An Effective Modification to Multiscale Elastic Network Model and Its Evaluation Based on Analyses of Protein Dynamics

Weikang Gong PhD student

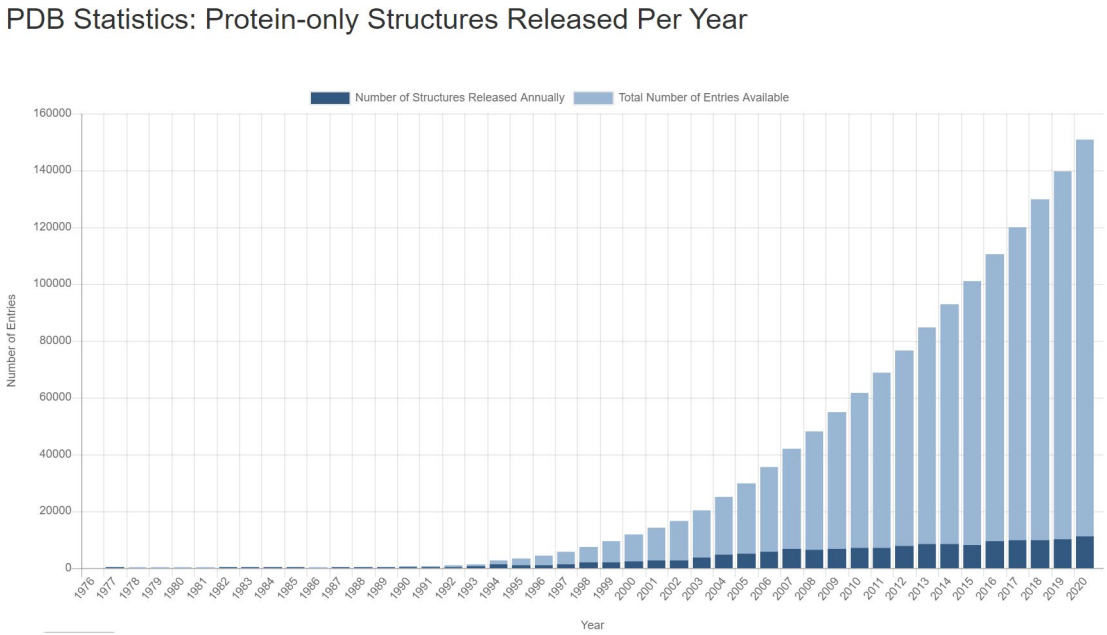
Faculty of Environmental and of Life Sciences

Beijing University of Technology

2020/12/10

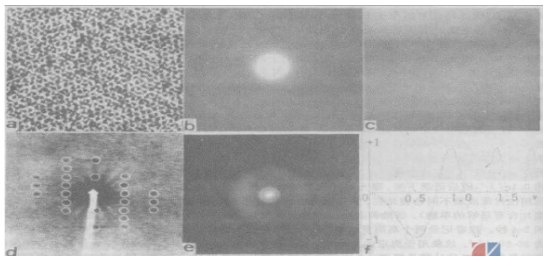
Background

- Protein structural dynamics is intimately related to their functions, which is reflected in many biological processes such as protein-ligand interactions, signal transduction, and assembly of macromolecular machines and allosteric regulation.
- Obtaining accurately protein dynamical characteristics is critical for understanding and deducing their functions.

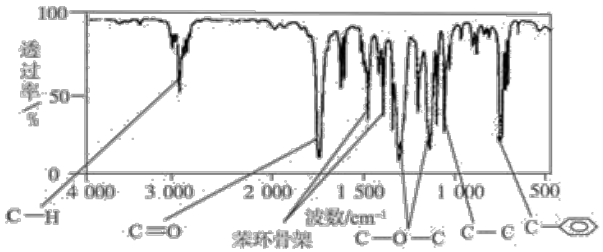


Kaynak *et al. J. Phys. Chem. B*. 2018; Wang *et al. Sci. Rep*. 2018; Cheng *et al. Nat. Struct. Mol. Biol*. 2019; Mikulska-Ruminska *et al. J. Chem. Inf. Model*. 2019; Zhang *et al. Mol. Biol. Evol*. 2019.

It is time-consuming and labor-intensive to study protein dynamics and conformational changes experimentally.



X-ray crystal diffraction



Nuclear magnetic resonance (NMR)

Cryo-electron microscopy (cryo- EM)

Advantages: reliable results;

Shortcoming: time-consuming and labor-intensive.

Kohn *et al. PLoS Comput. Biol.* 2010; Fenwick *et al. Proc. Natl. Acad. Sci. U. S. A.* 2014; Fernandez-Leiro *et al. Nature*. 2016.

Theory

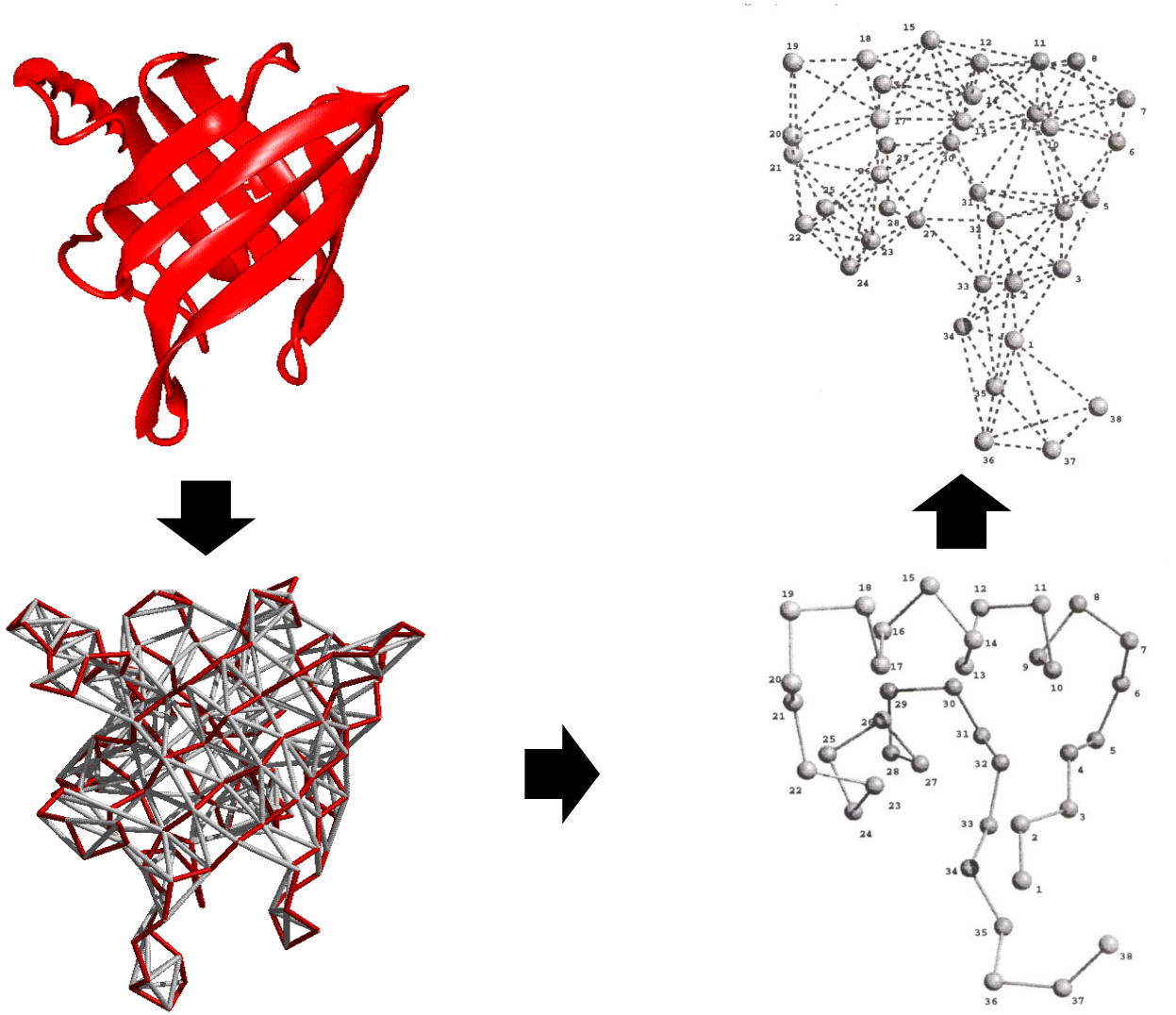
- **Molecular dynamics (MD) simulation provides a useful tool at the atomic level to analyze the mechanical, structural and thermodynamic properties of biomolecules. However, its application requires enormous computer resources, and does not always fully sample the entire conformational space accessible to a protein.**

