Supporting Information for:

Water-mediated long-range interactions between internal vibrations of remote proteins.

Anna Kuffel and Jan Zielkiewicz

Department of Chemistry, Gdansk University of Technology

Narutowicza 11/12, 80–233 Gdansk, Poland

e-mail: anna.kuffel@pg.gda.pl (A.K.), jaz@chem.pg.gda.pl (J.Z.)

1. Methods

Preparation of the systems. We investigated three systems composed of the catalytic domain of kinesin and a tubulin dimer. The proteins were separated by three different distances (0.8, 1.2 and 2.0 nm) and immersed in SPC/E water. Three structures from Protein Data Bank were used (PDB ID: 1BG2,[1] 1JFF,[2] 2P4N[3]). The 1BG2 structure is the kinesin catalytic domain, the 1JFF structure is a tubulin dimer, and the 2P4N structure consists of the kinesin catalytic domain resting on a tubulin dimer. Kinesin from the 2P4N structure was moved above the tubulin dimer by 0.8, 1.2 and 2.0 nm. The previously equilibrated structures made of 1BG2 (kinesin) and 1JFF (tubulin) were fitted to the coordinates of kinesin and tubulin separated by these distances. The three produced systems were solvated with SPC/E water, with a minimal distance between the proteins and the walls of octahedral boxes equal to about 1.2 nm. Also, two systems consisting solely of the kinesin catalytic domain in water and the tubulin dimer in water were constructed for comparative purposes. More details concerning the preparation of the aforementioned systems can be found in our previous paper.[4]

Computer simulations. Computer simulations were carried out using molecular dynamics package Amber12[5] and ff03 force field, which is appropriate for proteins.[6] The particle-mesh Ewald method was used for electrostatic interactions, and the lengths of chemical bonds involving hydrogen atoms were fixed using SHAKE; a 1.2 nm cutoff for nonbonding interactions was used.

Every system consisting of kinesin, the tubulin dimer and SPC/E water (or single kinesin molecule and SPC/E water) was initially equilibrated for at least 2 ns under NPT conditions, i.e. the temperature (298 K) was kept constant by the weak coupling to an external bath (τ_T = 1.0 ps), and the pressure (1 bar) was kept constant by using the weak coupling method (τ_p = 1.0 ps). During equilibration, the positions of α-carbon atoms of kinesin and tubulin were restrained by using harmonic forces, with the force constant k = 2 kcal·Å^{-2}·mol^{-1}, to prevent the proteins from displacing and changing the distance between them. Next, many short simulation runs (100 ps long) were performed under NPT conditions (298 K and 1 bar) and by restraining the positions of all atoms of kinesin and tubulin with the force constant k = 0.5 kcal·Å^{-2}·mol^{-1}. The atom velocities for each of these runs were independently generated from Maxwell distribution (at 298 K), using different seed values of a pseudo-random number generator.[5] These initial NPT runs were followed by 140 ps-long simulations under NVE conditions (and without any restrains). The terminal 120 ps of these runs were used to perform cross-correlation analysis and to calculate the covariance matrix. The total analyzed time period (the pooled times of all runs) equaled 40 ÷ 50 ns. The described procedure ensured that the proteins did not significantly change their mutual distance and orientation. Of course, we did not want to use any restraining forces during the production run because they could affect the protein dynamics. Trajectories for cross-correlation analysis were saved after each step, namely, every 2 fs, while in the case of the calculation of the covariance matrix, every 10 fs.

Definition of the surface atoms of proteins. Because of the irregular shape of used proteins, their surface atoms were selected in a two-step procedure. Firstly, water molecules embedded within a cylinder with a diameter of 2.2 nm were chosen. The cylinder axis was connecting kinesin and the tubulin dimer (see Figure 1 in our previous paper[4]). Secondly, the surface atoms of proteins were picked for which the atom-water distance did not exceed 0.6 nm (excluding hydrogen atoms of these proteins).

