Consistent View of Protein Fluctuations from All-Atom Molecular Dynamics and Coarse-Grained Dynamics with Knowledge-Based Force-Field

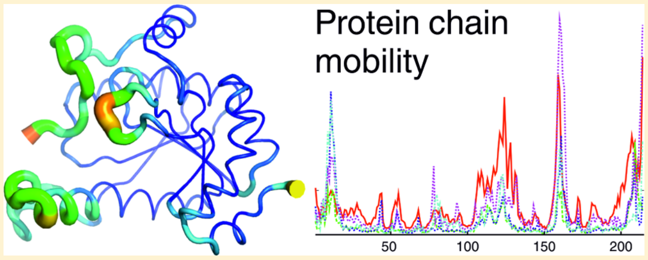
Michal Jamroz,[†] Modesto Orozco,[‡,¶] Andrzej Kolinski,[†] and Sebastian Kmiecik*[,†]

[†]Laboratory of Theory of Biopolymers, Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland

[‡]IRB - BSC Joint Research Program in Computational Biology, Institute for Research in Biomedicine, Josep Samitier 1-5, Barcelona 08028, Spain

[¶]Department of Biochemistry, Universitat of Barcelona, Gran Via de les Corts Catalanes, 585 08007 Barcelona, Spain

Supporting Information



ABSTRACT: It is widely recognized that atomistic Molecular Dynamics (MD), a classical simulation method, captures the essential physics of protein dynamics. That idea is supported by a theoretical study showing that various MD force-fields provide a consensus picture of protein fluctuations in aqueous solution [Rueda, M. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 796−801]. However, atomistic MD cannot be applied to most biologically relevant processes due to its limitation to relatively short time scales. Much longer time scales can be accessed by properly designed coarse-grained models. We demonstrate that the aforementioned consensus view of protein dynamics from short (nanosecond) time scale MD simulations is fairly consistent with the dynamics of the coarse-grained protein model - the CABS model. The CABS model employs stochastic dynamics (a Monte Carlo method) and a knowledge-based force-field, which is not biased toward the native structure of a simulated protein. Since CABS-based dynamics allows for the simulation of entire folding (or multiple folding events) in a single run, integration of the CABS approach with all-atom MD promises a convenient (and computationally feasible) means for the long-time multiscale molecular modeling of protein systems with atomistic resolution.

1. INTRODUCTION

Protein folding is a very complex process involving very fast local dynamics and long-time scale rearrangements of a large number of atoms. Local fluctuations (side-chains, loops) occur in picoseconds, while global rearrangements (folding/unfolding) require typically milliseconds, even for small globular proteins. No experimental or simulation technique is able to embrace all spatial and temporal scales relevant to process description.[1,2] Thus, complete characterization of the folding process requires proper integration of data from a variety of experimental and computational methods. Recent examples of such integrative characterization involve a description of the smallest systems and time scales[3] as well as large macromolecular machines in motion.[4]

As noted above, folding, and in fact most relevant biological processes involving protein conformational changes, takes place on large time scales (between 10 microseconds and milliseconds or even hours), making most of them inaccessible to atomistic MD simulation. Supercomputer efforts in the past few years established the limit of such simulations to be around 10 microseconds of simulated biological time.[5] Just very recently the 1-millisecond barrier was broken by the Shaw group thanks to a custom-built supercomputer.[6] The 1-ms simulation of folded protein BPTI (58 residues) revealed distinct separation of time-scales: hopping between structurally different conformational states on time scales of the order of 10 microseconds, whereas local relaxations occurred on a time scale at least 1000 times faster. The fast relaxations were found to originate primarily from side chain motions, whereas the slow relaxations corresponding to transitions between well separated basins originated mostly from backbone motions.[6] Shaw's group also succeeded in modeling the folding pathway of a 35-residue protein[6] and later continued folding simulation studies of larger and more complex fast-folding proteins.[7] The atomic MD simulations (over periods ranging between 100 microseconds and 1 ms) of 11 out of the 12 structurally diverse proteins studied (ranging from 10 to 80 residues) resulted in spontaneous and repeated folding to their experimentally determined native structures. Interestingly, for most cases, folding proceeded along a single, dominant route, where additional structural elements were formed in a well-defined sequence.[7] What is important is that these unique simulations (with respect to protein size and simulation time) were performed using a single force-field that was able to consistently fold a substantial number of proteins, representing major

Received: October 3, 2012
Published: December 3, 2012

structural classes, to their native states. This result suggests that current MD force-fields may be accurate enough for conducting long time-scale MD simulations. However, another study of the same group, using different force-fields to folding of the villin headpiece,[8] showed that even all studied force-fields were able to fold the protein with folding rates consistent with the experiment, the observed folding pathways depended on the choice of the force-field and the properties of the unfolded state were substantially different among various force-fields. Importantly, a number of other studies (applying atomistic MD and explicit representation of water molecules) confirmed a possibility to fold a protein into its native tertiary structure[6,9−13] and also the inconsistency of different force-fields in the description of a folding pathway.[6,14,15]

