J Mol Model (2011) 17:37–48
DOI 10.1007/s00894-010-0696-6

ORIGINAL PAPER

Structure and dynamics of human vimentin intermediate filament dimer and tetramer in explicit and implicit solvent models

Zhao Qin · Markus J. Buehler

Received: 22 December 2009 / Accepted: 22 February 2010 / Published online: 1 April 2010
© Springer-Verlag 2010

Abstract Intermediate filaments, in addition to microtubules and microfilaments, are one of the three major components of the cytoskeleton in eukaryotic cells, and play an important role in mechanotransduction as well as in providing mechanical stability to cells at large stretch. The molecular structures, mechanical and dynamical properties of the intermediate filament basic building blocks, the dimer and the tetramer, however, have remained elusive due to persistent experimental challenges owing to the large size and fibrillar geometry of this protein. We have recently reported an atomistic-level model of the human vimentin dimer and tetramer, obtained through a bottom-up approach based on structural optimization *via* molecular simulation based on an implicit solvent model (Qin et al. in PLoS ONE 2009 4(10):e7294, 9). Here we present extensive simulations and structural analyses of the model based on ultra large-scale atomistic-level simulations in an explicit solvent model, with system sizes exceeding 500,000 atoms and simulations carried out at 20 ns time-scales. We report a detailed comparison of the structural and dynamical behavior of this large biomolecular model with implicit and explicit solvent models. Our simulations confirm the stability of the molecular model and provide insight into the dynamical properties of the dimer and tetramer. Specifically, our simulations reveal a heterogeneous distribution of the bending stiffness along the molecular axis with the formation of rather soft and highly flexible hinge-like regions defined by non-*alpha*-helical linker domains. We report a comparison of Ramachandran maps and the solvent accessible surface area between implicit and explicit solvent models, and compute the persistence length of the dimer and tetramer structure of vimentin intermediate filaments for various subdomains of the protein. Our simulations provide detailed insight into the dynamical properties of the vimentin dimer and tetramer intermediate filament building blocks, which may guide the development of novel coarse-grained models of intermediate filaments, and could also help in understanding assembly mechanisms.

Keywords Bending stiffness · Biopolymer · Dimer · Explicit solvent · Implicit solvent · Intermediate filaments · Materiomics · Molecular dynamics · Molecular structure · Persistence length · Structural optimization · Tetramer · Vimentin

Z. Qin · M. J. Buehler (✉)
Laboratory for Atomistic and Molecular Mechanics,
Department of Civil and Environmental Engineering,
Massachusetts Institute of Technology,
77 Massachusetts Ave., Room 1-235A&B,
Cambridge, MA, USA
e-mail: mbuehler@MIT.EDU

Z. Qin · M. J. Buehler
Center for Materials Science and Engineering,
Massachusetts Institute of Technology,
77 Massachusetts Ave.,
Cambridge, MA, USA

M. J. Buehler
Center for Computational Engineering,
Massachusetts Institute of Technology,
77 Massachusetts Ave.,
Cambridge, MA, USA

Introduction

Intermediate filaments (IFs), in addition to microtubules and microfilaments, are one of the three major components of the cytoskeleton in eukaryotic cells [1] and are emerging as major players in cell biology, in particular in issues related to mechanobiological mechanisms. IFs are crucial in

defining key mechanical functions of cells such as cell migration, cell division and mechanotransduction, and have also been referred to as the "safety belts of cells" in the biology community as they play a structural role in preventing exceedingly large cell stretch [2–6]. Earlier studies focused on analyzing the mechanical signature of IFs have suggested that they are highly sensitive to applied forces at small levels, and that they can sustain extremely large deformation of up to 300% or four times their initial length [7, 8], an observation that was also observed in recent simulation studies [9]. Proteins of the IF family are also found in the cell's nucleus in the form of lamin, where they form a dense mesh-like network providing mechanical integrity and biochemical functions at the cytoskeleton-chromatin interface [10–12]. IFs have been associated with many genetic diseases, where single point mutations, domain deletions as well as sequence modifications lead to structural changes at different levels in the IFs' hierarchical organization. IF related diseases include muscle dystrophies, Alexander disease [13], epidermolysis bullosa simplex, as well as a broad class of disorders referred to as laminopathies (*e.g.*, rapid aging disease progeria) [14–17]. In extracellular materials, IF proteins are found in hair, hoof and wool, further emphasizing the broad significance of this protein material for a variety of applications in biology, medicine and engineering.

IFs form hierarchical structures as shown in Fig. 1, ranging from dimers, tetramers, unit length filaments, full-length filaments to the cellular level. Each vimentin dimer, the basic building block of this protein filament, contains 466 amino acids. Experimental studies suggested that they form four major structural segments linked in series in the sequence 1A, 1B, 2A and 2B, connected by linkers L1, L12 and L2 (see Fig. 2a). Up until now, a complete atomistic-level molecular model of the dimer or tetramer structure has remained elusive. So far, only parts of two of the four segments of the dimer, a section of the 1A and 2B domain of the vimentin dimer structure, have been crystallized and their atomic structure identified based on x-ray diffraction experiments [18–22] (the structures are found in Protein Data Bank (PDB) entries 1gk4, 1gk6, 3klt and 1gk7). Persistent experimental challenges have prevented the identification of the remaining parts of the IF structure using experimental approaches. This is because IFs are intrinsically disordered structures [23], so that x-ray diffraction studies on naturally occurring or recombinantly produced IF bundles do not provide sufficient data to produce a full atomistic model of IFs. Whereas solid state nuclear magnetic resonance (ssNMR) has been successfully utilized to derive atomistic models of amyloid fibrils [24], this approach has not yet worked for IF dimers or tetramer, likely because amyloid peptides are much smaller than IF dimers. Cryoelectron tomography has been suggested as a possible molecular-level imaging tool to visualize single IFs, but the highest resolution tomograms are still limited to a resolution around 5 nm [25].

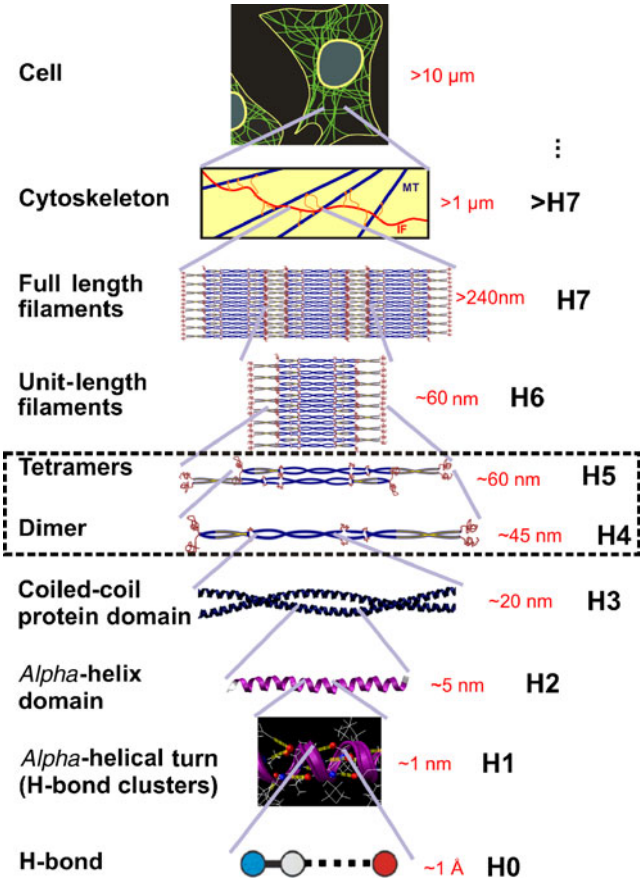

Fig. 1 The hierarchical structure of intermediate filaments, from atomic to cellular scales. The figure shows relevant structural levels (H0 to H7). The present paper focuses on structural analysis of levels H4 and H5 (dimer/tetramer structure). The regions marked in yellow in the dimer level are those that have been crystallized in earlier studies. The focus of the study reported in this paper is on vimentin dimers and tetramers, corresponding to hierarchy levels H4 and H5 as marked in the figure. Figure adapted from reference [9]