Modified water models. The SPC/E water model were modified by using the idea described by Sorin *et al.*[7] According to this concept, both ε and σ Lennard-Jones parameters of oxygen atom were modified, keeping the density of liquid water unchanged (the geometry of a water molecule and the partial charges of all atoms also remained unaltered). The selected physical properties of these models, including SPC/E, are listed in Table 1. The hydrogen bond definition (required for the calculation of n_{HB} and E_{HB} values) is the same as in our previous papers.[4]

2. Dependence of the phase spectra and amplitude spectra on the value of σ for proteins separated by three different distances

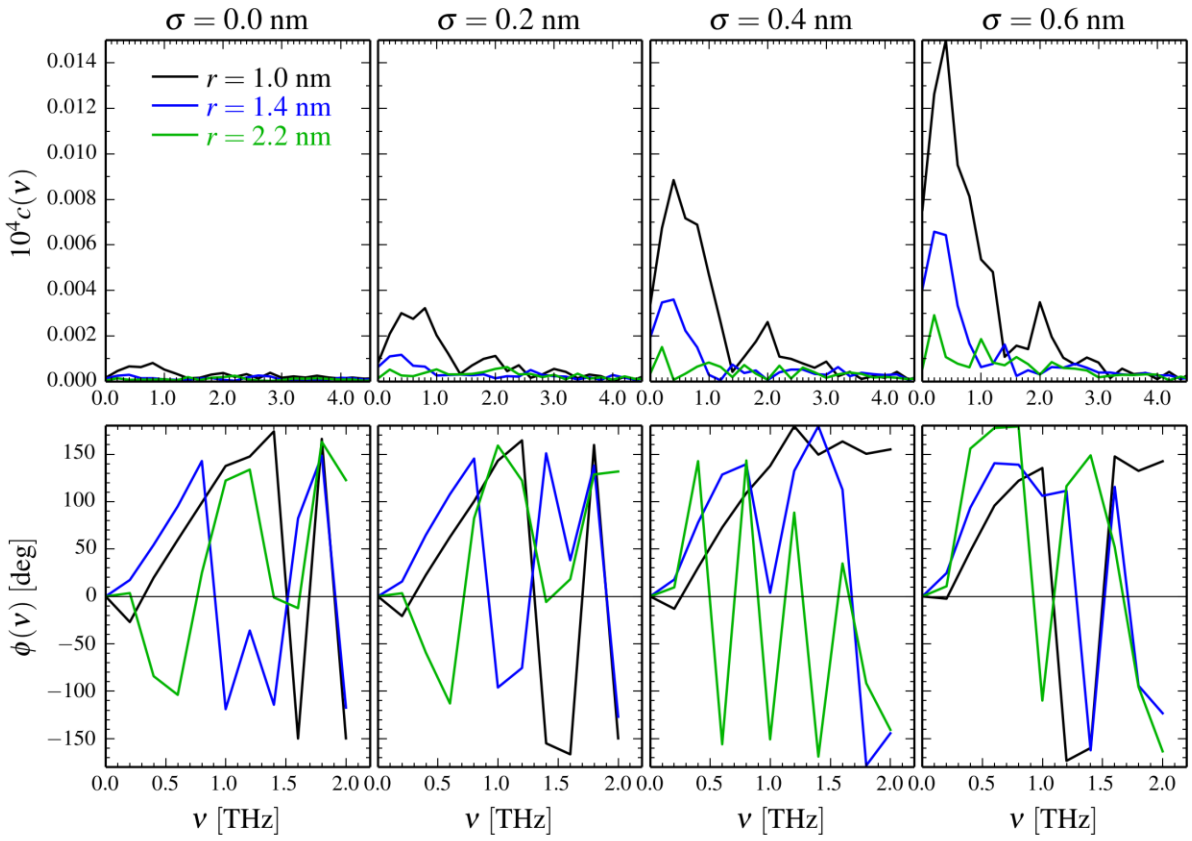

Figure S1. Amplitude spectra (top) and phase spectra (bottom) of the velocity correlation function of the kinesin and the tubulin surface atoms calculated according to equations (1), (2) and (3) in the main text, for various values of σ (0 nm, 0.2 nm, 0.4 nm, 0.6 nm). Proteins are separated by 0.8 nm (the maximum of the distribution function at r=1.0 nm), 1.2 nm (the maximum of the distribution function at r=1.4 nm) and 2.0 nm (the maximum of the distribution function at r=2.2 nm) and solvated with SPC/E water model.

