Published in final edited form as:

Protein Pept Lett. 2015 ; 22(7): 654–659.

Recapturing the Correlated Motions of Protein Using Coarse-Grained Models

Yan Lu and Freddie R. Salsbury Jr.

Department of Physics, Wake Forest University, Winston-Salem, NC 27106, USA

Freddie R. Salsbury: salsbufr@wfu.edu

Abstract

Long-range interactions and allostery are important for many biological processes. Increasing numbers of studies, both experimental and computational, show that internal dynamics may play an important role in such behaviors. Investigating the dynamical effects of proteins, however, is a challenging problem using all-atom molecular dynamics because of the length-scales and timescales involved. As a result, coarse-grained models are often implemented. Herein, we use three well-defined coarse-grained models: Go, Martini and Cafemol, and a small model protein Eglin C, which is readily studied via all-atom molecular dynamics, to examine if these coarse grained models can explore the dynamics of Eglin C accurately as well as to see how these models respond to mutations. We found that all three models can recapture the dynamics of Eglin C to a significant extent – where we focus on root-mean square fluctuations and correlated motions as dynamical measures – but that the Cafemol and Go models are superior. The best agreement with all-atom simulations is for structured regions of Eglin C.

Introduction

Allostery is one of the important mechanisms of cell signaling, which is the biological phenomenon where ligand binding or energetic perturbation at one molecular site results in structure or activity change at a second distinct site.¹ Two famous models, MWC (population shift or concerted mechanism)² and KNF (induced fit or sequential mechanism),³ have explained and predicted some experimentally observed aspects of allostery. Nevertheless, theory,⁴ experiments,⁵ and simulations⁶ have showed the existence of purely dynamical-driven allostery.

Several recent experimental examples include negative cooperative binding of cAMP to CAP mediated by conformational entropy in which allostery occurs in the absence of a conformational change,⁵ conformational entropy modulating peptide affinity for a PDZ domain that has been altered at a site distant to the peptide binding site,⁷ and combined NMR dynamic and thermodynamic evidence showing that changes in protein motions may activate allosteric proteins that are otherwise structurally inactive.⁸ Several recent computational papers, also show that allosteric communication can occur dynamically and/or via protein and protein-DNA interfaces. Such as a recent computational study of the

binding of Smad4 protein to DNA molecule in the heteromeric Smad4+DNA+Smad1/3 model inducing allosteric communication from the Smad4-DNA interface to Smad1/Smad3-DNA interface via DNA base-pair helical motions, surface conformation changes and new hydrogen bond formations.⁹ Another set of simulations show that small structural differences between ATP-bound and ATP-free MutS-DNA complex, but there are fluctuation couplings, especially in ATP-bound complex, show that allostery between the nucleotide and DNA binding sites in MutS can occur via dynamical allostery.¹⁰ Other studies show that responses to binding of damaged DNA are mediated by interfaces and often involved dynamic responses.^{11–14} Even studies of relatively small proteins, such as metallo-beta-lactamases, show that dynamic changes can occur upon binding even in the absence of structural changes.¹⁵ There are also studies showing that allostery can be modulated by intrinsic disordered protein.^{16,17} But how the dynamics transfer signals along the protein and affect protein function is still elusive.

All-atom molecular dynamics simulations can be used to study allosteric couplings, but are limited to relatively small systems. Allosteric proteins are usually large, therefore, a coarse-grained model becomes a good alternative, in which groups of atoms are treated as one single bead resulting in increasing the speed dramatically. Although coarse-grained models have been proved a valuable model, tests on dynamical measures and responses to mutation have been performed less-often. In order to see how well different coarse-grained models can reproduce correlated motions, we picked three popular models, the Go,¹⁸ Martini¹⁹ and Cafemol²⁰ models to explore the correlated motions using the small protein Eglin C as a model system. Clarkson et al. demonstrated that Eglin C while not a classically allosteric protein does exhibit allosteric behavior.²¹ This notion of Generalized allostery according to the results by Gunasekaran and co-workers,²² in which nonallosteric proteins can behave allosterically, due to perturbations causing population shifts is precisely what has been examined in Eglin C and makes it a good model system for testing coarse-grained models. While all three models performed well, or at least some measures, we find that the simplest model, Go, and the Cafemol model performed better. Overall, we find that Cafemol is the most accurate as well as an easy-to-use model to explore the protein dynamics easily and accurately.