The lack of a complete structural model of IFs has prevented us from addressing fundamental structure-function relationship questions related to the mechanical role of intermediate filaments, which is crucial to link structure and function in the protein filament's biological context. The availability of a structural model of IFs could also be the key to understanding the mechanisms of IF related genetic diseases, where structural flaws that originate at the genetic or molecular level may cause major structural changes of biologically relevant properties. To address these issues, here we provide a detailed analysis of a molecular model of IF dimers and tetramers with atomistic resolution that was recently obtained based on a bottom-up molecular simulation approach using an effective solvent model [9]. The earlier structure identification

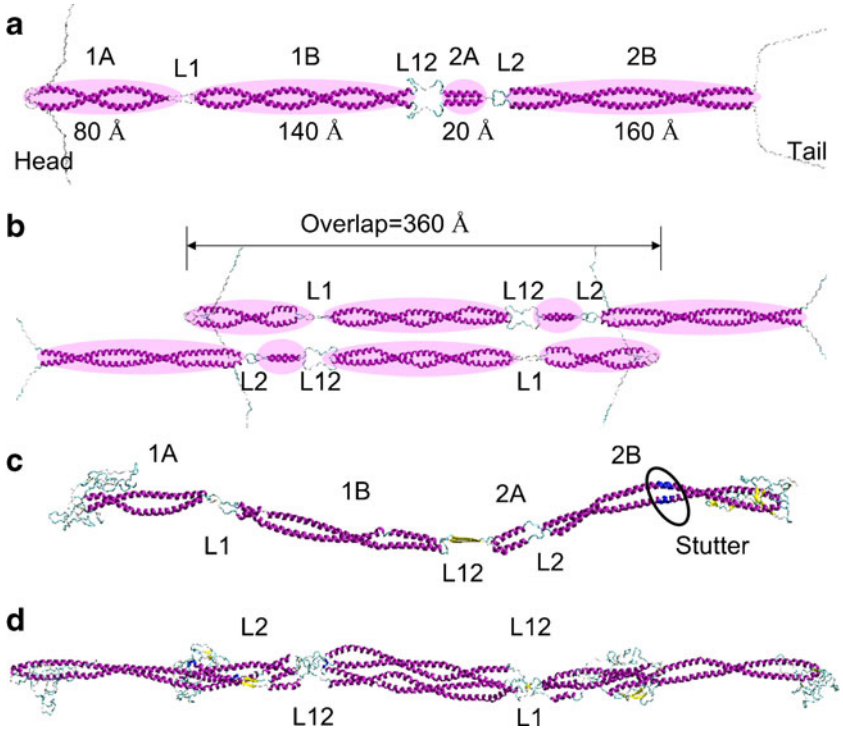

Fig. 2 Schematics of the molecular structure of the vimentin dimer (panel **a**) and the vimentin tetramer (panel **b**). The plot includes labels identifying the various segments and linker domains, as well as the head and tail domain. Panels **c** and **d** show the equilibrated structures of the dimer and tetramer, obtained from implicit solvent simulations as reported in [9]

approach utilized the available amino acid sequence and associated structural information from experimental results to generate an initial geometry that served as the starting point for a series of energy minimization and equilibration runs, which were subsequently performed using molecular dynamics. These simulations resulted in an equilibrated structure of both the vimentin dimer and tetramer.

Since the earlier study focused primarily on simulations in implicit solvent, an important outstanding question remains a detailed structural and dynamical study of the behavior of the resulting proteins in explicit solvent. Several earlier comparisons between explicit and implicit solvents have been reported in the literature, focused on exploring the effects of the solvents in conformation sampling and energy landscape in folding [26, 27]. In these studies it was found that simulation of proteins in implicit solvent can reproduce the set of local energy minimum in explicit solvent, but some energy terms such as salt bridge may be overestimated. However, most of the earlier works were focused on small peptides. The study reported here compares the effects of implicit versus explicit solvent for the cases of a IF dimer and tetramer, which are much larger proteins than the small peptides studied before (932 amino acids of a dimer and 1864 amino acids of a tetramer). To address this issue, here we present large-scale validation simulations and structural analyses of the atomistic model using structural equilibration in large-scale atomistic simulations in explicit solvent, with system sizes up to 500,000 atoms and simulations carried out at 20 ns time-scales. Based on these explicit solvent simulations, we report a detailed comparison of the dynamical behavior of this biomolecular model within implicit and explicit solvent models. Our studies confirm the stability of the molecular model, and provide insight into the dynamical properties of the dimer and tetramer.

Materials and methods

Background information on implicit solvent versus explicit solvent models

The solvent environment plays an important role in the structure, dynamics, and function of biomolecules. It is generally agreed upon that an explicit solvent model provides the most accurate treatment of solute-solvent interaction [28], because each explicit solvent molecule reflects its realistic chemical structure and resulting interactions with the solute and one another. However, the approach of relying on an explicit treatment of solvent molecules significantly increases the system size by several orders of magnitude. To achieve a better balance of computational efficiency and physical accuracy, implicit solvent models have been proposed as an alternative approach [29, 30]. Implicit solvent models add the effective solvent energy to the biological molecule, expressed indirectly as a function of the molecular structure and its conformation.

This strategy simplifies the numerical effort of simulating the physical system of interest and thereby speeds up

the simulation efficiency by averaging the solvent effect. Since no frictional effects are included in implicit solvent methods, such models can facilitate a more rapid structural organization toward the equilibrium geometry due to the lack of viscosity imposed by explicit water molecules. This effect can be advantageous in structure identification approaches [31–33] (such as in the study reported in [9]) since it accelerates the sampling of molecular configurations, so that more configurations are visited per simulated unit time. Whereas the approach of using implicit solvents is computationally efficient, it has been discussed that the implicit model may overestimate some energy terms and lead to inaccurate results, in particular regarding the detailed configurations of domains in a protein [27]. Therefore, further validation and comparison by explicit solvent models is needed to ensure the validity of the results obtained from implicit solvent simulations. To address these issues, a detailed comparison between implicit and explicit simulation models is provided here, applied specifically to a vimentin intermediate filament dimer and tetramer.

Molecular models and force field

The implicit solvent simulations whose results are analyzed here [9] were carried out with the CHARMM program by using the CHARMM19 all-atom energy function and an implicit Gaussian model for the water solvent [29, 30]. Explicit solvent simulations [34] are used to confirm the stability of the equilibrated structures, showing that the structures predicted by our approach are stable also in explicit solvent. We use the CHARMM force field and explicit TIP3 water [34] implemented in NAMD [35], starting from the equilibrated structure obtained using the approach described above. The time step used in all simulations is 1 fs, at a constant temperature (300 K) controlled by Langevin thermostat. A pressure of 1 atmosphere is used in the explicit water simulations. We use the particle mesh Ewald (PME) method in calculating the electrostatic interactions, with a grid size of 1 Å in all three directions. A cutoff length of 10 Å is applied to the van der Waals interactions and to switch the electrostatic forces from short range to long range forces. We use pure water in our simulations (no ions are included).