3. Distance distribution functions

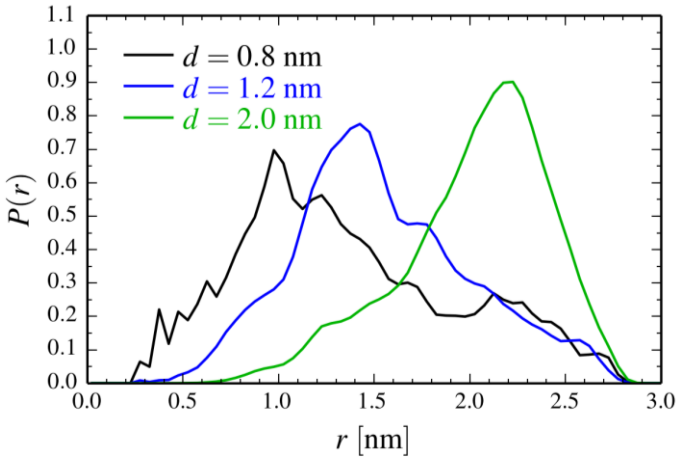

Figure S2. Distribution functions of distances between the selected surface atoms of the tubulin dimer and the nearest kinesin atoms (for the three analyzed systems with the proteins separated by 0.8 nm, 1.2 nm and 2.0 nm).

4. Principal component analysis (PCA)

One of the methods used to obtain valuable information on the dynamics of proteins is principal component analysis (PCA).[8] The assumption is that the analysis can be limited to a relatively small number of principal components, what noticeably reduces the complexity of the problem. As can be found in literature,[9-11] no more than several dozen principal components are often sufficient to describe the most characteristic aspects of the dynamics of proteins.

To calculate the principal components, a covariance matrix of internal atomic displacements has to be calculated (the coordinates of each atom were weighted by its mass):

$$\sigma_{ij} = \overline{(\xi_i - \bar{\xi}_i)(\xi_j - \bar{\xi}_j)}, \qquad \text{(e1)}$$

where ξ_i (i=1,2,…,3N) denote the mass-weighted Cartesian coordinate of atom i (total number of analyzed atoms is N).

First, the mean mass-weighted position of every one of all heavy atoms of the protein (that is: besides hydrogen atoms) has to be determined. These mean coordinates were calculated after removing the translational and rotational motion of the molecule as a whole. It was done by fitting the protein molecule to the indicated reference structure. We are aware of the fact that this procedure may rise some questions since the distinction between the overall translation and rotation and conformational changes may be ambiguous.[12] For this reason, it is sometimes proposed to use dihedral angles to construct a covariance matrix instead of Cartesian coordinates. Nevertheless, since we treat these results only as supplementary ones, we have decided to use the Cartesian coordinates for the simplicity of finding the frequencies of the modes. It needs to be stressed that PCA does not require the protein fluctuations to be harmonic. We have made an assumption of the harmonicity of the calculated components but it does not have to be fulfilled. The harmonicity assumption allows us to estimate (using equipartition theorem) the frequencies of the vibration modes of the kinesin from the eigenvalues obtained as a result of PCA.

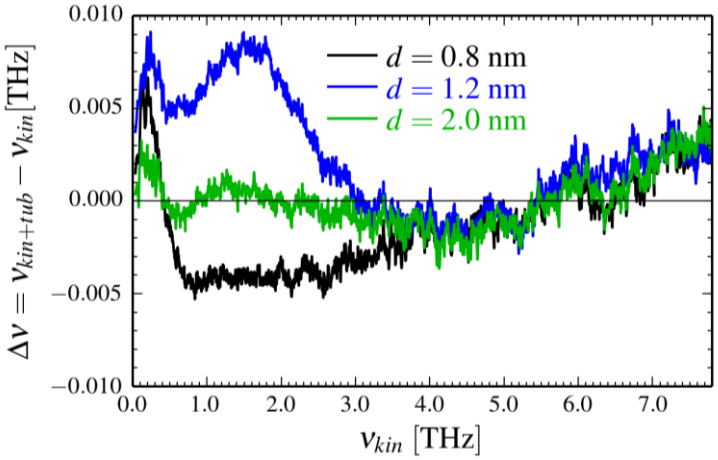

Figure S3. The differences between the frequencies of the consecutive vibration modes of the kinesin. The frequencies were calculated under the assumption of harmonicity. The frequency calculated for the single kinesin head was subtracted from the frequency calculated for the kinesin head in the neighborhood of the tubulin. The differences are plotted versus the frequency calculated for the single kinesin head.