Model and Methods

Eglin C, a small (70 residue) monomeric protein from the potato inhibitor I family of serine proteinase inhibitors, is a homologue of chymotrypsin inhibitor 2 which has been used in numerous studies on folding and stability.^{23–25} Eglin C is an alpha and beta protein with α -helix spanning residues 18–28, while the β -sheet spanning residues 33–38, 51–56, 68–69. The structure of Eglin C is known from NMR (PDB:1EGL)²⁶. The small size of this protein makes it particularly amenable to studies via all-atom simulations for comparison of both wildtype and mutation structures.

Go models are native-centric. They are directly constructed from the protein's native structure. The potential contains only pair-wise native interactions.¹⁸ The input files are produced by using Go model builder server.^{27,28} The simulations are performed by using

NAMD package²⁹ with Langevin dynamics. Temperature is set to 300 K. An integration timestep of 5 fs is used as recommended in CHARMM manual.³⁰

MARTINI potential for proteins is largely based on physico-chemical modeling with a weak bias to the native structure mostly via the secondary structure constraints.¹⁹ The backbone of each amino acid residue is collectively described by one bead, and the side chains are represented by a variable number of beads depending on the dimensions of the side chain of each amino acid. The force field is integrated into NAMD package according to the manual of Residue-based coarse graining.³¹ Langevin dynamics is used to investigate the dynamical behaviors. Before the production run, 20000 steps are performed for minimization. Temperature is set to 300 K. There are 5 simulations with each simulation 40 ns. An integration timestep of 10 fs is used. Switching function is adapted for Lennard-Jones potential with switchdist 9 and cutoff 12 Å.

Cafemol is a recently developed software to perform coarse grained molecular dynamics simulation. It can simulate proteins, nucleic acids, lipid membrane, and their complexes with various CG models.²⁰ Since the release of the program, it has been used to investigate various systems and has performed well.^{32–35} In this work, Cafemol version 2.1 is used to perform Langevin dynamics simulations. The temperature is set to 300 Kelvin. There are Go interactions within the protein. Non-local protein interactions are the repulsive interaction which adopts the form of $\epsilon(d/r_{ij})^{12}$. Ca atoms are used as the representative for the CG particle of protein. Friction coefficient is set to 0.25 as suggested by Cafemol, which works well for folding of single-domain protein. CafeMol implements the Debye-Huckel type electrostatic interactions. Time length in each MD step is 0.2 Cafe-time as recommended by Cafemol manual (1 Cafe-time $\approx$ 200fs). Totally 80000000 steps are performed for each simulation. Structure information is saved every 200 steps. Neighbor list is updated every 100 steps. 5 simulations are performed for each system with different random seeds.

As a reference, all-atom Langevin dynamics simulation (NPT) is also performed. Simulations were carried out with periodic boundary condition within a cubic cell with the minimal distance between the protein and the edge of water box being 8 Å. Ions are added to make the system neutral. All calculations are performed within the NAMD package,²⁹ utilizing the CHARMM22/CMAP parameter set.³⁶ To relax steric overlaps and optimize the hydrogen atom positions, each structure is minimized for 1000 cycles. The temperature is set to be 300 K. This temperature will be roughly maintained across the affected atoms through the addition of friction and random forces. The SHAKE algorithm³⁷ is used for restricting motions of all covalent bonds with hydrogen atoms, and the integration timestep is 2 fs. Switching function is adapted for Lennard-Jones potential with switching distance of 10 Å and cutoff 12 Å.

The initial structures for the wild type are downloaded from PDB website (PDBID: 1CSE and 1EGL). Mutated structures are constructed by using the Mutator plugin of VMD.³⁸ 5 simulations are performed for the four sequences with each simulation 40 ns by using different models. The first 5 ns data are excluded for analysis.

To detect the flexibility of different regions of the protein, RMSF (root mean square fluctuation of atomic positions) is used, which is defined as: $RMSF = \sqrt{\langle (x_i - \tilde{x}_i)^2 \rangle}$, where $\tilde{x}_i$ is the average value of x_i , brackets denote time average. Correlated motion is described in terms of correlated fluctuations. The covariance matrix, also known as dynamic cross-

correlation matrix (DCCM), is in the term of $DCCM_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}} \sqrt{C_{jj}}}$, where $C_{ij} = \langle (x_i - \tilde{x}_i)(x_j - \tilde{x}_j) \rangle$. These coefficients measure the linear correlation between two residues. A unitary value indicates a positive correlation (movements in the same direction), while -1 indicates a negative correlation (correlated motions in opposite directions).