While MD simulations of large structural rearrangements (such as entire folding processes) showed to be force-field dependent, the simulations of near-native dynamics seem to be essentially force-field independent, as shown by Orozco and colleagues.[16] The authors examined the consistency of different force-fields in the description of near-native protein dynamics (by state-of-the-art atomistic MD simulations with explicit water). The analysis revealed that most of the dynamics behavior is force-field independent. The four most popular force-fields were used: AMBER[17,18] (A), CHARMM[19,20] (C), GROMOS[21,22] (G), and OPLS[23,24] (O), in a massive supercomputer project for proteins with different folds. MD trajectories from the different force-fields provided a consensus picture of near-native protein dynamics by classical atomic MD in conditions close to physiological.[16] In this work, we use these trajectories as the reference data for comparison with our simulations conducted by a coarse-grained protein model with stochastic dynamics and statistical potentials − the CABS model. This work is a continuation of our previous studies of testing the capability of the CABS model which are successful examples of protein folding simulations from fully denatured to the near-native state.[25−29]

2. MATERIALS AND METHODS

2.1. Protein Data Set and MD Data. We used all the currently available MD trajectories from the Rueda et al.[16] MD dynamics analysis deposited in the microMoDEL subset of the MoDEL database.[30] The protein data set is listed in Table S1. For all the proteins in the data set 10-ns simulation MD runs were performed with explicit water (the TIP3P water model was used for A, C, and O simulations, and the SPC water model for G simulations) at constant pressure (1013.25 hPa) and temperature (300 K) using standard coupling schemes (the same in all cases).[16]

Experimental mobility profiles (Figure 2 and Figure S2) were derived from crystallographic B-factors ($\langle R^2 \rangle_i = (3B_i)/(8\pi^2)$, where B is the B-factor) or multimodel NMR structures (calculated in the same way as for trajectories, see eq 1). In the cases of NMR solved structures: 1BSN, 1SDF, 1IL6, and 1I6F (deposited in the PDB database as a single model), equivalent multimodel PDB data was used (1BSH, 2SDF, 2IL6, and 1I6G, respectively), except for 1FVQ for which multimodel data were not available.

2.2. CABS Dynamics. Coarse-grained models, employing united atom representation, offer substantial extension of the time scales of simulated systems compared to those of all-atom models.[2,31−33] The CABS model (described in detail elsewhere[34]) employs coarse-grained representation of a polypeptide chain that uses up to four atoms per residue. These are C^α and C^β atoms and two virtual pseudoatoms: one placed in the center of mass of a side-chain and the other in the center of the C^α−C^α virtual bond (see Figure 1). The CABS force-field is derived from statistical regularities seen in known protein structures, and it includes side-chain−side-chain mean field potentials, coarse-grained models of main chain hydrogen bonds, and local peptide-chain geometric preferences. The solvent effect is accounted for in an implicit fashion through protein structure statistics used in the derivation of the CABS force-field. The dynamics of CABS-based coarse-grained proteins is simulated by a random series of local conformational transitions (controlled by a Monte Carlo method). Thus, very short-time dynamics, relevant to small-distance local geometric changes, is not physical. However, long series of such random local transitions (modulated by the model force-field) define realistic long-time dynamics, as shown in previous reports.[25−29] Apart from the application to protein dynamics, it is worth noting that the CABS model proved very efficient and accurate in numerous protein structure prediction applications: *de novo* or comparative modeling (e.g.: ranked the best, or one of the best, among other approaches in CASP6 blind prediction experiments[35]) or loop modeling.[36] Importantly, the spatial resolution of CABS predictions enables computationally inexpensive conversion to realistic all-atom models (as shown in the application to high-resolution structure prediction[26] or all-atom description of a folding pathway[27]).

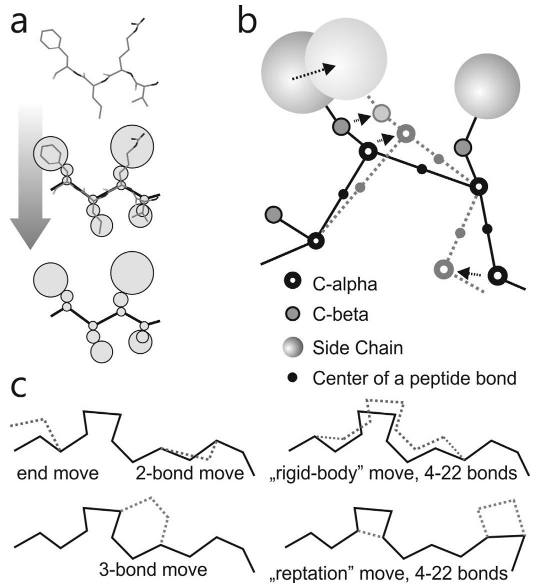


Figure 1. CABS model overview: (a) reduced representation, (b) single C-alpha kink move, (c) schematic illustration of larger scale moves of the Monte Carlo scheme.