5. Structural properties of modified water models

Modified water models differ from the original SPC/E one in terms of structural and dynamical properties; see Table T1 for illustration. To supplement the information on the number of hydrogen bonds given in the main text we have added the analysis of two more structural parameters, radial distribution function and LSI (local structure index).[13-16] LSI is a measure of heterogeneity of radial distribution of water molecules within short distances (up to about 0.37 nm). It was defined by Shiratani and Sasai[13] as follows. Water molecules surrounding some water molecule *i* are ordered according to the increasing radial distances r_j between the oxygen atom of the molecule *i* and the oxygen atom of the molecule *j*: $r_1 < r_2 < \ldots < r_j < \ldots < r_n < r_{n+1}$, and *n* is chosen so that $r_n < 0.37$ nm $< r_{n+1}$. Then, the value of LSI is calculated as:

$$\mathrm{LSI} = \frac{1}{n}\sum_{j=1}^{n}\left[\Delta(j) - \bar{\Delta}\right]^2 \qquad \text{(e2)}$$

where $\Delta(j) = r_{j+1} - r_j$ and $\bar{\Delta}$ is the average value of $\Delta(j)$ (over all analyzed molecules). Thus, LSI expresses the heterogeneity in the radial distribution of surrounding water molecules around the central one within the sphere of radius equal to about 0.37 nm. Increase of small values of LSI is an indicator of a more homogeneous (more disturbed, less tetrahedral) structure, while increase of large values can be interpreted as an indicator of less homogeneous (more ordered, more tetrahedral) structure. The ordered structure results in high LSI values because of the tetrahedral arrangement and hydrogen bonding of water molecules. If the liquid is ordered then up to the limit of the first solvation shell, we find water molecules at very similar distances from the central one. But the second solvation shell appears after a gap. Therefore, the distance to the first molecule in the second solvation shell is significantly different from the previous ones. This would have an impact on LSI, since the mean value of distances between aligned molecules would be quite different from each distance on its own.

The histograms given below support the statement that the structure of *mod*20 model is less distinct than the structure of SPC/E model (the increase of the values of the histogram for small LSI values and smaller peak on radial distribution function). The opposite is true for the *mod*10 model (the increase of the values of the histogram for large LSI values and taller peak on radial distribution function).

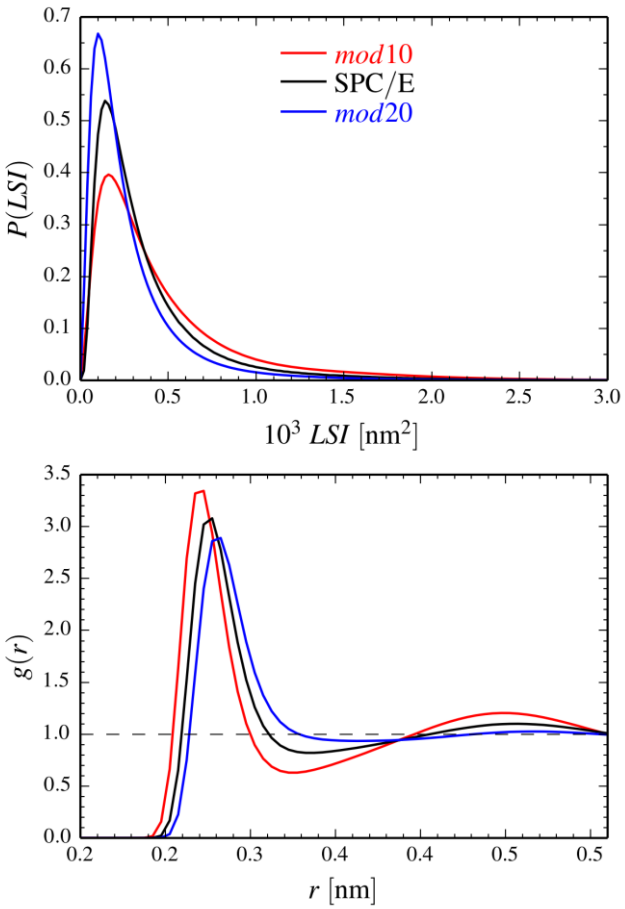

Figure S4. Comparison of properties of SPC/E water model and modified water models: *mod*10 and *mod*20 (LSI and radial distribution function).