Initial molecular structure

The structure of the dimer has been identified by a series of computational steps, as reported in [9]. Figure 2a and b show a schematic of the structural character of the dimer and the tetramer, with each domain labeled. Figure 2c and d depict a snapshot of the equilibrated dimer and tetramer structures as obtained from implicit solvent calculations. These geometries from earlier studies [9] are used as a starting point for the simulations reported here. The dimer structure is reported at http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0007294.s010, and the tetramer structure is reported at http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0007294.s011 is used. Explicit water is added in a rectangular water box with periodic boundary conditions in all three directions using visual molecular dynamics (VMD). We use a rectangular solvent box with dimension of 54 nm× 7.5 nm×7 nm for the dimer and 66 nm×9 nm×9 nm for the tetramer. Because the diameters of the dimer and tetrameter are estimated to be 2.4 nm and 4 nm, respectively, the nearest distances between two neighboring dimers and tetramers are 4.6 nm and 5 nm, respectively. These distances are much larger than the cutoff length of the pair list distance (12 Å) and thereby there is no interaction between the neighboring dimers or tetramers.

The dimer and tetramer systems in explicit solvent contain ≈250,000 in the dimer and ≈500,000 atoms in the tetramer case (in this case ≈30,000 atoms for the protein plus ≈470,000 for the water molecules). The use of parallelized supercomputers enables us to reach about 20 ns in about one month of simulation time.

Results

We begin with a comparison of the equilibrated vimentin dimer and tetramer geometries as obtained from implicit and explicit equilibration. Figure 3a shows snapshots of the molecular structure of the vimentin dimer, displaying the characteristic segmented geometry with coiled-coil regions connected through linker domains. The figure presents a comparison of the structure based on implicit solvent equilibration and explicit solvent equilibration, where the two geometries are superimposed in the plot. Figure 3b–e show detailed views of key domains in the protein structure, also in a visualization mode where the implicit and explicit solvent model results are superimposed. It can be seen that the linker L12 forms a parallel *beta*-sheet in both explicit and implicit solvent models. We find that in both cases the total length of the dimer without the head and tail domain is ≈49 nm. This result is in agreement with experimental results where a range of 46–49 nm has been reported [19, 36]. The visual comparison of the dimer structure between implicit and explicit solvent models shows that the overall geometry of the dimer is stable in either solvent model. We present a detailed analysis of the structural character of the equilibrated structure of the dimer in the two solvents in Fig. 3f and g, where the color along the molecular axis

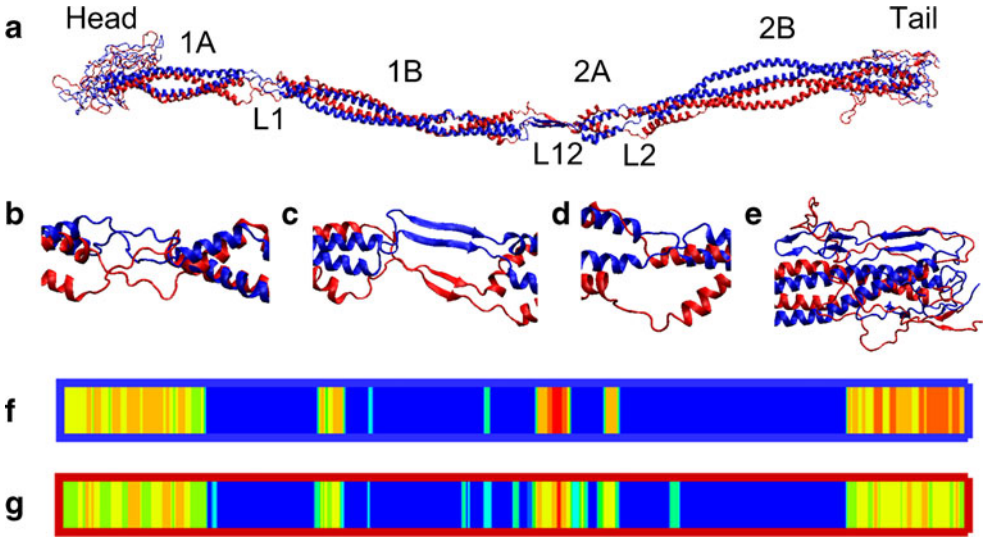

Fig. 3 Structure of the vimentin intermediate filament dimer, depicting a comparison between implicit and explicit solvent simulation results. Panel **a**: Snapshot of the atomistic model of the dimer after equilibration, in implicit water (in blue) and in explicit water (in red). Panels **b**–**e** show detailed views of the structure of linkers L1, L12, L2 and the tail domain, respectively. The *beta*-sheet structure of linker L12 is visible in panel **c** (we note that while the *beta*-sheet is rather stable in the implicit simulations, it is only marginally stable in the explicit solvent model). Panels **f** and **g** show the structural character of each amino acid along the polypeptide chain length, for the implicit (panel **f**) and explicit (panel **g**) solvent. *Alpha*-helical structures are colored in blue, *beta*-sheet structures are colored in red, and the other colors represent random coils

denotes the protein filament's secondary structure. By comparing the two structures, we find that the 2A segment (parallel helices) and L12 (*beta*-sheet) are only marginally stable, because the 2A segment lacks the hierarchical structure in enforcing its strength, and short *beta*-sheet segments are more energy favorable at the more flexible parts of the protein [37].

Figure 4a shows a comparison of the vimentin tetramer structure of the implicit solvent equilibration and explicit solvent equilibration results, where the two geometries are superimposed in the plot. The total length of the tetramer is ≈60 nm with a slightly smaller length of ≈59 nm in the explicit solvent case. These findings agree well with the length of unit length filaments observed in experiment,

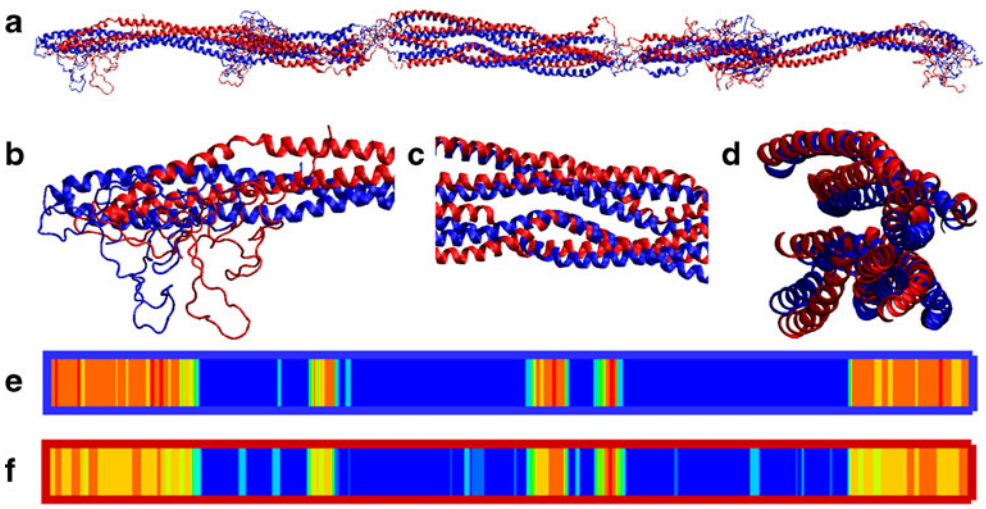

Fig. 4 Structure of the vimentin intermediate filament tetramer, depicting a comparison between implicit and explicit solvent simulation results. Panel **a**: Snapshot of the atomistic model of the dimer after equilibration, in implicit water (in blue) and in explicit water (in red). Panels **b**, **c**, **d** and **e** show detailed views of the structure of head and overlapped domain from two viewing directions, respectively. Panels **e** and **f** show the structural character of each amino acid along the polypeptide chain length, for the implicit (panel **e**) and explicit (panel **f**) solvent. *Alpha*-helical structures are colored in blue, *beta*-sheet structures are colored in red, and the other colors represent random coils