- **Some coarse-grained methods have been developed, and among them the elastic network model (ENM) is a harmonic potential-based and cost-effective computational method.**

- **The ENM has achieved great success in predicting the large-amplitude collective motion for proteins and even for RNAs. Gaussian network model (GNM) and anisotropic network model (ANM) are the two often-used ENM models.**

Zhang *et al. J. Biomol. Struct. Dyn.* 2019; Tirion *et al. Phys. Rev. Lett.* 1996; Han *et al. Biophys. J.* 2019; Li *et al. J. Chem. Phys.* 2016; Eyal *et al. Bioinformatics*. 2015.

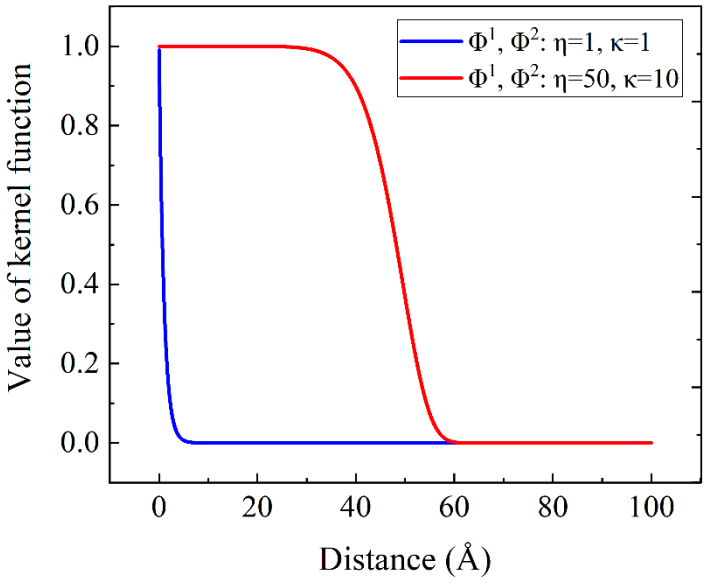
Traditional ENM



An Effective Modification to Multiscale Elastic Network Model (Equally-weighted mENM)

$$\Gamma^{\text{equally-weighted mGNM}} = \sum_n \Gamma_n^{\text{mGNM}}$$

$$\boldsymbol{H}^{\text{equally-weighted mANM}} = \sum_n \boldsymbol{H}_n^{\text{mANM}}$$



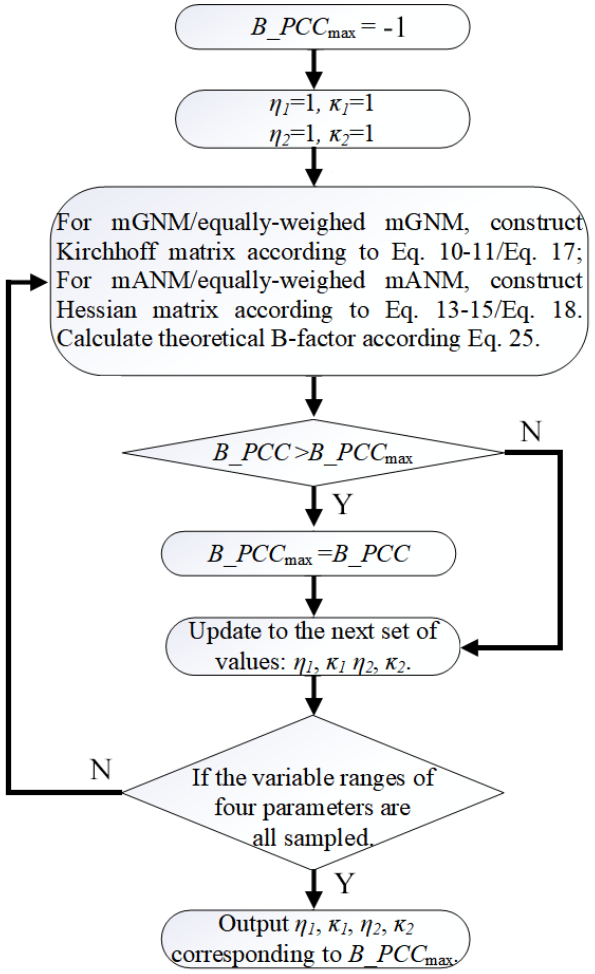


Figure 1. Optimization flowchart of η and κ parameters in mENM and equally-weighted mENM models