Table T1. Characteristics of the water models used in this work. Particular symbols denote: ε and σ are the Lennard-Jones parameters of the oxygen atom; d_W is the density of the liquid; D_T symbolizes the translational diffusion coefficient, while E_{HB} and n_{HB} symbolize the mean energy of a water-water hydrogen bond and the mean number of hydrogen bonds per water molecule, respectively. The physical properties: d_W, D_T and n_{HB}, measured at 298 K, are from reference 17.

model	ε [kJ/mol]	σ [nm]	d_W [g/cm^3]	D_T [10^{-9}m^2·s^{-1}]	n_{HB}	E_{HB} [kJ/mol]
*mod*10	0.41840	0.32277	0.999	1.52	3.48	21.20
SPC/E	0.65017	0.31656	0.999	2.75	3.28	20.12
*mod*20	0.83680	0.31431	0.998	3.71	3.09	19.31

6. Dependence of the phase spectra and amplitude spectra on the value of σ for SPC/E water model and modified water models for proteins separated by 1.2 nm

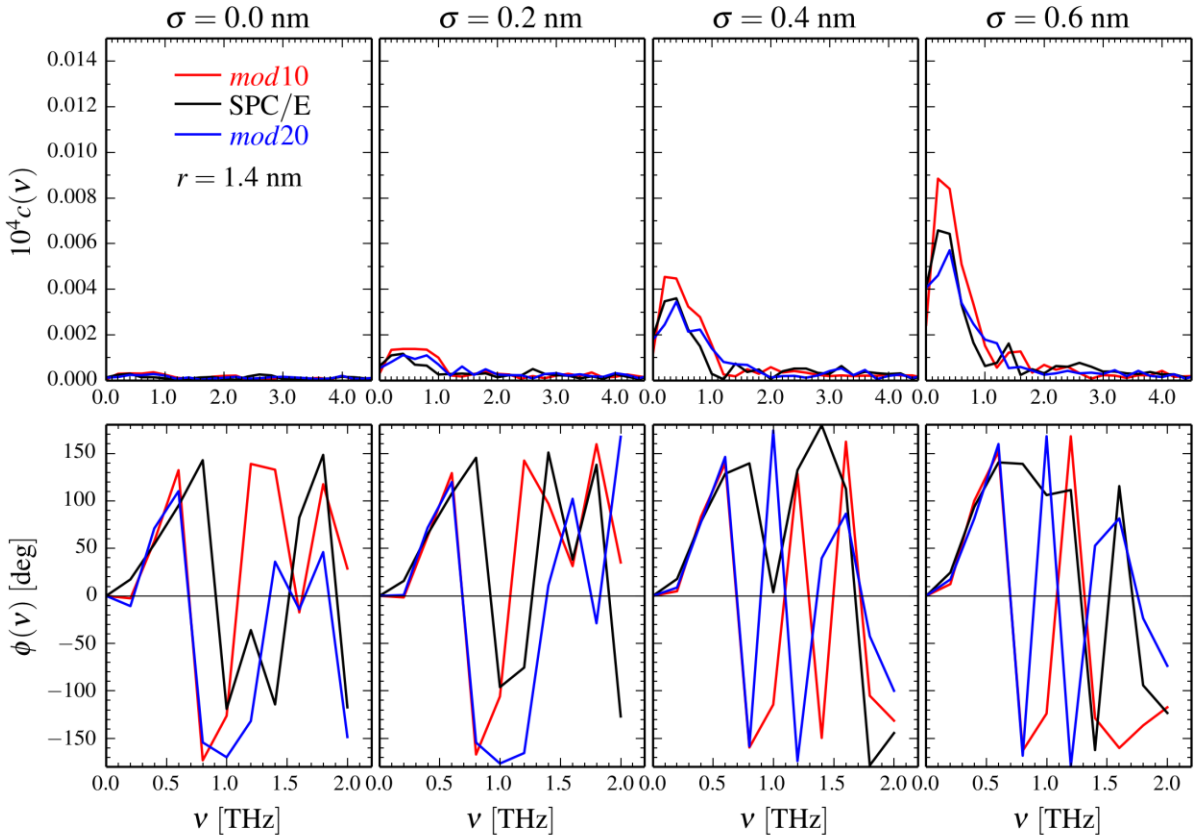

Figure S5. Amplitude spectra (top) and phase spectra (bottom) of the velocity cross-correlation function of the kinesin and the tubulin surface atoms calculated for various values of σ (0 nm, 0.2 nm, 0.4 nm, 0.6 nm). Proteins are separated by 1.2 nm (the maximum of the distribution function at r=1.4 nm) and solvated with SPC/E water model and modified water models: *mod*10 and *mod*20.

6. Maximum flow problem

The goal of solving the maximum flow problem is to find the maximum flow between a source and a sink that occurs through "pipes" represented as edges of the graph. Each edge and each vertex can have its own, unique capacity. The graph can be undirected or directed. In our case, the graph was undirected and the capacities of the vertices were not defined (*i.e*. they were infinite).

Our approach was slightly different from the one described by Ahmad *et al*.,[18] since we have taken into account water molecules in the first solvation shells of both analyzed protein surfaces (defined as those molecules that are present not farther than 0.4 nm from the selected patches on the surfaces of the proteins) and calculated the flow between a source connected with the molecules belonging to the solvation shell of tubulin and a sink connected with the molecules belonging to the solvation shell of kinesin. The edges connecting the water molecules