62 nm at pH 7.5 [36]. In both cases the overlap part of the tetramer has a length of ≈36 nm, where the experimental value is 30–36 nm [19, 36]. Segments 1A, 1B and 2A are fully contained in this overlapped part, but the 2B segments in the immediate vicinity of the two terminals are located completely outside. We find that the head segment of each dimer is coiled around the other dimer, increasing the contact surface area and thereby providing enhanced interdimer interactions, as shown in Fig. 4b. The structure of the coiled-coil dominated overlap region is rather stable, as shown in Fig. 4c and d (the images show the same part of the tetramer from different viewing directions). As for the dimer case discussed in the previous paragraph, here we also present a detailed analysis of the structural character of the equilibrated protein in Fig. 4e and f. The analysis shows that in comparison with the dimer case, the tetramer structure is more stable, and most of the structural features are well conserved after explicit solvent equilibration.

We analyze the structure of the dimer and tetramer by calculating the dihedral angle distribution as shown in the Ramachandran map [38] as depicted in Fig. 5. This analysis shows the probability of specific dihedral angles [φ, ψ] of all amino acid residues in the dimer/tetramer structure in equilibrium. The analysis shows that most of the residues (≈ 73%) belong to the left lower corner (φ< 0°, ψ<0°), and the peak value emerges at [φ≈−58°, ψ≈ −47°]. This data shows the character of a typical right-handed *alpha*-helix. Most other residues (≈ 25%) belong to the left upper corner (φ<0°, ψ>0°), and there is a much less intense peak that resembles the character of a *beta*-sheet in a Ramachandran map (−180°<φ<−50° and 50°<ψ<180°). The position of this peak center for the dimer/tetramer structure in implicit solvent ([φ≈−80°, ψ≈100°]) is slightly different from that of the explicit solvent ([φ≈ −80°, ψ≈140°]), indicating that the *beta*-sheet in explicit solvent has a more extended conformation than that in implicit solvent.

We further calculate the number of H-bonds in the dimer and tetramer structure during equilibration, and compare the data based on implicit and explicit solvent simulations as

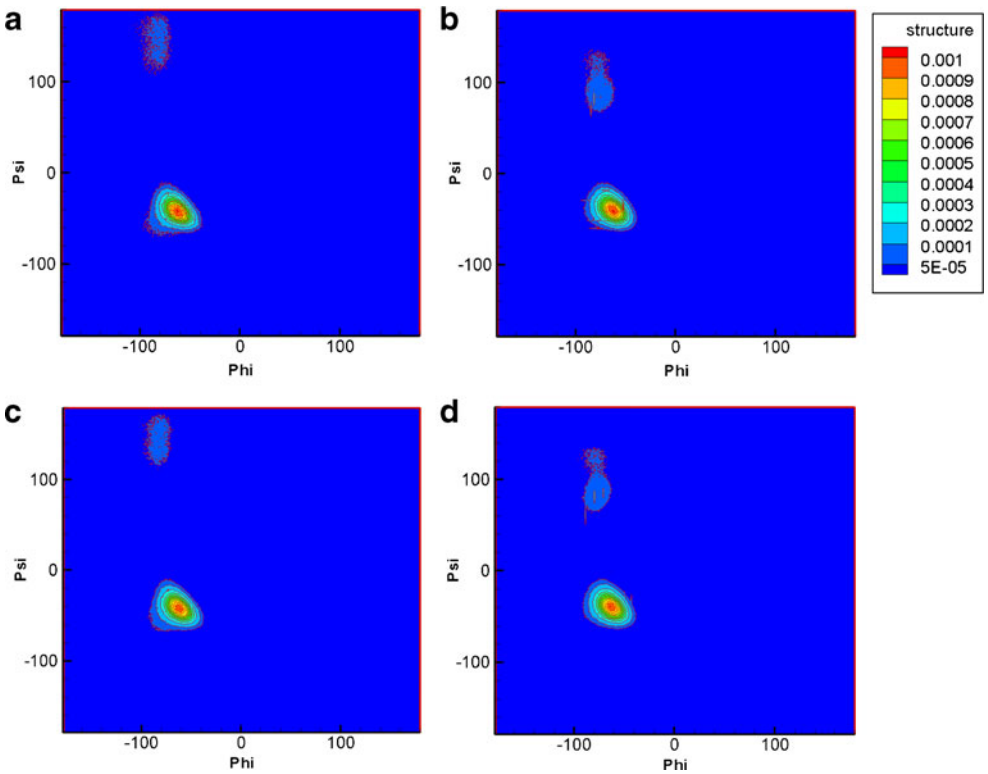

Fig. 5 Ramachandran maps that show the averaged dihedral angle of each amino-acid ([φ, ψ]) distribution during the equilibrium process, comparing implicit and explicit simulation results. Panels **a** and **b** show the results for the dimer in explicit solvent (**a**) and implicit solvent (**b**). Panels **c** and **d** show the results for the tetramer in explicit solvent (**c**) and implicit solvent (**d**). The analysis shows that most of the residues (≈ 73%) belong to the left lower corner (φ< 0°, ψ<0°), and the peak value emerges at [φ≈−58°, ψ≈−47°]. This data shows the characteristic of a right-handed *alpha*-helix. Most other residues (≈ 25%) belong to the left upper corner (φ<0°, ψ>0°), and there is a much less intense peak that resembles the character of a *beta*-sheet in a Ramachandran map (−180°<φ<−50° and 50°<ψ< 180°). The position of this peak center for the dimer/tetramer structure in implicit solvent ([φ≈−80°, ψ≈100°]) is slightly different from that of the explicit solvent ([φ≈−80°, ψ≈140°]), indicating that the *beta*-sheet in explicit solvent has a more extended conformation than that in implicit solvent

shown in Fig. 6. We observe that the number of H-bonds in both the dimer and tetramer structure decreases after energy minimization and converges asymptotically to a constant value in explicit solvent. The asymptotic number of H-bonds is also stable for the simulations in implicit solvent, as shown in Fig. 6a and b. By subtracting two times the number of H-bond in the dimer from that of the tetramer, we estimate the number of *inter*-dimer H-bonds, which varies with time as shown in Fig. 6c. The asymptotic value by the end of the simulation (by averaging over the last 20% of the total simulation time) is given in Fig. 6d, where it is observed that the number of *inter*-dimer H-bonds in explicit solvent (145±5) is greater than that obtained in implicit solvent (78±8), indicating that the interaction between two dimers in explicit solvent might be stronger than that of the implicit solvent.