Test cases

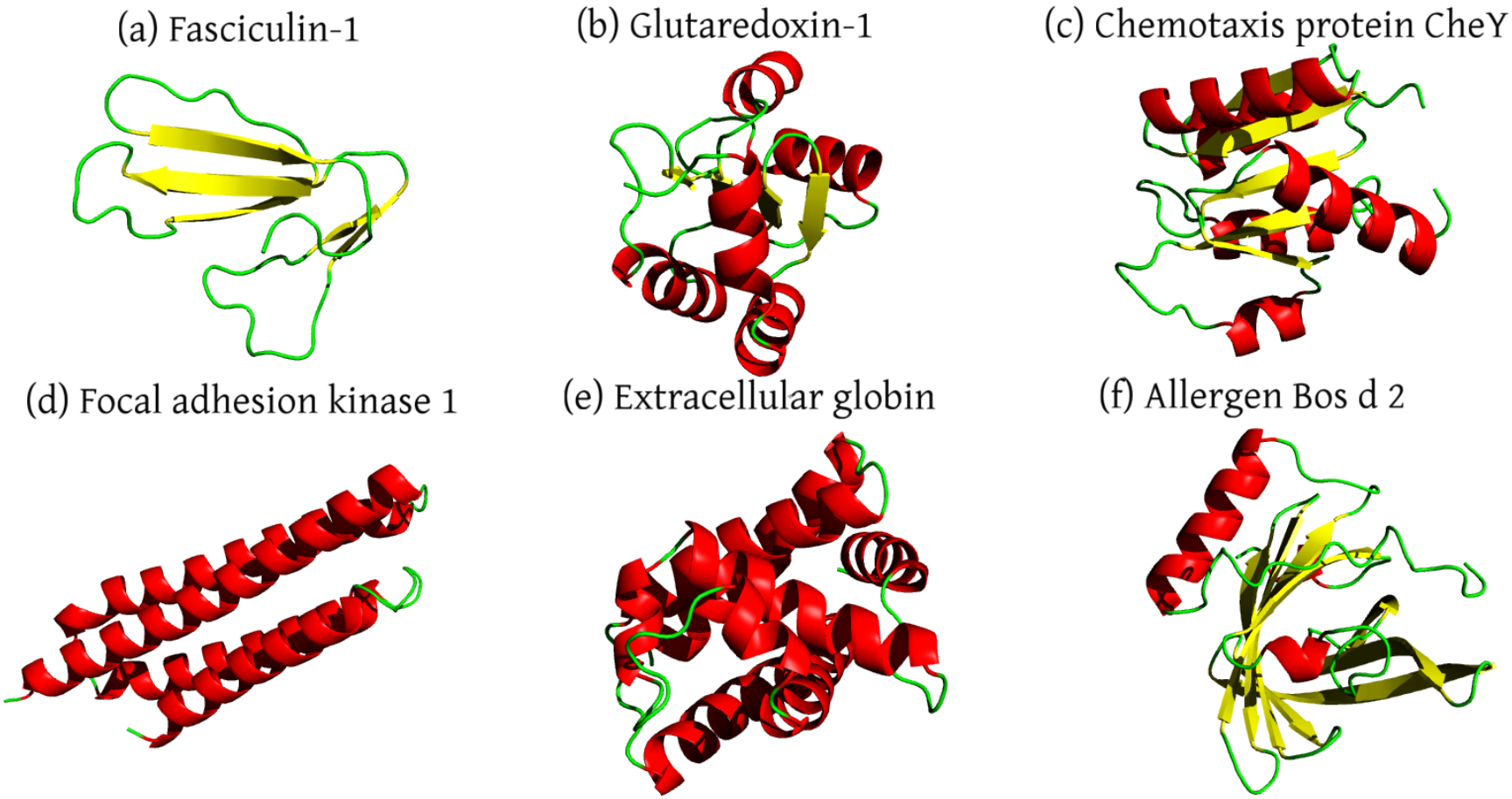


Figure 2. Six test proteins including Fasciculin-1 (a), Glutaredoxin-1(b), Chemotaxis protein CheY (c), Focal adhesion kinase 1 (d), Extracellular globin (e) and Allergen Bos d 2 (f) with PDB IDs being 1FAS, 1KTE, 1CHN, 1K40, 1ASH and 1BJ7, respectively.

Theoretical B-factor calculation

Table 1. *B_PCC* values between experimental and theoretical B-factors calculated by the four kinds of ENM models on the six proteins [a]

Model / PDB ID	traditional ENM		pfENM		mENM		equally-weighted mENM	
	traditional GNM	traditional ANM	pfGNM	pfANM	mGNM	mANM	equally-weighted mGNM	equally-weighted mANM
1FAS	**0.74**	0.69	0.63	0.39	**0.74**	**0.88**	0.73	0.78
1KTE	0.65	0.64	0.66	0.63	0.67	**0.72**	**0.70**	0.64
1CHN	0.66	0.69	0.72	0.71	**0.75**	**0.81**	0.74	0.73
1K40	0.79	0.70	0.72	0.58	**0.82**	**0.80**	0.81	0.78
1ASH	0.72	0.56	0.65	0.56	**0.76**	**0.79**	0.75	0.63
1BJ7	0.69	0.73	0.63	0.66	**0.79**	**0.78**	0.73	0.73

[a] Two highest *B_PCC* values from GNMs and ANMs respectively for each protein is shown in bold.

Comparing DCCMs from ENMs and MD ensembles

Table 2. Best *DCCM_PCC* values between the DCCMs from MD ensembles and four kinds of ENMs for the six test proteins [a]

Model / PDB ID	traditional ENM		pfENM		mENM		equally-weighted mENM	
	traditional GNM	traditional ANM	pfGNM	pfANM	mGNM	mANM	equally-weighted mGNM	equally-weighted mANM
1FAS	0.57	0.43	**0.66**	**0.70**	0.58	0.35	0.58	0.45
1KTE	0.48	0.80	0.61	0.77	0.20	0.17	**0.63**	**0.81**
1CHN	0.71	0.59	0.72	0.68	0.67	0.36	**0.75**	**0.77**
1K40	0.67	0.82	0.76	0.80	**0.78**	0.29	0.68	**0.84**
1ASH	0.58	0.69	**0.65**	**0.71**	0.55	0.32	0.56	0.66
1BJ7	0.56	0.66	**0.78**	**0.78**	0.23	0.34	0.62	0.69

[a] Two highest values obtained from GNMs and ANMs respectively for each protein are shown in bold.