Exploring the dynamics of Eglin C protein by using 5 different models

RMSFs may reflect local structural heterogeneity, especially experimentally, when derived from a molecular dynamics simulation it is precisely a measure of atomic fluctuations. Though atomic fluctuations by themselves are not indicative of allostery, they are an important measure of dynamics that need be evaluated for different models. Figure 1 shows the RMSF profile for the 5 models: all-atom, Cafemol, Go, Martini and NMR. NMR results are calculated from all the models of pdb 1EGL. From this figure, we can see that Cafemol and NMR have the most similar profile with all-atom model, with a Spearman correlation coefficient of 0.84 and 0.90 respectively after excluding the 3 very mobile residues at the termini. Although the Go model has a larger average RMSF than that of the all-atom model, it has a Spearman correlation coefficient of 0.89 with the all-atom model. Martini has a Spearman correlation coefficient of 0.30 with all-atom model. This suggests that while Martini is good at exploring protein assembly process,^{39,40} it is not as good as other models for studying fluctuations. Cafemol has a correlation of 0.90 with NMR which is higher than that of Cafemol and all-atom model. Go model has a correlation of 0.85 with NMR and Martini has a correlation of 0.15 with NMR. Go model and Martini have a smaller correlation coefficient with NMR than that with all-atom model. The standard deviation of RMSF for the five simulations of all-atom simulations is less than 0.3 Å except for the first 8 amino acids and the last amino acid which are at the ends. For Cafemol, the standard deviation of RMSF is less than 0.06 Å. For Go model, that is less than 0.3 Å. For Martini model, the standard deviation of RMSF is around 1 Å. This result means that different simulations alter much while using Martini model, but not much while using the other three models, especially Cafemol has the smallest standard deviation.

Figure 2 shows the covariance matrix for the three models (Go, Martini, Cafemol) on the bottom below the diagonal, with the all-atom model on the top above the diagonal. While it is true that a correlation matrix is indexed by residues, and in a folded structure residues which are distant in sequence may be close in space, the figures show all the correlations including both short-range and long-range correlations. Correctly modeling the latter is critical in understanding allostery. As a comparison, a covariance matrix is also constructed from the NMR models. Bars in the left and bottom are the secondary structure indicators, red bars indicate α -helix and brown ones indicate β -strand. From the covariance matrix of all-atom model, we can see that there are correlations between the following pairs: N-

terminal coil- β_3 , β_1 - β_2 , β_1 - β_3 , β_2 - β_3 . All of the four models recapture these correlations, except that Martini shows a negative correlation in N-terminal coil- β_3 instead of positive. It is worth noting that there are several regions showing increased negative correlation in Cafemol and NMR than in all-atom model. Most of them involve the coil part, and Cafemol has a correlation of 0.9 with NMR. The RMSF of varied conformations from NMR has a similar profile with the RMSF from dynamical models, the same is true for covariance matrix, suggesting that there is an intrinsic relation between native-state dynamics and conformation heterogeneity in crystal structures. This result is consistent with the result given by Friedland et. al.⁴¹ To check if there is any structural deviation, RMSD between structures at 40 ns of different models are calculated. RMSD of the structures at 40 ns between Go and all-atom is 0.51 nm, that between Cafemol and all-atom is 0.37 nm, while 0.91 nm for Martini and all-atom model. This suggests that along with dynamical change, Martini model also explores structures that are deviating from the native structure.

Test the sensitivity of Cafemol for detecting the difference caused by altered protein length, as well as the effect of mutation

To test the sensitivity of Cafemol for detecting the difference caused by altered protein length, another form of Eglin C is introduced. It exists in complex with various proteinases from X-ray crystallography (PDB:1CSE),⁴² containing 63 residues with F10W mutation introduced to facilitate fluorescence monitoring of chemical denaturation-based stability measurements⁴³. Research shows that the structural and dynamic properties of wild-type and pseudo-wild-type Eglin C (F10W) are virtually identical.⁴⁴ F25A mutation is also introduced to explore the mutation effect. Sequences are shown in Figure 3. Red and blue letters show the differences in sequences. The full length is named as 1EGL containing 70 amino acids, and the truncated one is called ec containing 63 amino acids with a F10W mutation. The other two sequences are the F25A mutations.