The solvent accessible surface area (SASA), which reflects the surface area that is accessible to a water molecule with radius of 1.4 Å, of the dimer and tetramer is shown in Fig. 7a and b. It is shown that the SASA in the dimer and tetramer structures in explicit solvent increases after energy minimization, while it is relatively stable in implicit solvent. This phenomenon agrees with the observations made in the [φ, ψ] analysis, and suggests that the molecular conformation in explicit solvent tends to become more extended leading to an increase in the SASA value. We calculate the covered *inter*-dimer surface area by subtracting the SASA value of the tetramer from two times of the SASA value of the dimer, as shown in Fig. 7c. The asymptotic value by the end of the simulation (by averaging over the last 20% of the total simulation time) is shown in Fig. 7d, where it can be seen that the two dimers in a tetramer model become less adhesive in explicit solvent as they allow for more water to fill the gap between the dimers as shown in Fig. 7c. This phenomenon is not observed in implicit solvent. By the end of the simulation, the hidden surface areas of the implicit and explicit solvent simulations converge to 13,800 $Å^2$ and 10,800 $Å^2$, respectively, as shown in Fig. 7d. This finding suggests that the hidden area between two dimers in explicit solvent is 22% smaller than that observed under implicit solvent conditions.

From dynamical analyses of the trajectory of the dimer at 300 K, we observe that the linker domains represent rather soft, hinge-like structures that connect much stiffer coiled-coil segments. The heterogeneous distribution of the bending stiffness along the molecule's axis strongly affects the filaments' motion by enhancing its mobility due to an effectively reduced persistence length, a mechanism that might be significant for IF self-assembly mechanisms. Thereby, the linker domains act as hinges around which the molecular structure can more easily bend and rotate. The particular distribution of soft and stiff elements along the dimer axis suggests a simplified model of the protein structure as shown in Fig. 8a. In this representation, each coiled-coil is represented by a rod and each linker is represented by a hinge. Thereby the distances *L1*, *L2*, *L3* and *L4*, as well as the angles *θ1*, *θ2* and *θ3* describe the key geometrical parameters of the dimer structure. We

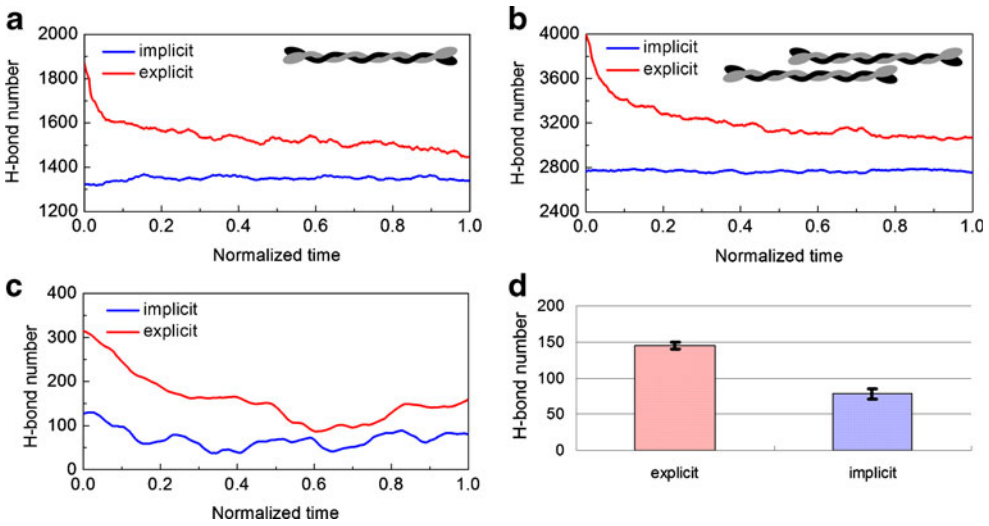

Fig. 6 Number of H-bonds of the vimentin intermediate filament dimer and tetramer, showing a comparison between implicit and explicit solvent simulation results. Panel **a**: Number of H-bonds in a dimer. Panel **b**: Number of h-bonds in a tetramer. Panel **c**: Number of H-bonds between the two dimers within a tetramer. Each of the panels **a**, **b** and **c** are obtained during the equilibration process, in both implicit and explicit solvent, respectively. Panel **d**: Comparison between the asymptotic values of the number of interdimer H-bonds in both implicit and explicit solvent. The main bar gives the average value during the last 20% of the total simulation time while the corresponding error bar corresponds to the standard deviation. In panels **a**–**c**, the simulation time is normalized by the length of the total equilibration (10 ns for the implicit solvent for both dimer and tetramer, 16 ns for the explicit solvent for the dimer, and 19 ns for explicit solvent simulations for the tetramer)

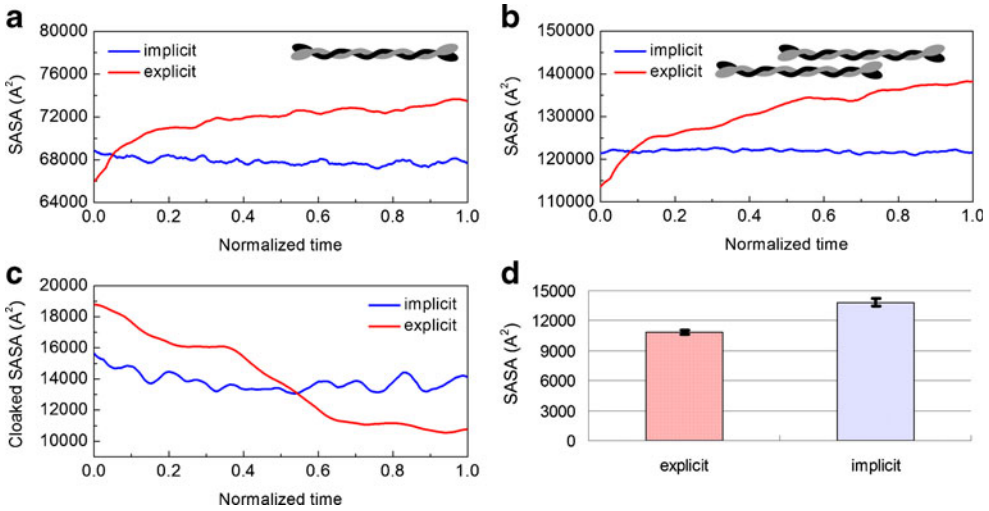

Fig. 7 Solvent accessible surface area (SASA) of the vimentin intermediate filament dimer and tetramer, showing a comparison between implicit and explicit solvent models. Panel **a**: SASA of the dimer. Panel **b**: SASA of the tetramer. Panel **c**: SASA between the two dimers within a tetramer. Each of the Panel **a**, **b** and **c** report results obtained during the equilibrium process, in implicit and explicit solvent, respectively. Panel **d**: Comparison between the asymptotic values of the between-dimer-SASA in implicit and explicit solvent. The main bar gives the average value during the last 20% of the total simulation time, and the corresponding error bar depicts the standard deviation. The simulation time is normalized by the duration of the complete equilibration (10 ns for the implicit solvent for both the dimer and the tetramer, 16 ns for the explicit solvent for the dimer, and 19 ns for explicit solvent simulations for the tetramer)

measure the value of the distances *L1*, *L2*, *L3* and *L4*, as well as the angles *θ1*, *θ2* and *θ3* during both the implicit and explicit equilibration and summarize the resulting mean and standard deviation values in Table 1. Figure 8b