Comparing DCCMs from ENMs and MD ensembles

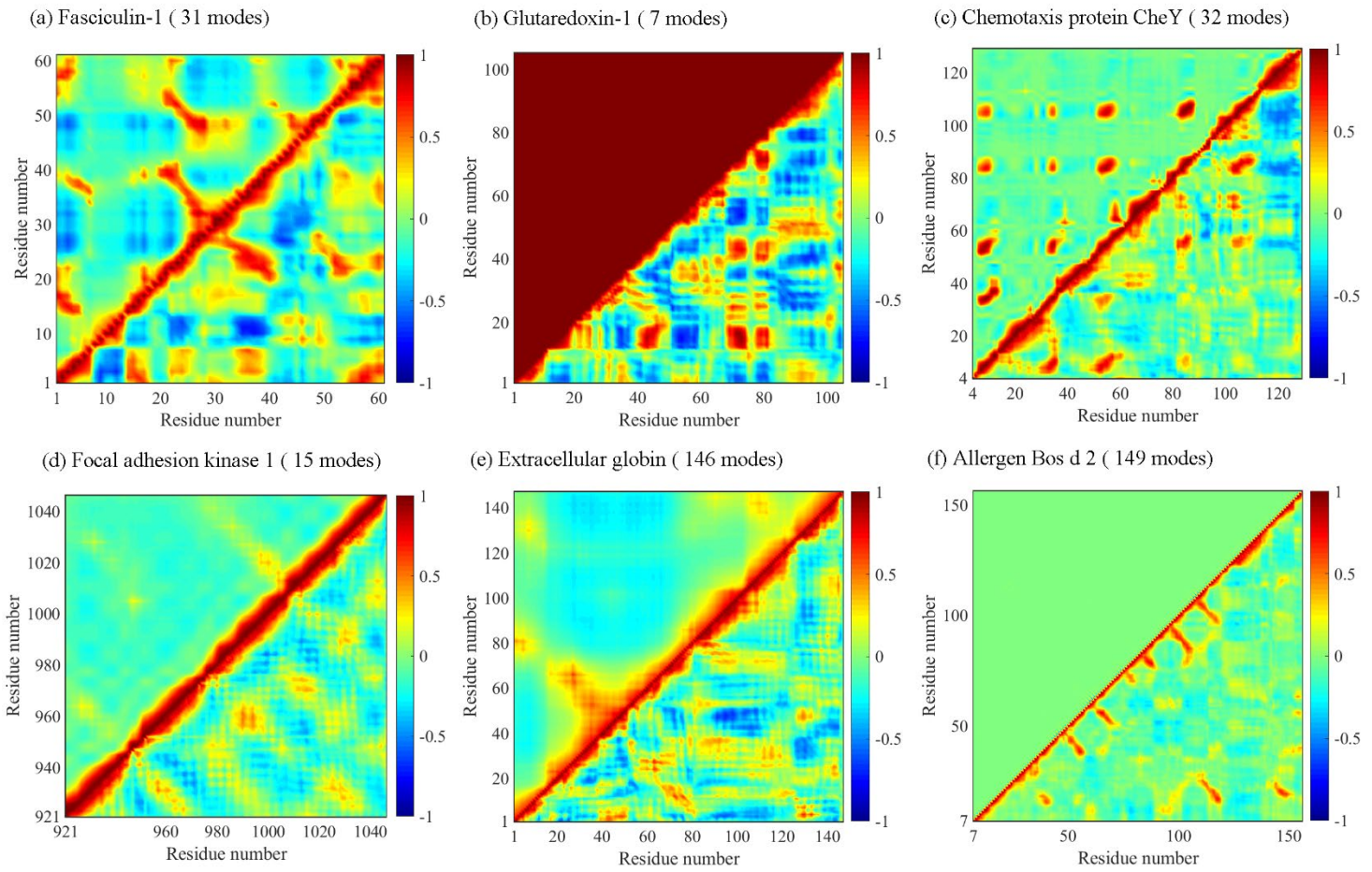


Figure 3. DCCMs obtained from MD ensembles (lower right triangle) and mGNM (upper left triangle) at the best *DCCM_PCC* value (corresponding number of motional modes given in parentheses) for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e) and 1BJ7 (f).

Comparing DCCMs from ENMs and MD ensembles

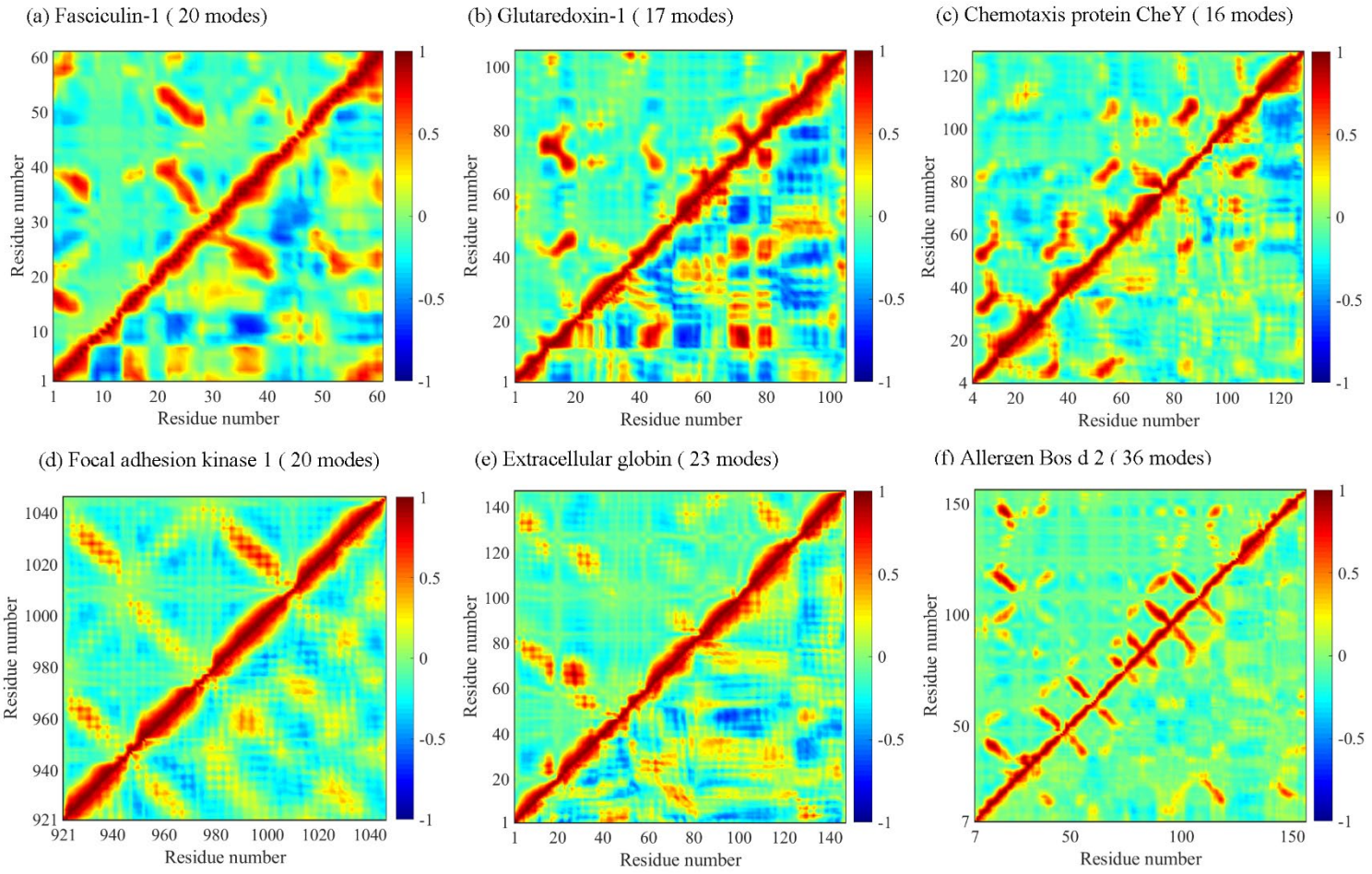


Figure 4. DCCMs obtained from MD ensembles (lower right triangle) and the equally-weighted mGNM (upper left triangle) at the best *DCCM_PCC* value (corresponding number of motional modes given in parentheses) for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e) and 1BJ7 (f).