2.3. CABS Simulation Setup. For the whole protein data set, more than a hundred simulation setups were performed to test various temperature values, scaling coefficients of force-field components, and versions of distance restraints (unrestrained simulations were also performed) to find the best correlation coefficient for residue fluctuation profiles with the MD trajectories. The highest Spearman's correlation coefficient was found for the simulations with local native-like restraints put exclusively on pairs of residues under two conditions: (1) the distance between their C^α atoms was smaller than 8 Å, (2) both residues were assigned to belong to secondary structure elements. Therefore, loop regions were completely unrestrained and regions of secondary structure locally only. The applied distance restraints softly penalized the position of restrained residues if their distance differed from the distance in the native structure by more than 1 Å.

2.4. Analysis of MD and CABS Trajectories. MD and coarse-grained trajectories were analyzed on the level of C^{α} trace representation to obtain their structural and dynamics characteristics together with their consistency measures.

The mobility of residue i was defined as

$$\left\langle R_i^2 \right\rangle = \frac{1}{N}\sum_{j}^{N}((p_{j,x}^{i} - c_{j,x}^{i})^2 + (p_{j,y}^{i} - c_{j,y}^{i})^2 + (p_{j,z}^{i} - c_{j,z}^{i})^2) \qquad (1)$$

where j - trajectory frame, i - residue index, c - position of the C^{α} atom in the average structure, and N - number of trajectory models.

The global similarity of structures generated by different MD force-fields and the CABS model was obtained by computing the RMSD between all of the snapshots collected in the two trajectories and related similarity index Ω

$$\Omega_{AB} = (\alpha_{AA} + \alpha_{BB})/2\alpha_{AB} \qquad (2)$$

where

$$\alpha_{AB} = \frac{1}{M_A M_B}\sum_{i}^{M_A}\sum_{j}^{M_B}\left(\frac{1}{N}\sum_{t}^{3N}(x_{i,t} - x_{j,t})^2\right)^{1/2} \qquad (3)$$

where N is the number of atoms, M is the number of frames in the compared trajectories, and x is the residue coordinate. The similarity index was computed using the Bioshell package.[37]

The mean-square displacement autocorrelation function $acorr(\tau)$ was defined as

$$acorr(\tau) = \left\langle \left\langle R^2 \right\rangle \right\rangle_{\tau} = \frac{1}{M}\sum_{t}^{M-\tau}\left(\frac{1}{N}\sum_{i}^{N}(\vec{p}_{i,t} - \vec{p}_{i,t+\tau})^2\right) \qquad (4)$$

where $p_{i,t}$ - position of residue i at time t, N - number of protein residues, M - number of trajectory frames, and τ - time frame (Δt).

Global flexibility was shown by the Lindemann's disorder index[38]

$$\Delta_L = \frac{\sqrt{\frac{1}{N}\sum_{i}^{N}\left\langle R_i^2 \right\rangle}}{a'} \qquad (5)$$

where N is the number of atoms, a' is the most-probable nonbonded near-neighbor distance, and $\langle R_i^2 \rangle$ is the fluctuation of the residue i (see eq 1).[38] Lindemann's disorder index was calculated using the PCASuite package.[39]

Commonly used deformation space overlap was defined using root-mean-square inner product γ[40]

$$\gamma_{AB} = \frac{1}{n}\sum_{i,j}^{n}(\vec{v}_i^{A}\cdot\vec{v}_j^{B})^2 \qquad (6)$$

where A and B index the two compared methods, i and j index the eigenvectors (ranked on the basis of their contribution to structural variance), and n is the minimum number of eigenvectors needed to explain 90% of total variance.

Deformation space overlap was defined using root-mean-square inner product "s overlap"[41]

$$s(A, B) = 1 - \frac{d(A, B)}{\sqrt{\operatorname{tr} A + \operatorname{tr} B}} \qquad (7)$$

and

$$d(A, B) = [\sum_{i}^{n}(\lambda_i^{A} + \lambda_j^{B}) - 2\sum_{i,j}^{n}\sqrt{\lambda_i^{A}\lambda_j^{B}}(\vec{v}_i^{A}\cdot\vec{v}_j^{B})^2]^{1/2} \qquad (8)$$

where A and B index covariance matrices of the two compared methods, tr is the trace of a matrix, λ are index eigenvalues, and v are index eigenvectors.

This measure has several advantages over the commonly used subspace overlap[41] (the overlap between the subspaces of the first n_A and n_B eigenvectors of matrix A and B, employed also in the study by Rueda et al.[16]) which depends strongly on n_A and n_B and ignores the eigenvalues. "s overlap" metric takes into account differences between eigenvectors with small and large eigenvalues and treats more correctly degenerate subspaces.

3. RESULTS AND DISCUSSION

3.1. Maximizing MD and CABS Convergence. We started the CABS simulations of the proteins with the optimization of CABS parameters (simulation time, temperature) to obtain the best possible convergence with the available MD data[16] (see Materials and Methods). As a convergence criterion we used the average Spearman's correlation coefficient (r_s) for residue mobility between different MD force-fields and the CABS model. The residue mobility reflects the oscillations of the C^{α} atom of a residue around its mean position (averaged over the whole trajectory, see eq 1).