Figure 4(a) shows the RMSF as a function of residues for the four sequences, namely ec wild, ec mut, 1EGL wild and 1EGL mut from the all-atom simulations. Black stars indicate the position of F25A mutation, which is located in the helix. We can see that F25A mutation increases the RMSF of residues 25–33, which is partly helix and partly coil shown in Figure 4(d) in blue. But F25A has no effect on 1EGL. Figure 4(b) is the RMSF as a function of residues from Cafemol simulations, which shows that ec wild and ec mut have a similar profile and 1EGL wild and 1EGL mut have a similar profile. This means that Cafemol did not detect the mutational effect for ec sequence, but it did detect the difference between 1EGL and ec. To evaluate the quality of the detection of the differences, correlation of $RMSF_{1EGL} - RMSF_{ec}$ between all-atom model and Cafemol model is performed (shown in figure Figure 4(c)), which gives a correlation coefficient of 0.35. The RMSD between the initial structure of ec and 1EGL is 0.912 Å by using the common 63 amino acids, suggesting that the different results between ec and 1EGL is not due to the difference of the initial structures.

To check the impact of mutation on the allosteric conformational equilibrium population, clustering analysis is performed for the all-atom simulations. The RMSD between the central structure of the first cluster of ec wild type and ec mutant is 1.361 Å, which is 1.784

Å for 1EGL by using a threshold of 1.5 Å for clustering. (that is 1.361 Å and 2.137 Å using a threshold of 2.0 Å for clustering.) This suggests that the mutation does not affect the conformation but only the dynamics of ec.

Figure 5 shows the covariance matrix for the four sequences by using all-atom and Cafemol model, with all-atom model located at the top left, while Cafemol at the right bottom. As mentioned before, 1EGL wild type has correlations: N-terminal coil- β_3 , β_1 - β_2 , β_1 - β_3 , and β_2 - β_3 . The truncated sequence has a decreased correlation for all the pairs, and Cafemol only keeps β_1 - β_3 correlation, but explored some negative correlated regions. The effects of F25A appear to be modest on the covariance matrix.

All-atom model is physically accurate. It can reproduce protein behavior as observed in experiments, so can be used as a golden standard when discussing coarse-grained simulations. Go and Cafemol models are topology-based model. The assembly is around the native structure. Martini model probably explores intermediate states. The different levels of matching to the AA simulations for the coarse-graining models probably have some relationship with the explored structures. The three coarse-graining models are all implicit solvent, yet Go and Cafemol reproduce the protein behavior as observed in all-atom model, showing that implicit solvent is a good approximation in this case.

Conclusions

To study the dynamical mechanisms of allostery efficiently, we use three well-defined coarse grained models: Go, Martini and Cafemol, and a small model protein Eglin C to examine if these coarse grained models can explore the dynamics of Eglin C efficiently as well as to see how these models respond to mutations. By comparing the fluctuations and correlation matrix from different models, We find that Cafemol and Go have the highest correlation with all-atom model, with Cafemol showing the best agreement with all-atom model. Furthermore, Cafemol can be easily extended to study protein-DNA complex, hence it is a good candidate to study the dynamical property of proteins. Nevertheless, while Cafemol works well to capture the effects of a truncation mutant, it is less well suited for studying single-point mutant.

Acknowledgments

All calculations were performed on the DEAC cluster supported by the WFU Information systems and the WFU Provost's office. This research was partially supported by NIH R01CA129373 to FRS and Crystallography & Computational Biosciences Core services supported by the Comprehensive Cancer Center of Wake Forest University NCI CCSG P30CA012197 grant.