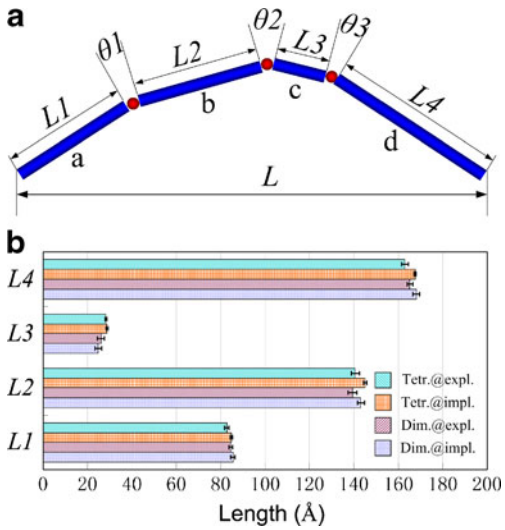

Fig. 8 Analysis of the geometric parameters of a dimer. Panel **a**: Simplified dynamical model of a dimer, which is considered here as a collection of rod-like domains connected by hinge-like linkers. The points a, b, c and d correspond to the mid points of the corresponding rod-like domains. Panel **b**: the length of each rod-like domain within dimer/tetramer in different solvents (implicit/explicit). The mean value is given by the main bar, and standard deviations are plotted *via* error bars

compares the length of each rod-like domain in both the dimer and the tetramer based on the implicit and explicit solvent model. It is shown that the lengths of all the four coiled-coil domains remain almost constant during simulation, providing evidence that both implicit and explicit models provide a similar description of the structure of the rod-like domains. The geometric location of the bending angles *θ1*, *θ2* and *θ3* as a function of the normalized simulation time are shown in Fig. 9a–c. We observe that the bending angle of each linker is fluctuating around an average value during the simulation. This average value and the fluctuation range is summarized in Fig. 9d, in which we plot the mean value and standard deviation of the three angles. We observe that the fluctuations of the angles in the tetramer are smaller than in the dimer, indicating that the structure is stabilized in the assembling process. Using the definition of the persistence length L_p

$$\cos(\theta_{mean}) = \left\langle \vec{t}(s) \cdot \vec{t}(s') \right\rangle = \exp\left[-\frac{(s-s')}{L_p}\right] \qquad (1)$$

where s and s' are coordinates along the contour lengths of the chain. Using the mean angle values in the tetramer obtained from the explicit solvent simulations we compute the persistence length of the three parts, each of which has a linker at its center, as $L_p^{ab} = 322\text{nm}$, $L_p^{bc} = 234\text{nm}$ and $L_p^{cd} = 392\text{nm}$. The effective persistence length of a full dimer protein within the tetramer structure is estimated to be $L_p^{ad} = 402\text{nm}$.

Table 1 Structure analysis of each domain within the dimer and tetramer in implicit solvent and explicit solvent during equilibrium, the geometric parameters are as shown in Fig. 5a

Geometric parameter		Dimer in implicit solvent	Dimer in explicit solvent	Tetramer in implicit solvent	Tetramer in explicit solvent
L1	Average (Å)	85.5	84.6	84.8	82.8
	SD. ±(Å)	1.0	0.9	0.5	1.0
L2	Average (Å)	143.2	139.6	145.1	140.7
	SD. ±(Å)	1.6	2.0	0.8	1.8
L3	Average (Å)	24.8	26.1	28.8	28.3
	SD. ±(Å)	1.5	1.5	0.5	0.5
L4	Average (Å)	168.0	165.2	167.5	162.9
	SD. ±(Å)	1.5	1.4	0.5	1.5
L	Average (Å)	487.0	479.7	613.0	593.4
	SD. ±(Å)	–	2.2	–	3.2
θ1	Average (°)	10.7	12.0	15.8	9.8
	SD. ±(°)	4.7	5.7	2.9	3.1
θ2	Average (°)	13.9	12.7	7.9	15.3
	SD. ±(°)	6.9	6.0	4.4	6.2
θ3	Average (°)	13.7	11.6	11.2	12.6
	SD. ±(°)	7.1	6.3	4.1	4.1

It is noted that except for *L* (which represents the end-to-end length of a dimer or tetramer), all other parameters are measured based on the rod-like domains and linkers as identified in the dimer structure. SD = standard deviation, used here to define error bars

Based on the estimated persistence length, the bending modulus of the dimer can be expressed as [39]

$$E_{bend} = k_B T L_p / I \tag{2}$$

where $I = \frac{\pi d^4}{64}$, and d=2.4 nm is the diameter of the dimer including the vdW radius (obtained from geometric analysis of the protein structure). Thereby, the bending modulus is found to be E_{bend}=1 GPa. The bending modulus obtained from experimental bending was reported to be around 900 MPa in earlier studies [39], close to our results.

In extending this analysis, we estimate the angular stiffness of each hinge shown in Fig. 8a. Using a beam theory for a clamped beam, the angle change is given by:

$$\Delta\theta = \frac{Ml}{E_{bend} I} \tag{3}$$

where l is the contour length of the domain considered (for example, in studying $\overline{ab}$, we have $l=(l_1+l_2)/2$). Since $K_\theta = \frac{M}{\Delta\theta}$, combining with Eq. 2 and we obtain

$$K_\theta = \frac{k_B T L_p}{l}. \tag{4}$$

Based on this analysis the angular stiffness for each of the three hinges is calculated to be $K_{\theta 1} = 1.72\text{kcal/mol}$ (for L1), $K_{\theta 2} = 1.65\text{kcal/mol}$ (for L12), and $K_{\theta 3} = 2.44\text{kcal/mol}$ (for L2). It is observed that all these values are relatively small, which explains why each rod-like domain can easily rotate around its linker, in agreement with the observations made in the dynamical simulations.

Finally, Fig. 10 shows the evolution of the root mean square deviation (RMSD) for the dimer (Fig. 10a) and the tetramer (Fig. 10b) in explicit solvent (compared with the initial structure as obtained from implicit solvent equilibration). The results show that after 12 ns, the RMSD values for both the dimer and tetramer converge to a constant value, suggesting that both structures have reached a stable configuration in the explicit solvent model (the final structures are the structures shown in Figs. 3 and 4, respectively). Furthermore, the overall relatively small deviation of the structure provides evidence for the fact that the implicit model is a reasonable approximation for the dimer and tetramer structure, at least at the time-scales considered here.

Additional analysis of the dynamical properties of each amino-acid in the dimer and tetramer during the simulation is shown in Fig. 11, where the RMSD value for each residue in the simulation is depicted. These results also reflect the relative flexibility of each domain during the equilibration process. We observe that the peaks shown in Fig. 11a correspond to the linker, head and tail segments, providing the evidence that these domains are among the most flexible ones in the dimer structure and as such undergo the largest structural fluctuations. The peak RMSD values P corresponding to the linkers have a relation as $P_{L12} > P_{L1} > P_{L2}$, which agrees with the relation of the angular stiffness of each linker $K_{\theta 2} K_{\theta 1} K_{\theta 3}$. It is also