The highest mean correlations for the completely unrestrained simulations (average over all simulations) between CABS and A, C, G, and O force-fields were the following: 0.62, 0.61, 0.64, 0.63, respectively (see Table S3 for individual protein values). This level of similarity to all-atom MD were also recently achieved by other prediction methods: Support Vector Regression[42] and Gaussian Network Model[43] (mean correlation coefficients respectively: 0.67 and 0.64, as given in ref 42).

Further examination of the CABS mobility profiles revealed, in comparison to the MD trajectories, sometimes smaller stability of the secondary structure elements, particularly visible at elevated temperatures due to long-distance and very fast motions of more flexible parts of protein structures. Furthermore, relying on this observation, we attempted to increase the CABS and MD convergence by repeating the optimization of CABS parameters (simulation time, temperature) and introduction of various types of distance restraints (derived from native structures) to avoid any long-distance and very fast motions of protein structure (see the SI text for optimization procedure of CABS parameters for simulations with distance restraints and predictive power test).

The optimization results showed that the same parameters setup as trained on the whole protein set was found when the method was optimized on randomly chosen half of the protein set. The predictive strength of the method is slightly lower when evaluated on the remaining half of the protein set, than as tested on the whole set (r_s = 0.70 and 0.74, respectively).

The optimal parameters setup, which yielded overall the highest mean correlations (on the level of 0.74), were obtained with weak native-like restraints applied only locally and between coordinates belonging to the secondary structure elements (alpha or beta) (see the SI text for the parameters details). Therefore, loop-forming residues remained completely unrestrained (for the restraints description see Materials and Methods). That was for the setup with significantly higher temperature than the optimal in unrestrained simulations described above. Thus, in comparison to the unrestrained simulations (optimal with regard to temperature and simulation time), the optimal restrained ones allowed for the following: (1) enhanced mobility of at least loop fragments (higher temperature increases the overall acceptance rate of the moves in the Monte Carlo scheme), (2) additional stabilization of the secondary structure and its motifs, and (3) overall decrease in CABS fluctuation level (see the mean RMSD in Table S3 for unrestrained and restrained CABS simulations).

The average correlation coefficients for residue mobility between different MD force-fields and the CABS-simulations with the optimal setup found are listed in Table 1(detailed results for each protein are listed in Table S3). Note that, in this manuscript we report the average statistics for the entire protein set (for the most comprehensive comparison of the methods), however the average from the predictive power test (0.7) should be considered as the estimate of the CABS ability to predict fluctuations from the MD (see the SI text for the optimization details). As highlighted above, the mean correlation between CABS and MD force-fields is on the level of 0.7, which is on a slightly lower level with respect to correlations among different MD force-fields (which varied between 0.75 and 0.82). The average similarity between the mobility profiles measured by RMSD (Table 1) shows more significant differences between CABS and MD force-fields (in the range of 2.8−3.1 Å) than among different MD force-fields (1.8−2.5 Å) which is due to higher average residue oscillations observed in CABS than in MD simulations. For the examples of the mobility profiles with the highest (1FAS, r_s = 0.86) and the lowest (1PDO, r_s = 0.49) correlation coefficients between CABS and MD, see Figure 2. For the mobility profiles visualized on example 3D structures, see Figure 3. As analyzed by Rueda et al.,[16] there is a good agreement between X-ray B-factors and MD-derived mobility profiles, which is also the case of the similarity between experimental (X-ray or NMR) and CABS derived fluctuations (see Figure 2 and Figure S2).

Table 1. Average Spearman's Correlation Coefficient and Mean RMSD (in Brackets) between MDs (A, C, G, O) and CABS Mobility Profiles[a]

	A	C	G	O
CABS	**0.74 (3.12)**	**0.74 (2.84)**	**0.72 (2.91)**	**0.75 (2.92)**
A	1	0.80 (1.75)	0.78 (2.49)	0.82 (1.76)
C		1	0.75 (2.23)	0.81 (1.59)
G			1	0.75 (2.43)

[a] The mean values for the whole test set are shown. Individual values for each protein are reported in Table S3.