Comparing DCCMs from ENMs and MD ensembles

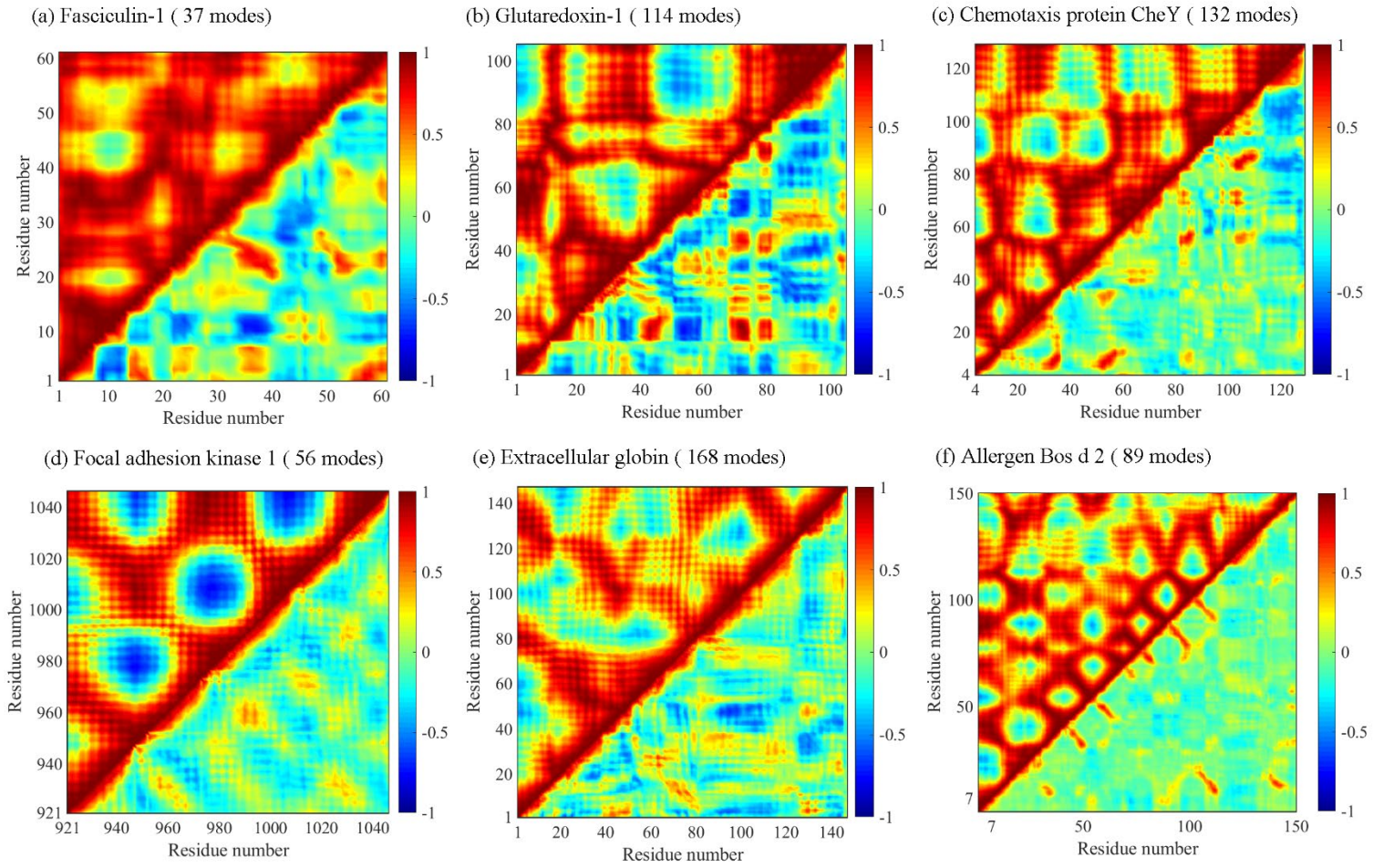


Figure 5. DCCMs obtained from MD ensembles (lower right triangle) and the mANM (upper left triangle) at the best *DCCM_PCC* value (corresponding number of motional modes given in parentheses) for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e) and 1BJ7 (f).

Comparing DCCMs from ENMs and MD ensembles

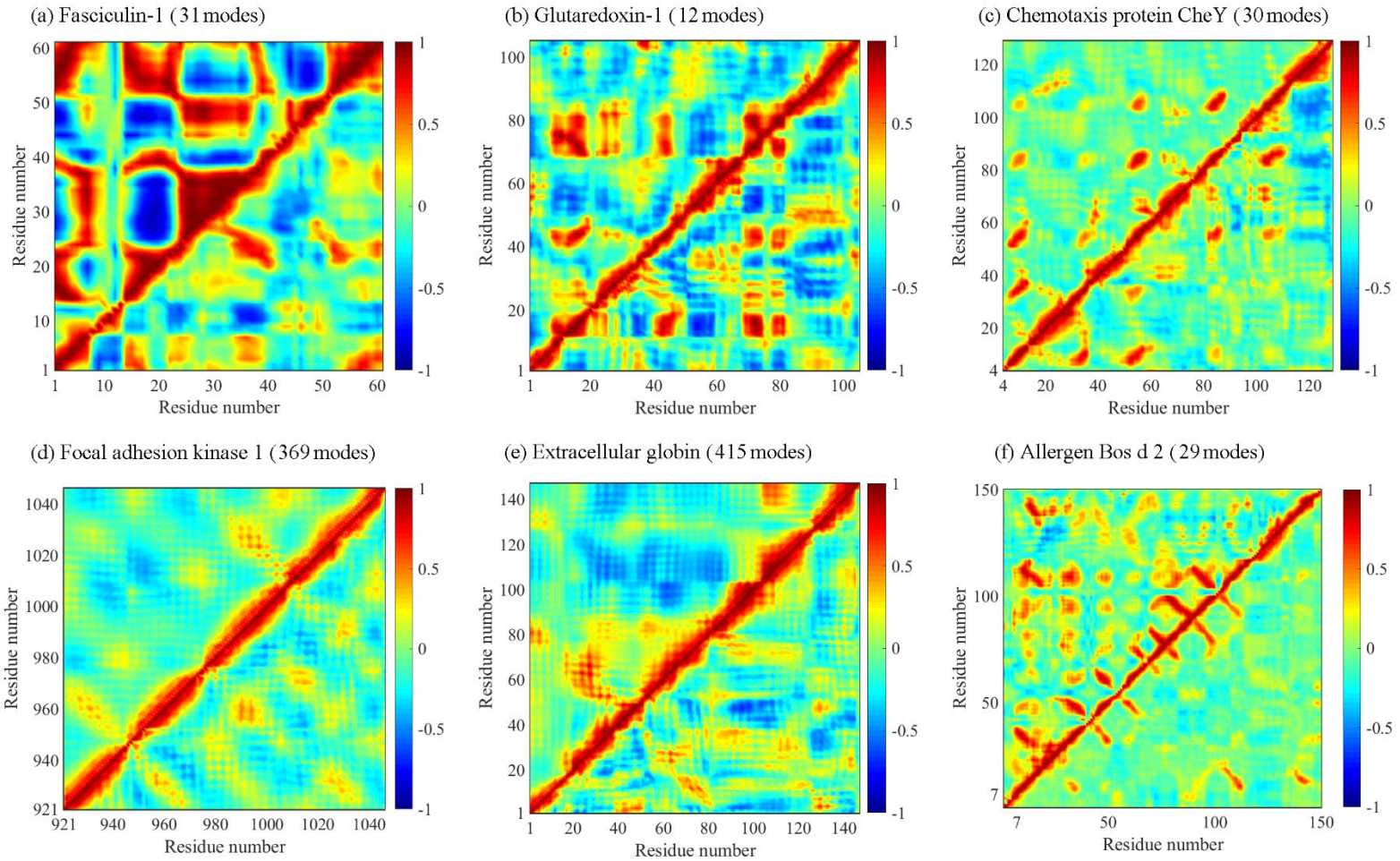


Figure 6. DCCMs obtained from MD ensembles (lower right triangle) and the equally-weighted mANM (upper left triangle) at the best *DCCM_PCC* value (corresponding number of motional modes given in parentheses) for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e) and 1BJ7 (f).