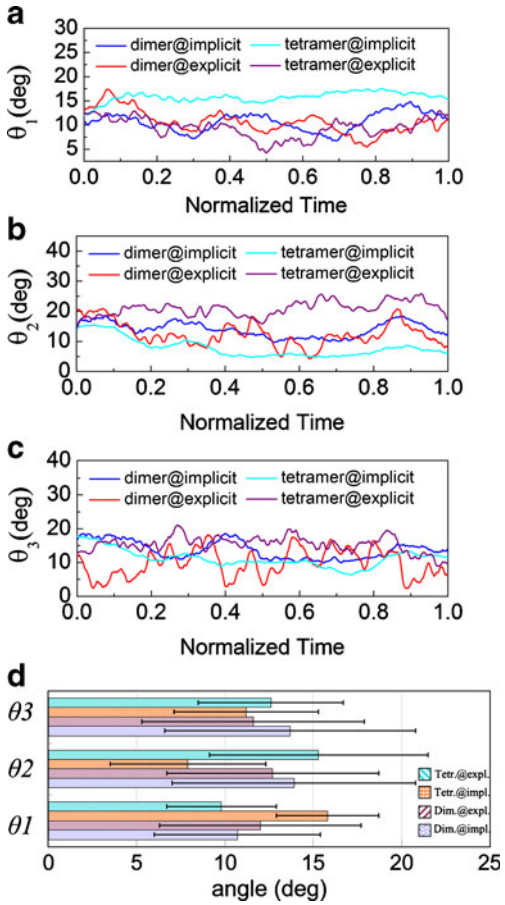

Fig. 9 Analysis of hinge angles of a dimer and tetramer. Panels **a**, **b** and **c** show how the angles θ_1, θ_2 and θ_3 change during the simulation (for geometry see Fig. 5a). The simulation time is normalized by the total equilibration time length (10 ns for the implicit solvent for both dimer and tetramer, 16 ns for the explicit solvent for the dimer, and 19 ns for explicit solvent simulations for the tetramer). Panel **d** summarizes the mean value and standard deviations of the three angles during the simulation. The mean value is given by the main bar, and standard deviations are plotted *via* error bars

interesting to note that the right part (amino acid numbers 291 to 351) of the 2B segment are more stable than the left part (amino acid numbers 351 to 405), as shown by the peaks in these regions. Experiments have revealed that a stutter defect is located around amino acid number 351. This information provides evidence that the 2B segment is marginally stable from the stutter region on to the beginning of the tail domain, but more flexible before the stutter region. The RMSD value of the tetramer (Fig. 11b) shows that the interaction between two dimers significantly affects the flexibility and stability. As the two *anti*-parallel dimers combine, the overlapped part becomes more stable than the other domains. Moreover, we do not find a peak at the L2 domain, which confirms that this linker interacts with the rod like segment of the other dimer and its flexibility is significantly reduced, reflecting an enhanced stability in this assembly stage.

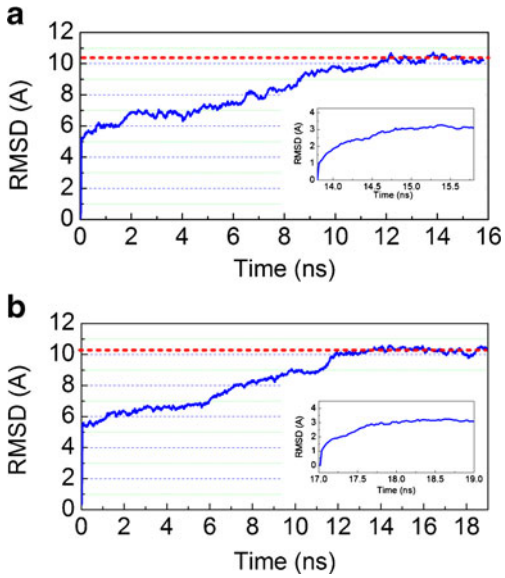

Fig. 10 Root mean square displacement (RMSD) analysis in explicit solvent, for the dimer (panel **a**) and the tetramer (panel **b**). The deviation is calculated with respect to the starting configuration, which is the result of equilibration in implicit solvent. Both systems are seen to converge to a constant RMSD value at approximately 12 ns time-scale, with an overall structural deviation of approximately 10 Å. The inserts in each panel depict the RMSD analysis during the last 2 ns of equilibration for the dimer and tetramer, respectively

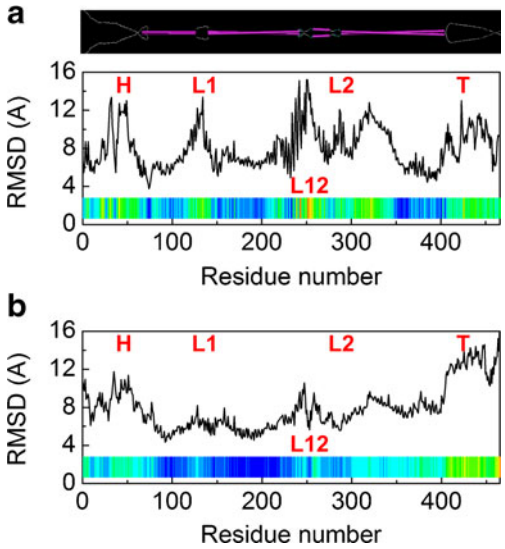

Fig. 11 Root mean square displacement (RMSD) analysis during the equilibrium process in explicit solvent (during the period as shown in Fig. 6), for the dimer (panel **a**) and the tetramer (panel **b**). The residues in the range from 1 to 466 refer to the number in one of the polypeptide chains. It corresponds to the average of RMSD values for two chain (dimer) and to four chains (tetramer), respectively. The color bars show the magnitude of RMSD value from low (blue) to high (red). The location of head (H), tail (T) and linker domains (L1, L12, L2) are indicated in the plot. A schematic of the overall dimer structure is shown above panel **a**

Discussion and conclusions

Based on molecular simulations, we analyzed and compared the structural and dynamical properties of the vimentin dimer and tetramer in implicit and explicit solvent model. In implicit solvent, a long time equilibration was performed in earlier studies reach a minimum energy state. A continued equilibration in explicit solvent as reported in this paper for up to 20 ns reveals that the structural model developed earlier [9] appears to be stable also in explicit solvent, as confirmed in the analyses shown in Figs. 3 and 4, as well as through the RMSD analyses reported in Fig. 10. The analysis of Ramachandran maps as reported in Fig. 5 provides important insight into structural parameters in implicit *vs*. explicit solvents, and could perhaps be used as a method to compare against similar experimental data. The detailed analysis of the number of H-bonds within each dimer and the tetramer as a whole, as well as between the two dimers in the tetramer comparing implicit *vs*. explicit solvent models (see Fig. 6), reveals some differences between the two models, suggesting that the explicit solvent model tends to feature a greater H-bond density. The detailed analysis of the solvent accessible surface area as shown in Fig. 7 suggests that the hidden area between two dimers in explicit solvent is smaller than that of the implicit solvent. The long time-scale simulations provided us with ensembles of conformations, which enabled us to analyze the dynamical properties of the dimer and tetramer around the equilibrium state. This analysis revealed a heterogeneous distribution of the bending stiffness along the molecular axis, where highly flexible hinge-like regions defined by non-*alpha*-helical linkers connect stiffer regions of the protein, as shown in Figs. 8 and 9.

Overall, our simulations further provided detailed insight into the dynamical properties of the vimentin dimer and tetramer intermediate filament building blocks, which may guide the development of novel coarse-grained models of intermediate filaments that could help in understanding assembly mechanisms. The analysis method used in this paper could be used to gain insight into the structural and dynamical properties of other large biological molecules. However, it should be noted that the analysis reported here has limitations, in particular with respect to accessible time-scales. Accurate calculations can only be performed starting from an equilibrated or near-equilibrated structure, and the simulation time must be sufficient to include enough possible conformations that represent a reasonable sampling of the configurational space. The use of explicit solvents in particular is computational very expensive. In light of these limitations, the implicit solvent calculations show advantages in terms of efficiency at sufficient levels of structural and energetic accuracy.