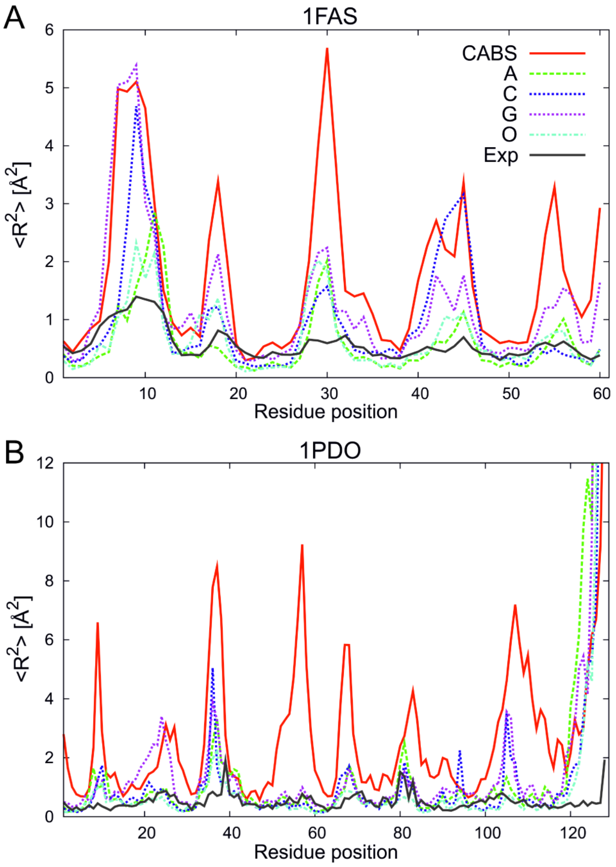


Figure 2. Example mobility profiles by the CABS model, different MD force-fields (A, C, G, O) and experimental data (derived from crystallographic B-factors). The profiles are shown for the following: (A) 1FAS (example of the highest correlation between CABS and MD: 0.86) and (B) 1PDO (example of the lowest correlation between CABS and MD: 0.49). See also 1FAS and 1PDO profiles visualized on 3D structures in Figure 3. The profiles for the remaining proteins in the test set are shown in the SI text (Figure S2) together with Spearman's correlation coefficient values for the whole test set (Table S3).

Recently, two similar studies of the suitability of coarse-grained techniques for the prediction of protein dynamics were conducted by Emperador et al.[44,45] The studies tested two Gō-like models:[45] Brownian dynamics (BD[46]) and discrete molecular dynamics (DMD[47]) with a Gō-like Hamiltonian and a DMD model based on a simple pseudophysical force-field[44] (a hybrid between the physical potential and empirical Gō-like model). The simulation results were compared to fully atomistic MD simulations (the same as used in our study). The comparison showed that the coarse-grained models delivered in general similar protein dynamics features as the atomistic MD simulations. The force-field of the CABS model is not limited to native-like interactions, and, therefore, the results obtained in folding simulations are not assumed *a priori*. It should be noted, that in the case of the restrained simulations (described above) a part of the protein residues forming native-like interactions were weakly restrained toward their native distance (those between or within secondary structure elements), while the rest of them remained completely unrestrained (those between or within loops or between loops and secondary-structure elements).

3.2. MD and CABS Convergence. In addition to residue mobility analysis, we provide below further convergence

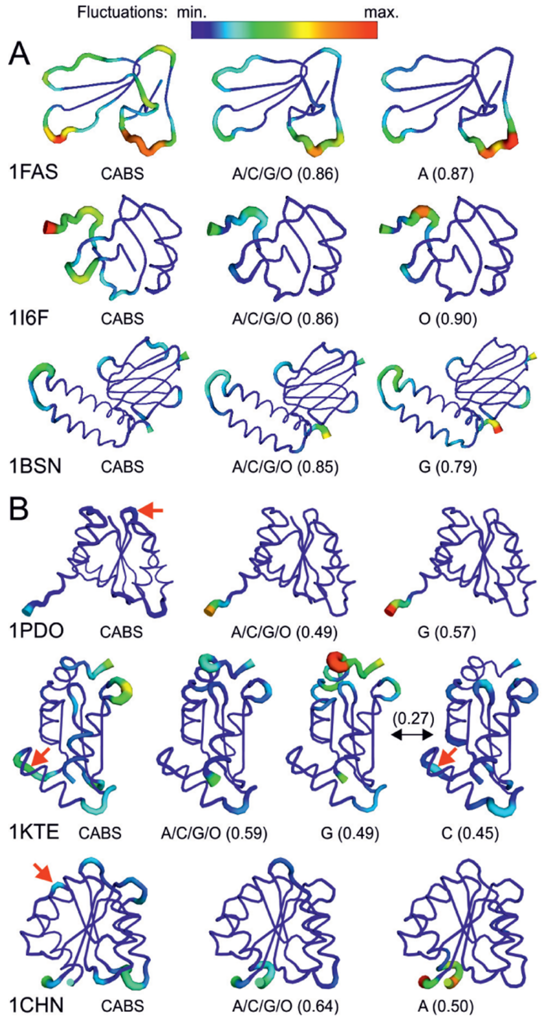

Fluctuations: min. max.
A
1FAS CABS A/C/G/O (0.86) A (0.87)
1I6F CABS A/C/G/O (0.86) O (0.90)
1BSN CABS A/C/G/O (0.85) G (0.79)
B
1PDO CABS A/C/G/O (0.49) G (0.57)
(0.27)
1KTE CABS A/C/G/O (0.59) G (0.49) C (0.45)
1CHN CABS A/C/G/O (0.64) A (0.50)