Comparing DCCMs from ENMs and MD ensembles

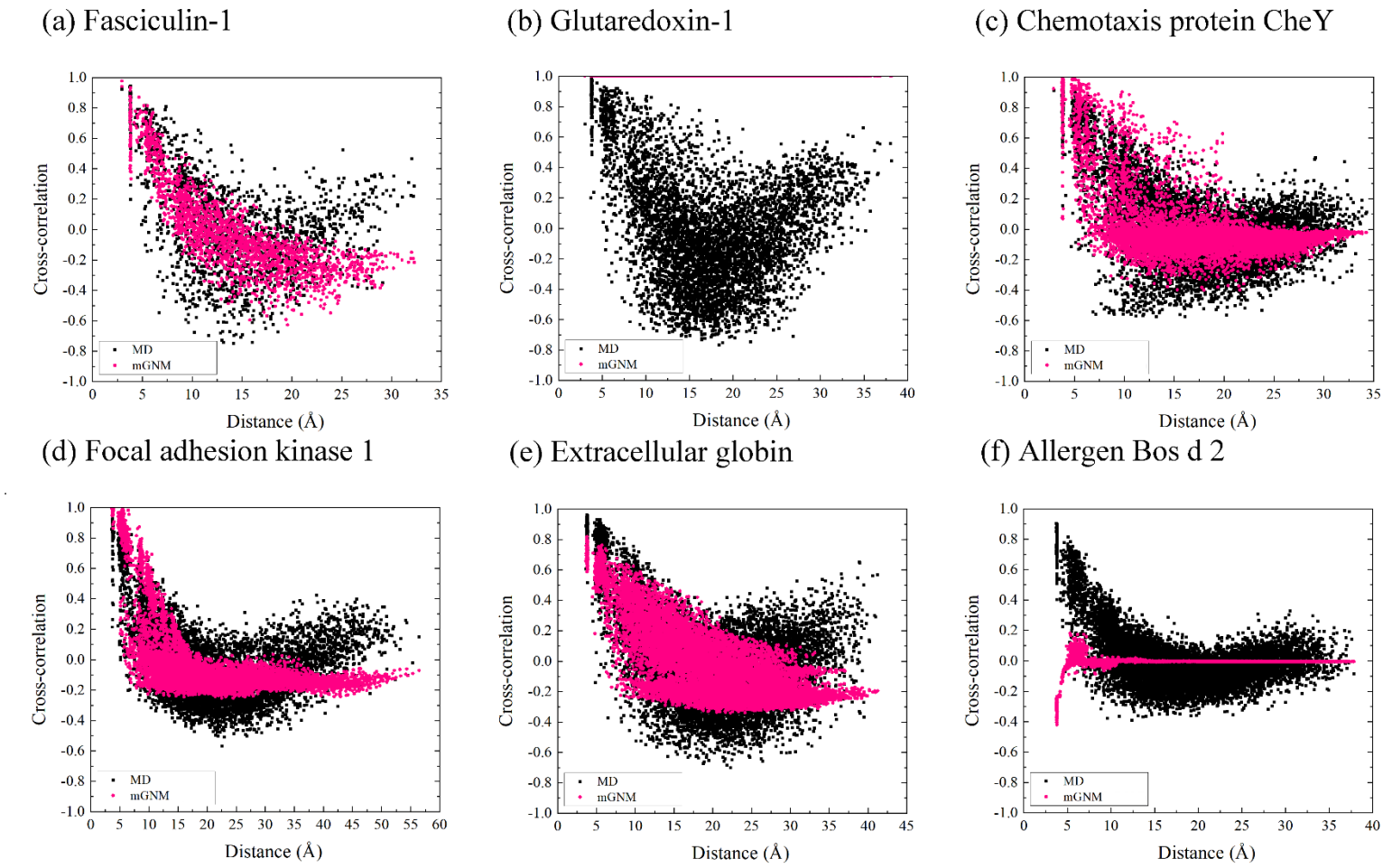


Figure 7. Distributions of dynamical cross-correlations obtained from MD ensembles (black) and mGNM (pink) (at the best *DCCM_PCC* value) with respect to the inter-residue distance for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e) and 1BJ7 (f).