Future studies of tensile stretching with explicit models could be carried out, albeit these simulations would likely contain millions of atoms due to very large water boxes needed to accommodate the large deformation of the molecules. Other future work could be focused on setting up a coarse-grained *n*-body structural model by treating each dimer as rod-like coiled-coils connected by hinge-like linkers (similar to the suggestion put forth in Fig. 8a), or perhaps using methods that provide enhanced sampling of structural configurations (*e.g*., replica exchange method) to accelerate the overall simulation speed. The computational expense of full atomistic simulations in explicit solvent is significant in particular if long time-scales are considered. Additional studies could be carried out with advanced time-scale sampling techniques such as replica exchange simulations.

Acknowledgments ZQ and MJB acknowledge support by Air Force Office of Scientific Research (AFOSR) (FA9550-08-1-0321) and National Science Foundation (NSF) (MRSEC DMR-081976). This research was supported by an allocation of advanced computing resources supported by the National Science Foundation (TeraGrid, grant # TG-MSS080030). The authors acknowledge support from the TeraGrid Advanced Support Program. The authors declare no conflict of interest of any sort.

References

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K et al. (2002) Molecular biology of the cell. Taylor & Francis, New York
2. Herrmann H, Bar H, Kreplak L, Strelkov SV, Aebi U (2007) Intermediate filaments: from cell architecture to nanomechanics. Nat Rev Mol Cell Biol 8(7):562–573
3. Wang N, Butler JP, Ingber DE (1993) Mechanotransduction across the cell surface and through the cytoskeleton. Science 260 (5111):1124–1127
4. Wang N, Stamenovic D (2002) Mechanics of vimentin intermediate filaments. J Muscle Res Cell Motil 23(5–6):535–540
5. Fudge D, Russell D, Beriault D, Moore W, Lane EB et al. (2008) The intermediate filament network in cultured human keratinocytes is remarkably extensible and resilient. PLoS ONE 3(6): e2327
6. Qin Z, Buehler MJ, Kreplak L (2010) A multi-scale approach to understand the mechanobiology of intermediate filaments. J Biomech 43(1):15–22
7. Lewis MK, Nahirney PC, Chen V, Adhikari BB, Wright J et al. (2003) Concentric intermediate filament lattice links to specialized Z-band junctional complexes in sonic muscle fibers of the type I male midshipman fish. J Struct Biol 143(1):56–71
8. Kreplak L, Herrmann H, Aebi U (2008) Tensile properties of single desmin intermediate filaments. Biophys J 94(7):2790–2799
9. Qin Z, Kreplak L, Buehler MJ (2009) Hierarchical structure controls nanomechanical properties of vimentin intermediate filaments. PLoS ONE 4(10):e7294
10. Dahl KN, Kahn SM, Wilson KL, Discher DE (2004) The nuclear envelope lamina network has elasticity and a compressibility limit suggestive of a molecular shock absorber. J Cell Sci 117 (20):4779–4786

11. Wilson KL, Zastrow MS, Lee KK (2001) Lamins and disease: Insights into nuclear infrastructure. Cell 104(5):647–650
12. Aebi U, Cohn J, Buhle L, Gerace L (1986) The Nuclear Lamina is a Meshwork of Intermediate-Type Filaments. Nature 323 (6088):560–564
13. Brenner M, Johnson AB, Boespflug-Tanguy O, Rodriguez D, Goldman JE et al. (2001) Mutations in GFAP, encoding glial fibrillary acidic protein, are associated with Alexander disease. Nat Genet 27(1):117–120
14. Brown CA, Lanning RW, McKinney KQ, Salvino AR, Cherniske E et al. (2001) Novel and recurrent mutations in lamin A/C in patients with Emery-Dreifuss muscular dystrophy. Am J Med Genet 102(4):359–367
15. Bonne G, Mercuri E, Muchir A, Urtizberea A, Becane HM et al. (2000) Clinical and molecular genetic spectrum of autosomal dominant Emery-Dreifuss muscular dystrophy due to mutations of the lamin A/C gene. Ann Neurol 48(2):170–180
16. Broers JLV, Hutchison CJ, Ramaekers FCS (2004) Laminopathies. J Pathol 204(4):478–488
17. Omary MB, Coulombe PA, McLean WH (2004) Intermediate filament proteins and their associated diseases. N Engl J Med 351 (20):2087–2100
18. Parry DAD, Strelkov SV, Burkhard P, Aebi U, Herrmann H (2007) Towards a molecular description of intermediate filament structure and assembly. Exp Cell Res 313(10):2204–2216
19. Sokolova AV, Kreplak L, Wedig T, Mucke N, Svergun DI et al. (2006) Monitoring intermediate filament assembly by small-angle x-ray scattering reveals the molecular architecture of assembly intermediates. Proc Natl Acad Sci USA 103(44):16206–16211
20. Strelkov SV, Herrmann H, Geisler N, Lustig A, Ivaninskii S et al. (2001) Divide-and-conquer crystallographic approach towards an atomic structure of intermediate filaments. J Mol Biol 306(4):773–781
21. Parry DAD (2006) Hendecad repeat in segment 2A and linker L2 of intermediate filament chains implies the possibility of a right-handed coiled-coil structure. J Struct Biol 155(2):370–374
22. Strelkov SV, Herrmann H, Geisler N, Wedig T, Zimbelmann R et al. (2002) Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. EMBO J 21(6):1255–1266
23. Rafik ME, Doucet J, Briki F (2004) The intermediate filament architecture as determined by X-ray diffraction modeling of hard alpha-keratin. Biophys J 86(6):3893–3904
24. Luca S, Yau WM, Leapman R, Tycko R (2007) Peptide conformation and supramolecular organization in amylin fibrils: Constraints from solid-state NMR. Biochemistry 46(47):13505–13522
25. Goldie KN, Wedig T, Mitra AK, Aebi U, Herrmann H et al. (2007) Dissecting the 3-D structure of vimentin intermediate filaments by cryo-electron tomography. J Struct Biol 158(3):378–385
26. Huang A, Stultz CM (2007) Conformational sampling with implicit solvent models: Application to the PHF6 peptide in tau protein. Biophys J 92(1):34–45
27. Zhou RH, Berne BJ (2002) Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? Proc Natl Acad Sci USA 99(20):12777–12782
28. Nymeyer H, Garcia AE (2003) Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. Proc Natl Acad Sci USA 100(24):13934–13939
29. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Protein Struct Funct Genet 35(2):133–152
30. Lazaridis T, Karplus M (1997) "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 278(5345):1928–1931
31. Best RB, Merchant KA, Gopich IV, Schuler B, Bax A et al. (2007) Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. Proc Natl Acad Sci USA 104 (48):18964–18969
32. Paci E, Karplus M (2000) Unfolding proteins by external forces and temperature: the importance of topology and energetics. Proc Natl Acad Sci USA 97(12):6521–6526
33. Paci E, Karplus M (1999) Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. J Mol Biol 288(3):441–459
34. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102 (18):3586–3616
35. Nelson MT, Humphrey W, Gursoy A, Dalke A, Kale LV et al. (1996) NAMD: A parallel, object oriented molecular dynamics program. Int J Supercomput Appl 10(4):251–268
36. Mucke N, Wedig T, Burer A, Marekov LN, Steinert PM et al. (2004) Molecular and biophysical characterization of assembly-starter units of human vimentin. J Mol Biol 340(1):97–114
37. Keten S, Buehler MJ (2008) Geometric confinement governs the rupture strength of H-bond assemblies at a critical length scale. Nano Lett 8(2):743–748
38. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99
39. Guzman C, Jeney S, Kreplak L, Kasas S, Kulik AJ et al. (2006) Exploring the mechanical properties of single vimentin intermediate filaments by atomic force microscopy. J Mol Biol 360 (3):623–630