Figure 3. Example mobility profiles visualized on 3D structures. Profiles are shown for the three proteins with the highest (A) and the three with the lowest (B) correlation values between CABS and MD. For each protein mobility profiles are presented from the CABS model, the four MD force-fields (averaged mobility profile marked as A/C/G/O) and a single MD force-field (A, C, G, or O - always the one which yielded the highest fluctuation value for any single residue). Correlation coefficients for residue mobility between CABS and presented MD simulation are given in brackets. Colors denote fluctuation values scaled from the maximum (red) to minimum fluctuation value (blue) observed in any of the simulations. Protein chain thickness indicates the largest (thick tube) and smallest fluctuations (thin tube), for the given simulation only. Additionally, for the weakest correlation cases (B) protein fragments with the largest contribution to CABS and MD fluctuation inconsistency are marked with red arrows. For 1KTE, an additional fourth fluctuation profile is shown (from C simulations) to indicate significant inconsistency between G and C simulations (r_s = 0.27) and consistency between CABS and C simulations in the marked region. The correlation coefficients and RMSD for the whole test set are given in Table S3.

analysis (for the optimal CABS setup) with different metrics for trajectory comparison. The metrics applied here are the same, or similar, as those used in the study investigating A, C, G, and O force-fields consistency.[16]

The global similarity between the structures obtained by different MD force-fields and CABS is shown in Table 2 according the similarity index Ω. The analysis shows that all simulations produce a similar picture of protein structural diversity, with CABS and G-simulations being slightly less similar to others than A, C, and O simulations to each other. The average effective distance (Ω^{-1}) between pairs of A, C, and O simulations is around 1.4 Å, while that between pairs of CABS and G-simulations with others is around 1.7 Å . The examination of average divergences between different types of simulations (α_{AB} in eq 3) shows that the largest deviations are found between CABS and MD simulations (3.3−3.5 Å), while among MD force-fields the divergences are in the range of 2.2−2.9 Å (the largest for G-simulations).

Table 2. Ω Similarity Index between MDs (A, C, G, O) and CABS Simulations[a]

	A	C	G	O
CABS	**0.6**	**0.6**	**0.6**	**0.6**
A	1.0	0.7	0.6	0.7
C		1.0	0.6	0.7
G			1.0	0.7

[a] Ω = 1 indicates that the simulations sample identical conformational space (in terms of pair-cross RMSD between trajectory structures), while Ω close to zero means that structural diversity is very high.

The CABS trajectories appeared to be different from the different MDs and most similar to G by way of the average Lindemann's disorder index. The index provides a global measure of protein flexibility compared with that of macroscopic solids or liquids[38] (see eq 5). The average Δ_L indexes are as follows: 0.21 ± 0.03 for O; 0.22 ± 0.03 for A, C; 0.24 ± 0.03 for G; and 0.26 ± 0.03 for CABS trajectories. The slight difference in the calculated Δ_L compared to the data presented in ref 16 (average Δ_L = 0.28 ± 0.06) may result from considering only C^α atoms here, with more flexible portions of proteins (such as exposed side chains for which the Δ_L found[16] was 0.38 ± 0.07) being neglected.

Furthermore, we computed the overlap of deformation space (indicating similarity between the modes of two trajectories) using γ and s overlap (see eq 6 and eq 7). The similarity indexes presented in Table 3 indicate the same level of deformation space overlap between CABS and MD as among different MDs. The complexity of the deformability space (measured as the minimum number of eigenvectors needed to explain 90% of total variance) is higher in the case of CABS simulations than different MDs (see Figure 4). This is a similar observation to that shown in the study of coarse-grained BD and DMD models (already mentioned above), indicating that essential movements from coarse grained models might not be accurate individually, but when considered together (in the

Table 3. Average Deformation Space Overlaps γ (First Number) and s (After the Slash Number) between MDs (A, C, G, O) and CABS Simulations[a]

	A	C	G	O
CABS	**0.6/0.3**	**0.6/0.4**	**0.6/0.4**	**0.6/0.4**
A	1.0/1.0	0.6/0.4	0.6/0.3	0.7/0.4
C		1.0/1.0	0.6/0.4	0.7/0.4
G			1.0/1.0	0.6/0.4

[a] Similarity index γ was computed for the minimum number of eigenvectors required to explain the 90% of variance. Note that when the compared trajectories span the same conformational space, the overlap value is equal 1; when the overlap value is zero, the sampled spaces are completely orthogonal (γ and s indices are described in the Materials and Methods, see eqs 6 and 7).

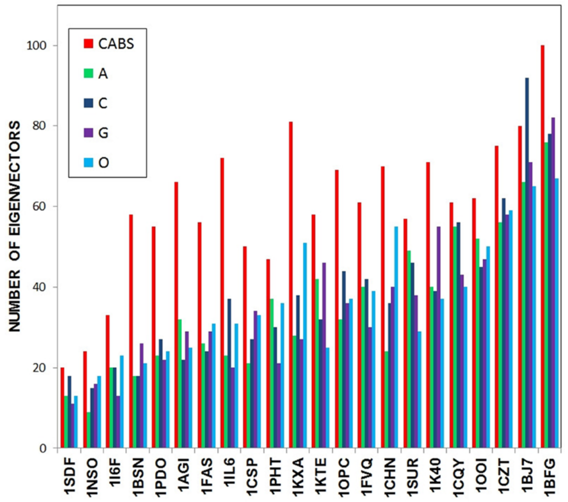


Figure 4. Number of essential modes required to explain 90% of variance (for each protein from the set), using CABS (shown in red) and different MDs (A, C, G, O - shown in green, blue, purple, and cyan, respectively). The proteins are listed according to the average value of essential modes for A, C, G, and O.

essential deformation space) they provide a similar description to that obtained by MD.[45] Interestingly, the similarity index γ between MD and CABS observed in our study (for 90% of the essential space, Table 3) is on a similar level but slightly higher (0.59) than the same index between MD and coarse grained BD and DMD models (0.51 and 0.55, respectively) observed in the Emperador et al.[45] study.