References

1. Hilser VJ, Wrabl JO, Motlagh HN. Structural and Energetic Basis of Allostery. *Annu Rev Biophys.* 2012; 41:585–609. PMID: 22577828. [PubMed: 22577828]
2. Monod J, Wyman J, Changeux J-P. On the nature of allosteric transitions: A plausible model. *J Mol Biol.* 1965; 12:88–118. [PubMed: 14343300]
3. Koshland DE, Némethy G, Filmer D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry.* 1966; 5:365–385. [PubMed: 5938952]

4. Cooper A, DT D. Allostery without conformational change. A plausible model. *Eur Biophys J*. 1984; 11:103–109. [PubMed: 6544679]
5. Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. *Nature structural & molecular biology*. 2006; 13:831–838.
6. Yonetani T, Laberge M. Protein dynamics explain the allosteric behaviors of hemoglobin. *Biochim Biophys Acta*. 2008; 1784:1146–1158. [PubMed: 18519045]
7. Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL. Hidden dynamic allostery in a PDZ domain. *Proc Nat Acad Sci U S A*. 2009; 106:18249–18254.
8. Tzeng S-R, Kalodimos CG. Dynamic activation of an allosteric regulatory protein. *Nature*. 2009; 462:368–372. [PubMed: 19924217]
9. Wang G, Li C, Wang Y, Chen G. Cooperative assembly of Co-Smad4 MH1 with RSmad1/3 MH1 on DNA: a molecular dynamics simulation study. *PLoS One*. 2013; 8:e53841. [PubMed: 23326519]
10. Pieniazek SN, Hingorani MM, Beveridge DL. Dynamical allostery in the mechanism of action of DNA mismatch repair protein MutS. *Biophys J*. 2011; 101:1730–1739. [PubMed: 21961599]
11. Negureanu L, Salsbury FR. Insights into protein - DNA interactions, stability and allosteric communications: a computational study of mutSα-DNA recognition complexes. *J Biomol Struct Dyn*. 2012; 29:757–776. [PubMed: 22208277]
12. Negureanu L, Salsbury FR. The molecular origin of the MMR-dependent apoptosis pathway from dynamics analysis of MutSα-DNA complexes. *J Biomol Struct Dyn*. 2012; 30:347–361. [PubMed: 22712459]
13. Salsbury FR, Clodfelter JE, Gentry MB, Hollis T, Scarpinato KD. The molecular mechanism of DNA damage recognition by MutS homologs and its consequences for cell death response. *Nucleic Acids Res*. 2006; 34:2173–2185. [PubMed: 16648361]
14. Salsbury FR. Effects of Cisplatin binding to DNA on the dynamics of the E. Coli MutS dimer. *Protein Pept Lett*. 2010; 17:744–750. [PubMed: 19995335]
15. Salsbury FR, Crowder MW, Kingsmore SF, Huntley JJA. Molecular dynamic simulations of the metallo-beta-lactamase from *Bacteroides fragilis* in the presence and absence of a tight-binding inhibitor. *J Mol Model*. 2009; 15:133–145. [PubMed: 19039608]
16. Ferreón ACM, Ferreón JC, Wright PE, Deniz AA. Modulation of allostery by protein intrinsic disorder. *Nature*. 2013; 498:390–394. [PubMed: 23783631]
17. Hilser VJ, Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc Nat Acad Sci U S A*. 2007; 104:8311–8315.
18. Estácio SG, Fernandes CS, Krobath H, Faísca PFN, Shakhnovich EI. Robustness of atomistic G models in predicting native-like folding intermediates. *The Journal of Chemical Physics*. 2012; 137:085102. [PubMed: 22938266]
19. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-j. The MARTINI Coarse-Grained Force Field : Extension to Proteins. *J. Chem. Theory and Comput*. 2008; 4:819–834. [PubMed: 26621095]
20. Kenzaki H, Koga N, Hori N, Kanada R, Li W, Okazaki K-i, Yao X-Q, Takada S. CafeMol: A Coarse-Grained Biomolecular Simulator for Simulating Proteins at Work. *J Chem Theory Comput*. 2011; 7:1979–1989. [PubMed: 26596457]
21. Clarkson MW, Gilmore SA, Edgell MH, Lee AL. Dynamic Coupling and Allosteric Behavior in a Nonallosteric Protein. *Biochemistry*. 2006; 45:7693–7699. [PubMed: 16784220]
22. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? *Proteins Struct Funct Bioinf*. 2004; 57:433–443.
23. Clarkson MW, Lee AL. Long-range dynamic effects of point mutations propagate through side chains in the serine protease inhibitor eglin c. *Biochemistry*. 2004; 43:12448–12458. [PubMed: 15449934]
24. Materese CK, Goldmon CC, a Papoian G. Hierarchical organization of eglin c native state dynamics is shaped by competing direct and water-mediated interactions. *Proc Nat Acad Sci U S A*. 2008; 105:10659–10664.
25. Boyer JA, Lee AL. Monitoring Aromatic ps-ns Dynamics in Proteins via 13 C Relaxation: Expanding Perturbation Mapping of the Rigidifying Core Mutation, V54A, in eglin c. *Biochemistry*. 2008; 47:4876–4886. [PubMed: 18393447]