Comparing DCCMs from ENMs and MD ensembles

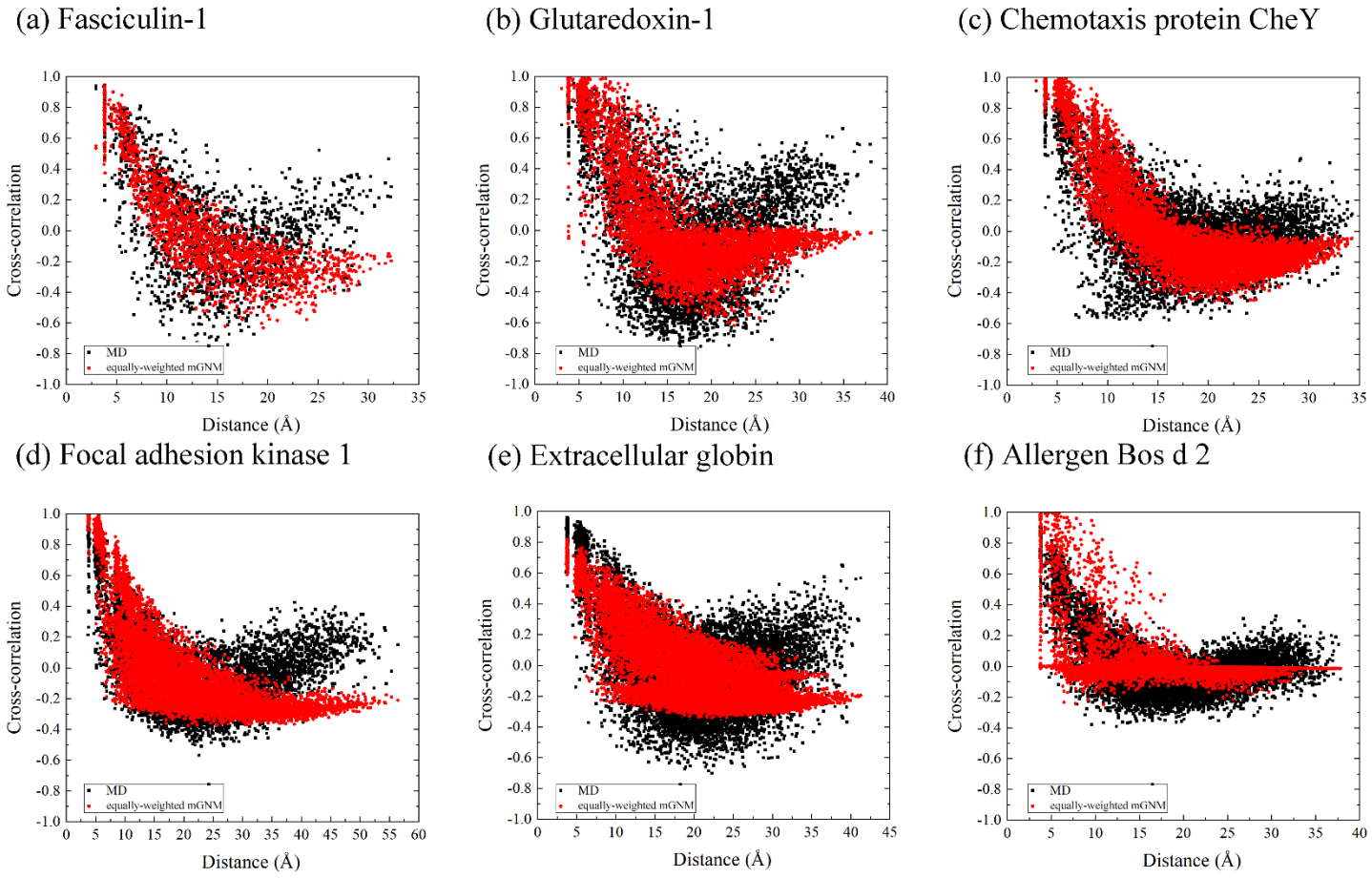


Figure 8. Distributions of dynamical cross-correlations obtained from MD ensembles (black) and equally-weighted mGNM (pink) (at the best *DCCM_PCC* value) with respect to the inter-residue distance for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e) and 1BJ7 (f).

Comparing DCCMs from ENMs and MD ensembles

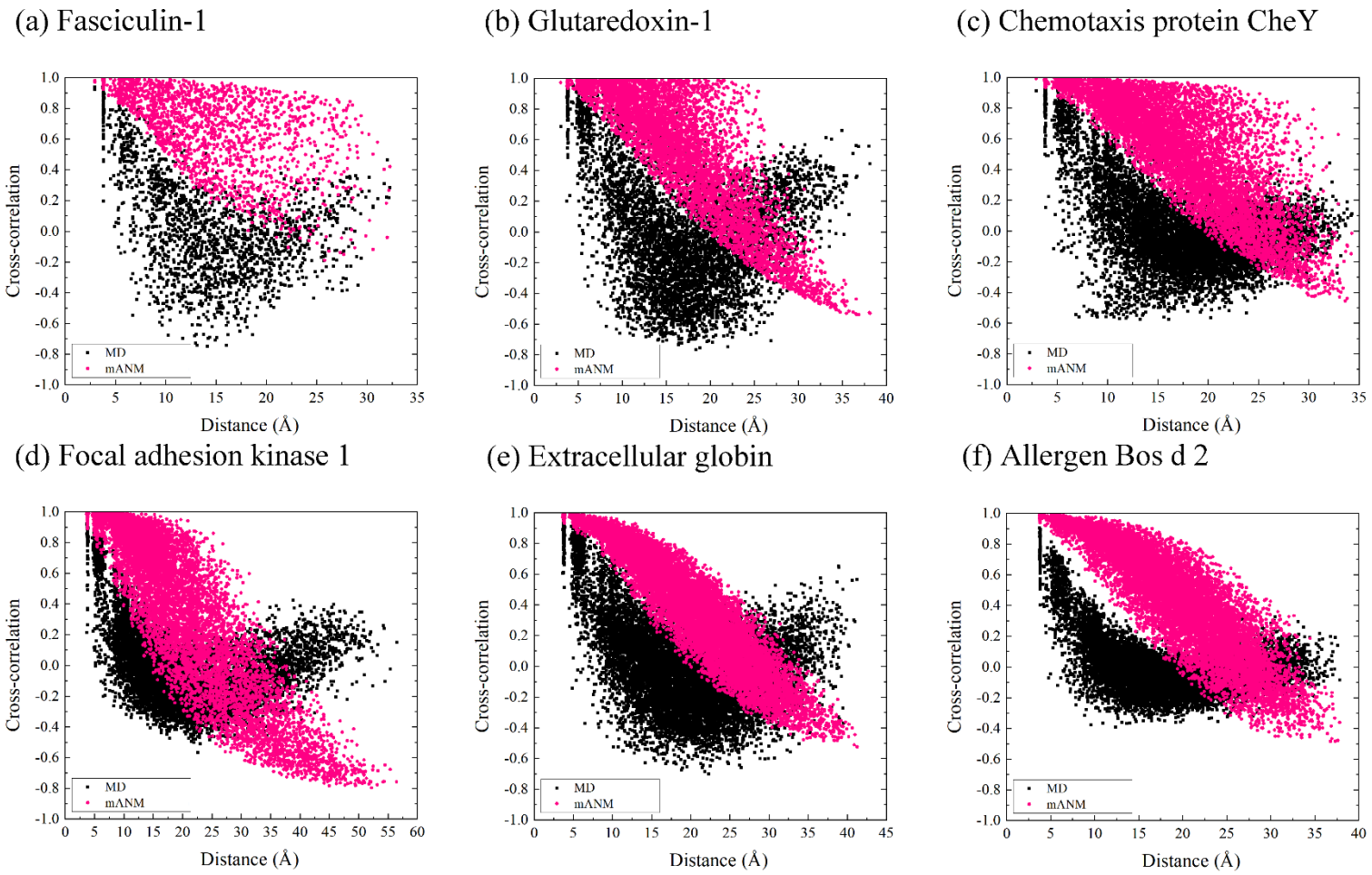


Figure 9. Distributions of dynamical cross-correlations obtained from MD ensembles (black) and mANM (pink) (at the best *DCCM_PCC* value) with respect to the inter-residue distance for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e) and 1BJ7 (f).