3.3. Diffusion Properties.

Protein folding dynamics can be described as diffusion on a free energy landscape (considered as a function of one or a few chosen reaction coordinates).[48] Diffusive dynamics is characterized by mean square displacement (MSD) linearly growing with time $\langle \Delta x^2(t) \rangle = 2Dt^\alpha$, where $\alpha = 1$ and D is the diffusion coefficient. The nonlinear relationship with time is described as "anomalous diffusion". α exponent values <1 and >1 correspond to subdiffusion and superdiffusion, respectively. Subdiffusion indicates that a system is trapped in local minima and the dynamics "has memory", whereas superdiffusion denotes long jumps of a system in conformational space. We studied the diffusion properties with the MSD autocorrelation function (see eq 4) to compare MD and CABS dynamics. The MSDs of all MD trajectories exhibit a power law dependence on time, with an exponent of around 0.3, just as in the CABS model (see Figure 5). This suppressed diffusion is a consequence of the relatively short time scale of the MD trajectories (the proteins are trapped in their near-native minimum) in atomic MD simulations and soft restraints imposed on protein structures (which force near-native trapping) in the case of CABS modeling.

4. CONCLUSIONS

A great effort has been expended in recent decades to the quest for efficient and accurate algorithms for protein dynamics simulation. Numerous methods have been exercised utilizing various sampling, representation, and force-field models. Atomistic MD, employing Newton's laws of motion and empirical energy functions, emerged as gold standard of protein dynamics simulations. Apart from the improvement of many problems of classical MD techniques, current research seeks for novel computational approaches capable of moving protein simulations to higher coverage of conformational space and better accuracy. We attempted to examine and maximize the consistency of short-time protein dynamics by atomistic MD and the CABS model, two qualitatively different approaches with respect to sampling, representation, and force-field. Considering the conceptual difference between the methods, they both offer a surprisingly similar picture of protein structure flexibility (the average Spearman's correlation coefficient for the fluctuations along protein chains from the protein set is 0.7).

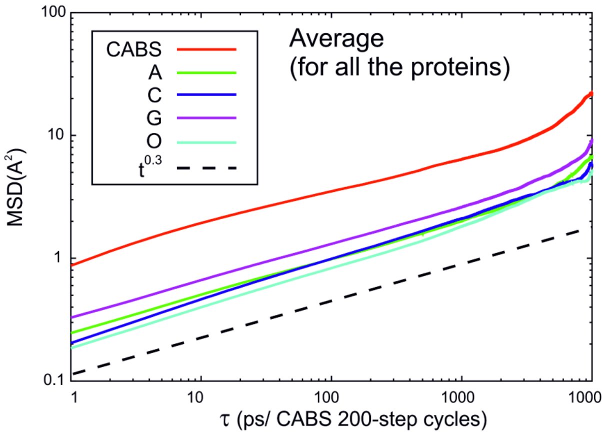


Figure 5. Autocorrelation function - mean square displacement (MSD) (see eq 4) of all protein residues (log scale) at different time intervals, averaged over all proteins studied. A single time unit on abscissa corresponds to 1 ps in MD simulations and 1 CABS time unit (time interval between each frame of the CABS trajectories, set to 200 MC CABS macrocycles).

This work offers promising prospects for the following: (1) fast prediction of MD results by the CABS model (for at least short time scale dynamics) and (2) bridging the CABS and atomistic MD into a single multiscale protocol for the simulation of protein dynamics in atomic resolution (in which MD could be bootstrapped from representative models from the CABS dynamics). Development of such multiscale procedures offers simulation approach of similar quality to atomic MD but many times faster. We roughly estimate CABS dynamics to be 6 × 10^3 cheaper in terms of computational cost than the classical MD (based on Rueda et al.[16] estimations giving on average 3650 CPU hours for single protein simulation from a test set, compared to 0.6 CPU hour by the CABS model).

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Tables: S1 (protein data set), S2 (the five top ranked parameters setups and respective r_s values for the training, the test, and the whole protein set), S3 (Spearman's correlation coefficients and mean RMSD (after the slash) between MDs (A, C, G, O) and CABS mobility profiles), and Figures: S1 (restraints map for 1I6F), S2 (mobility profiles by the CABS model and MD force-fields, for the protein test set). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sekmi@chem.uw.edu.pl.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. Tim Meyer for critical reading of the manuscript. We would like to acknowledge support from a Project operated within the Foundation for Polish Science MPD Programme and TEAM project (TEAM/2011-7/6) cofinanced by the EU European Regional Development Fund operated within the Innovative Economy Operational Program, and from Polish National Science Center (Grant No. NN301071140), and from Polish Ministry of Science and Higher Education Grant No. IP2011 024371.