26. Hyberts SG, Goldberg MS, Havel TF, Wagner G. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* 1992; 1:736–751. [PubMed: 1304915]
27. Karanicolas J, Brooks CL. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 2002; 11:2351–2361. [PubMed: 12237457]
28. Karanicolas J, Brooks CL. Improved G-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J Mol Biol.* 2003; 334:309–325. [PubMed: 14607121]
29. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005; 26:1781–1802. [PubMed: 16222654]
30. http://www.charmmtutorial.org/index.php/Coarse_Grained_Go_Models
31. <http://www.ks.uiuc.edu/Training/Tutorials/>
32. Terakawa T, Takada S. Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain. *Biophys J.* 2011; 101:1450–1458. [PubMed: 21943426]
33. Terakawa T, Kenzaki H, Takada S. p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains. *J Am Chem Soc.* 2012; 134:14555–14562. [PubMed: 22880817]
34. Li W, Terakawa T, Wang W, Takada S. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc Natl Acad Sci U S A.* 2012; 109:17789–17794. [PubMed: 22753508]
35. Okazaki, K-i; Sato, T.; Takano, M. Temperature-enhanced association of proteins due to electrostatic interaction: a coarse-grained simulation of actin-myosin binding. *J Am Chem Soc.* 2012; 134:8918–8925. [PubMed: 22559201]
36. Brooks BR, et al. CHARMM: the biomolecular simulation program. *J Comput Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
37. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys.* 1977; 23:327–341.
38. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996; 14:33–38. 27–28. [PubMed: 8744570]
39. Bond PJ, Sansom MSP. Insertion and assembly of membrane proteins via simulation. *J Am Chem Soc.* 2006; 128:2697–2704. [PubMed: 16492056]
40. Bond PJ, Holyoake J, Ivetac A, Khalid S, Sansom MSP. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *J Struct Biol.* 2007; 157:593–605. [PubMed: 17116404]
41. Friedland GD, Lakomek N-A, Griesinger C, Meiler J, Kortemme T. A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. *PLoS Comput Biol.* 2009; 5:e1000393. [PubMed: 19478996]
42. Bode W, Papamokos E, Musil D. The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. Structural analysis, subtilisin structure and interface geometry. *Eur J Biochem.* 1987; 166:673–692. [PubMed: 3301348]
43. Yi F, Sims DA, Pielak GJ, Edgell MH. Testing hypotheses about determinants of protein structure with high-precision, high-throughput stability measurements and statistical modeling. *Biochemistry.* 2003; 42:7594–7603. [PubMed: 12809516]
44. Fetrow JS, Knutson ST, Edgell MH. Mutations in alpha-helical solvent-exposed sites of eglin c have long-range effects: evidence from molecular dynamics simulations. *Proteins.* 2006; 63:356–372. [PubMed: 16342264]

Figure 1.
RMSF profile for the 5 models. Bars at the bottom of the figure indicate secondary structures. Red indicates α -helix, and brown indicates β -sheet.

Figure 2.
Covariance matrix of all the models. For comparison, top left triangle shows the covariance matrix of all-atom model.

Figure 4. RMSF from all-atom (a) and Cafemol (b) model for the four sequences. (c) Correlation of $RMSF_{1EGL} - RMSF_{ec}$ between all-atom model and Cafemol model. (d) Structure of eglin C. Blue part is the region where F25A mutation take effect, which has a larger RMSF than in the wild type of *ec*, but not in *1EGL*.

Figure 5.

Covariance matrix for the four sequences. For comparison, the triangle at top left shows the covariance matrix from all-atom model, while the one at bottom right shows the covariance matrix from Cafemol model.