 370-382.

49. D. J. Wales JPKD, A. Dullweber, M. P. Hodges, F. Y. Naumkin F. Calvo, J. Hernández-Rojas and T. F. Middleton The Cambridge Cluster Database.

50. Xiang YH, Jiang HY, Cai WS, Shao XG (2004) An efficient method based on lattice construction and the genetic algorithm for optimization of large Lennard-Jones clusters. Journal of Physical Chemistry A 108: 3586-3592.

51. Xiang YH, Cheng LJ, Cai WS, Shao XG (2004) Structural distribution of Lennard-Jones clusters containing 562 to 1000 atoms. Journal of Physical Chemistry A 108: 9516-9520.

52. Accelrys Inc. (2008) Materials Studio, Release 4.4. San Diego: Accelrys Inc.

53. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 26: 1781-1802.

54. Tsolou G, Harmandaris VA, Mavrantzas VG (2006) Temperature and pressure effects on local structure and chain packing in cis-1,4-polybutadiene from detailed molecular dynamics simulations. Macromolecular Theory and Simulations 15: 381-393.

Figure Captions:

Figure 1. Averaged k_{nn} vs. k plots for RN with $N = 190\text{-}210$ (29 proteins), MN with $v = 0.60$ (cylindrical micelle is formed in the core), LJC ($N = 500$), HCP ($N = 500$), and PBD systems. Using equation 14, the values for C and z are predicted and compared with the actual values of the network in Table 1. For RN, nodes with degree 1, 13, 14 and 15 are omitted since there is relatively small number of nodes with such degrees (< 25) to provide meaningful statistics. Similarly, for MN, nodes with degree less than 5 and greater than 15 are omitted to provide meaningful statistics.

Figure 2. Averaged clustering vs. degree plots for RN ($N = 190\text{-}210$), MN ($v = 0.60$), LJC ($N = 500$), HCP ($N = 500$), and PBD on the left y-axis. Degree distributions are superposed (shaded) and labeled on the right y-axis.

Figure 3. Radial distribution function $g(r)$ calculated for sample systems in the current work. Distance r is in Å for RN, PBD and LJC structures, and is in reduced units (bead size $= 1$ unit) for the other cases. The cutoff distances, r_c, utilized for network construction are also marked on the figures. An example network construction is displayed for the residue network (RN) of the sample protein (PDB code **1esl**) as an inset; protein structure in ribbon diagram is on the left, the constructed network at the r_c value selected for all residue networks is on the right. Also shown as inset are the MN structures formed at various concentrations ($v = 0.30$, spherical; $v = 0.60$ cylindrical; $v = 0.90$, lamellar).

Figure 4. Comparison of predicted versus calculated values of the ratio of second to first moments of the degree distributions, $<k^2>/z$.

Table 1. Network parameters $\langle C \rangle$ and $\langle k^2 \rangle/z$ computed from the generated graphs and predicted from the least squares linear fit to k_{nn} vs. k curves.

		Calculated		Predicted[c]	
		$\langle C \rangle$	$\langle k^2 \rangle/z$	$\langle C \rangle$	$\langle k^2 \rangle/z$
Residue Networks[a]	595 Proteins; $\langle N \rangle = 254$	0.38	6.2	0.35±0.01	5.8±0.2
	$N = 140\text{-}160$	0.38	6.1	0.32±0.01	5.7±0.2
	$N = 190\text{-}210$	0.39	6.2	0.32±0.02	5.8±0.4
	$N = 290\text{-}310$	0.37	6.6	0.36±0.01	6.2±0.2
Micellar Networks[a]	$v = 0.3$	0.45	10.3	0.40±0.02	10.5±0.8
	$v = 0.6$	0.43	9.9	0.51±0.02	10.2±0.8
	$v = 0.9$	0.41	9.4	0.51±0.02	9.6±0.6
Lennard-Jones Clusters[b]	$N = 350$	0.47	15.1	0.33±0.07	14.4±1.4
	$N = 400$	0.47	15.3	0.31±0.06	14.5±1.1
	$N = 450$	0.46	15.4	0.33±0.07	14.6±1.3
	$N = 500$	0.46	15.5	0.33±0.07	14.6±1.4
	$N = 550$	0.47	15.6	0.37±0.12	15.3±2.6
HCP[b]	$N = 500$	0.41	10.2	0.38±0.06	9.9±0.8
PBD[b]	$T = 430$ K, $P = 100$ GPa	0.45	12.8	0.52±0.03	12.4±0.7

[a] Degree distribution is well-described by Poisson; therefore predictions by eq. 14 and 15 lead to the same result. $z = \langle k \rangle = \langle k^2 \rangle/z$ for these systems.
[b] Degree distributions are not well-described by Poisson. Predictions are made through eq. 14.
[c] Error margins on the predicted values are reported.