with the source and with the sink all had the same capacity, equal to a mean energy of a hydrogen bond of SPC/E water model. As mentioned in the main text, the capacities of the rest of the edges, representing hydrogen bonds, were equal to the energies of the bonds. The absolute values of the flow are not given in the text, since they are virtually meaningless on their own. They depend strongly on the specific procedure used to determine them. We were only interested in differences between their values.

REFERENCES

1. F. J. Kull, E. P. Sablin, R. Lau, R. J. Fletterick and R. D. Vale *Nature* 1996, **380**, 550-555.

2. J. Löwe, H. Li, K. H. Downing ans E. Nogales *J. Mol. Biol.* 2001, **313**, 1045-1057.

3. C. V. Sindelar and K. H. Downing *J. Cell Biol.* 2007, **177**, 377-385.

4. A. Kuffel and J. Zielkiewicz *Phys. Chem. Chem. Phys*. 2013, **15**, 4527-4537.

5. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P.A. Kollman *Amber 12. Reference manual* 2012, AMBER 12, University of California, San Francisco.

6. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang,P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman *J. Comput. Chem*. 2003, **24**, 1999-2012.

7. E. Sorin, Y. M. Rhee, M. R. Shirts and V. S. Pande *J. Mol. Biol*. 2006, **356**, 248-256.

8. E. C. Dykeman and O. F. Sankey *J. Phys. Cond. Matter* 2010, **22**, 423202.

9. Z. Yang, B. Shi, H. Lu, P. Xiu and R. Zhou *J. Phys. Chem. B* 2011, **115**, 11137–11144.

10. A. Kitao, S. Hayward and N. Go *Proteins*, 1998, **33**, 496-517.

11. A. Amadei, A. B. Linssen and H. J. Berendsen *Proteins*, 1993, **17**, 412-425.

12. F. Sittel, A. Jain and G. Stock *J. Chem. Phys.* 2014, **141**, 014111.

13. E. Shiratani and M. Sasai *J. Chem. Phys*. 1996, **104**, 7671-7680.

14. E. Shiratani and M. Sasai *J. Chem. Phys*. 1998, **108**, 3264-3276.

15. G. A. Appignanesi, F. A. Rodriguez Fris and F. Sciortino *Eur. Phys. J. E: Soft Matter Biol. Phys*. 2009, **29**, 305-310.

16. S. R. Accordino, F. A. Rodriguez Fris, F. Sciortino and G. A. Appignanesi *Eur. Phys. J. E: Soft Matter Biol. Phys*. 2011, **34**, 48.

17. A. Kuffel and J. Zielkiewicz *Phys. Chem. Chem. Phys*. 2012, **14**, 5561-5569.

18. M. Ahmad, W. Gu, T. Geyer and V. Helms *Nat. Comm.* 2011, **2**, 261.