Comparing DCCMs from ENMs and MD ensembles

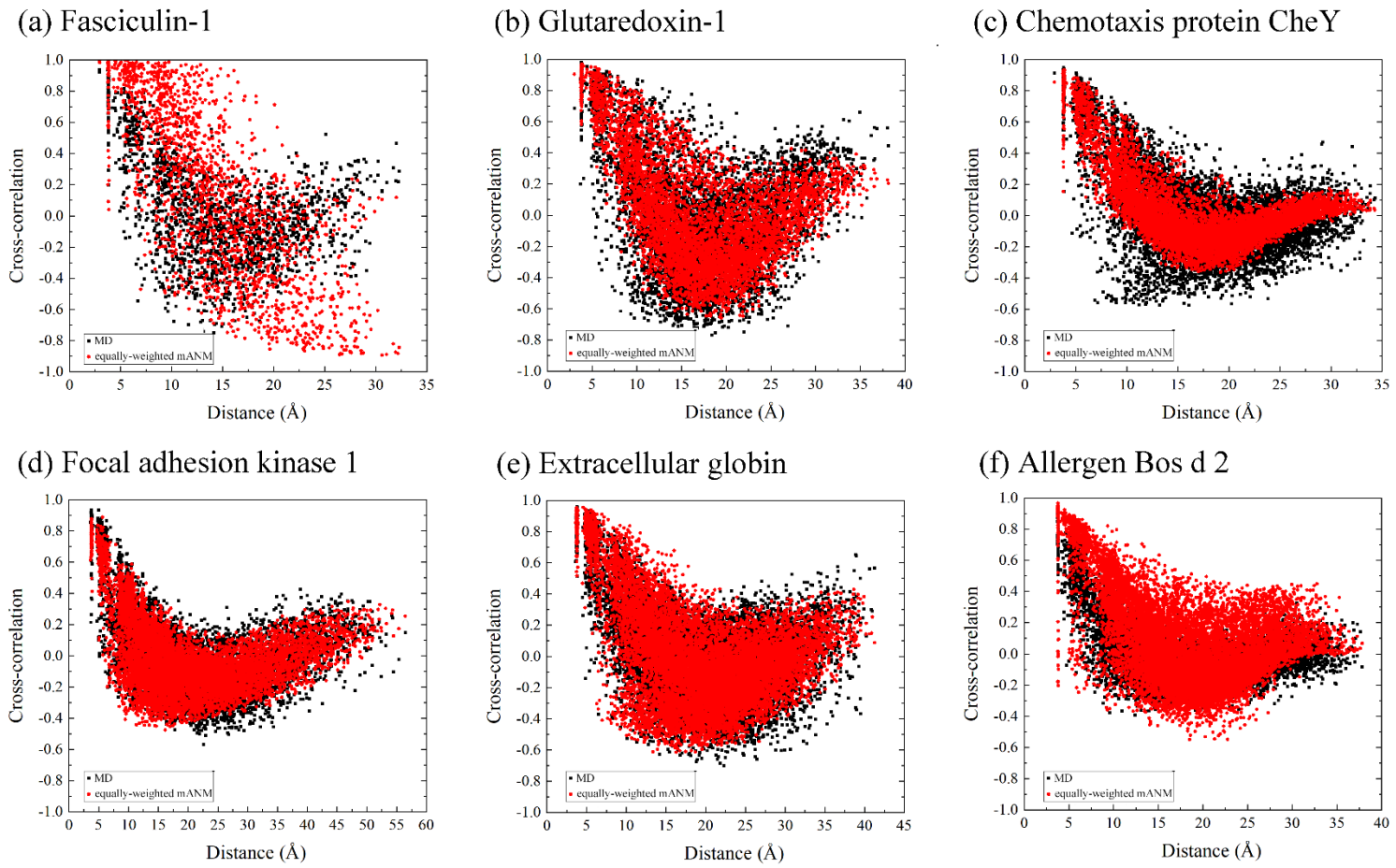


Figure 10. Distributions of dynamical cross-correlations obtained from MD ensembles (black) and equally-weighted mANM (pink) (at the best *DCCM_PCC* value) with respect to the inter-residue distance for proteins 1FAS (a), 1KTE (b), 1CHN (c), 1K40 (d), 1ASH (e) and 1BJ7 (f).

Comparing ANM modes with motions present in MD ensembles

Table 3. Average values of overlaps and *RMSIPs* between motional modes from ANMs and the principle components of motions sampled by MD simulations for the six proteins [a]

Metrics \ Model	traditional ANM	pfANM	mANM	equally-weighted mANM
O_1^{max}	0.31 (0.14)	**0.34** (0.14)	0.24 (0.15)	0.32 (0.10)
O_2^{max}	0.27 (0.08)	**0.29** (0.11)	0.18 (0.12)	**0.29** (0.05)
O_3^{max}	**0.35** (0.09)	0.30 (0.06)	0.19 (0.13)	0.31 (0.07)
CO_1^{20}	0.58 (0.20)	**0.60** (0.16)	0.38 (0.22)	0.59 (0.13)
CO_2^{20}	0.54 (0.12)	0.51 (0.14)	0.34 (0.21)	**0.56** (0.08)
CO_3^{20}	0.59 (0.07)	0.57 (0.09)	0.36 (0.25)	**0.63** (0.07)
$RMSIP_3^{20}$	0.57 (0.10)	0.57 (0.10)	0.36 (0.22)	**0.60** (0.04)
$RMSIP_6^{20}$	0.31 (0.03)	0.31 (0.03)	0.23 (0.09)	**0.32** (0.01)
$RMSIP_{10}^{20}$	0.54 (0.07)	0.54 (0.08)	0.36 (0.21)	**0.57** (0.04)
$RMSIP_{20}^{20}$	0.50 (0.07)	0.50 (0.05)	0.36 (0.20)	**0.53** (0.03)

[a] The highest value for each of the ten metrics is shown in bold. Standard deviations are given in parentheses

Comparing ANM modes with motions present in MD ensembles

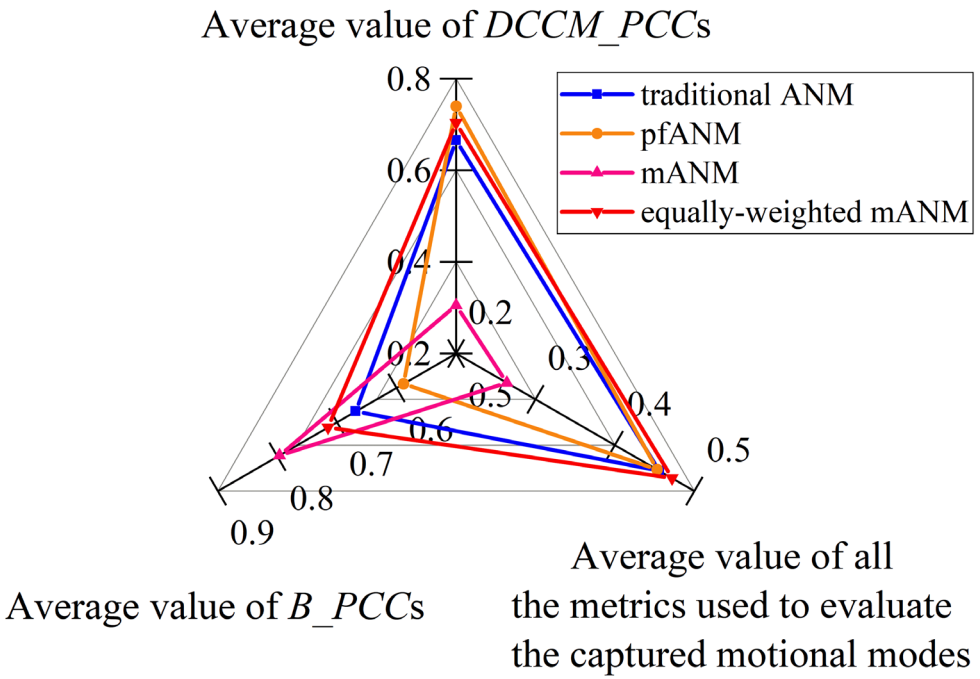


Figure 11. Performance comparison among traditional ANM, pfANM, mANM, and equally-weighted mANM in the calculations of B-factor, DCCM, and motional mode with the average values of B_PCC, DCCM_PCC and all the metrics describing the correlations between motional modes from ANMs and PCs of motions sampled by MD simulations over the six proteins. The three axes extend in the positive direction from the origin. The lines connect the values obtained from the same model. Traditional ANM, pfANM, mANM and equally-weighted mANM are colored in blue, orange, pink, and red, respectively.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31971180, 11474013)

Thanks to my supervisor Professor Chunhua Li and the co-authors of this work : Yang Liu, Yanpeng Zhao, Shihao Wang, Zhongjie Han, Faculty of Environmental and of Life Sciences, Beijing University of Technology

Thanks for your time !