■ REFERENCES

(1) Russel, D.; Lasker, K.; Phillips, J.; Schneidman-Duhovny, D.; Velazquez-Muriel, J. A.; Sali, A. *Curr. Opin. Cell Biol.* **2009**, *21*, 97−108.

(2) Kmiecik, S.; Jamroz, M.; Kolinski, A. In *Multiscale Approaches to Protein Modeling*; Kolinski, A., Ed.; Springer: New York, 2011; Chapter 12, pp 281−294.

(3) Lin, M. M.; Mohammed, O. F.; Jas, G. S.; Zewail, A. H. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16622−16627.

(4) Zhang, J.; Baker, M. L.; Schroder, G. F.; Douglas, N. R.; Reissmann, S.; Jakana, J.; Dougherty, M.; Fu, C. J.; Levitt, M.; Ludtke, S. J.; Frydman, J.; Chiu, W. *Nature* **2010**, *463*, 379−383.

(5) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008**, *94*, L75−L77.

(6) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341−346.

(7) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.

(8) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47−L49.

(9) Freddolino, P. L.; Schulten, K. *Biophys. J.* **2009**, *97*, 2338−2347.

(10) Ensign, D. L.; Kasson, P. M.; Pande, V. S. *J. Mol. Biol.* **2007**, *374*, 806−816.

(11) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.

(12) Mittal, J.; Best, R. B. *Biophys. J.* **2010**, *99*, L26−L28.

(13) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. *J. Mol. Biol.* **2011**, *405*, 43−48.

(14) Noe, F.; Schutte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19011−19016.

(15) Ensign, D. L.; Pande, V. S. *Biophys. J.* **2009**, *96*, L53−L55.

(16) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 796−801.

(17) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(18) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(19) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

(20) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(21) Ott, K. H.; Meyer, B. *J. Comput. Chem.* **1996**, *17*, 1068−1084.

(22) Hermans, J.; Berendsen, H. J. C.; Vangunsteren, W. F.; Postma, J. P. M. *Biopolymers* **1984**, *23*, 1513−1518.

(23) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(24) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(25) Kmiecik, S.; Kolinski, A. *Biophys. J.* **2008**, *94*, 726−736.

(26) Kmiecik, S.; Gront, D.; Kolinski, A. *BMC Struct. Biol.* **2007**, *7*, 1−11.

(27) Kmiecik, S.; Gront, D.; Kouza, M.; Kolinski, A. *J. Phys. Chem. B* **2012**, *116*, 7026−7032.

(28) Kmiecik, S.; Kolinski, A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 12330−12335.

(29) Kmiecik, S.; Kolinski, A. *J. Am. Chem. Soc.* **2011**, *133*, 10283−10289.

(30) Meyer, T.; D'Abramo, M.; Hospital, A.; Rueda, M.; Ferrer-Costa, C.; Pérez, A.; Carrillo, O.; Camps, J.; Fenollosa, C.; Repchevsky, D.; Gelpí, J. L.; Orozco, M. *Structure* **2010**, *18*, 1399−1409.

(31) Kolinski, A.; Skolnick, J. *Polymer* **2004**, *45*, 511−524.

(32) Liwo, A.; He, Y.; Scheraga, H. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16890−16901.

(33) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57−83.

(34) Kolinski, A. *Acta Biochim. Pol.* **2004**, *51*, 349−371.

(35) Kolinski, A.; Bujnicki, J. M. *Proteins* **2005**, *61*, 84−90.

(36) Jamroz, M.; Kolinski, A. *BMC Struct. Biol.* **2010**, *10*, 1−9.

(37) Gront, D.; Kolinski, A. *Bioinformatics* **2008**, *24*, 584−585.

(38) Zhou, Y.; Vitkup, D.; Karplus, M. *J. Mol. Biol.* **1999**, *285*, 1371−1375.

(39) Meyer, T.; Ferrer-Costa, C.; Perez, A.; Rueda, M.; Bidon-Chanal, A.; Luque, F. J.; Laughton, C. A.; Orozco, M. *J. Chem. Theory Comput.* **2006**, *2*, 251−258.

(40) Hess, B. *Phys. Rev. E* **2000**, *62*, 8438−8448.

(41) Hess, B. *Phys. Rev. E* **2002**, *65*, 031910.

(42) Jamroz, M.; Kolinski, A.; Kihara, D. *Proteins* **2012**, *80*, 1425−1435.

(43) Yang, L.; Song, G.; Jernigan, R. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12347−12352.

(44) Emperador, A.; Meyer, T.; Orozco, M. *J. Chem. Theory Comput.* **2008**, *4*, 2001−2010.

(45) Emperador, A.; Carrillo, O.; Rueda, M.; Orozco, M. *Biophys. J.* **2008**, *95*, 2127−2138.

(46) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, UK, 1987.

(47) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459−466.

(48) Krivov, S. V. *PLoS Comput. Biol.* **2010**, *6*, e1000921.