Comparative Investigation of Normal Modes and Molecular Dynamics of Hepatitis C NS5B Protein

M S Asafi[1], A Yildirim[2] and M Tekpinar[1]

[1]Yuzuncu Yil University, Van, Turkey
[2]Siirt University, Siirt, Turkey

E-mail: tekpinar@buffalo.edu

Abstract. Understanding dynamics of proteins has many practical implications in terms of finding a cure for many protein related diseases. Normal mode analysis and molecular dynamics methods are widely used physics-based computational methods for investigating dynamics of proteins. In this work, we studied dynamics of Hepatitis C NS5B protein with molecular dynamics and normal mode analysis. Principal components obtained from a 100 nanoseconds molecular dynamics simulation show good overlaps with normal modes calculated with a coarse-grained elastic network model. Coarse-grained normal mode analysis takes at least an order of magnitude shorter time. Encouraged by this good overlaps and short computation times, we analyzed further low frequency normal modes of Hepatitis C NS5B. Motion directions and average spatial fluctuations have been analyzed in detail. Finally, biological implications of these motions in drug design efforts against Hepatitis C infections have been elaborated.

1. Introduction

Hepatitis C virus (HCV) is an infection affecting millions of people worldwide. Unfortunately, there is not an effective treatment for this fatal viral disease. One of the drug targets in HCV infections is NS5B polymerase protein, which replicates the viral genome. NS5B polymerase protein performs conformational changes during genome replication. Therefore, understanding NS5B protein dynamics is vital for developing effective inhibitors against this protein.

Molecular dynamics and elastic network methods are widely used methods to investigate dynamics of proteins. There are a few molecular dynamics studies on HCV NS5B protein [1, 2]. However, there is not any study of HCV NS5B within framework of elastic network models. How do molecular dynamics simulation results are related with elastic network model data is one of answers we look for in this study. As a result, we employed both molecular dynamics and elastic network model to understand dynamics of HCV NS5B polymerase protein. We compared results of B-factors, normalized cross-correlations. Moreover, we compared principal components obtained from a molecular dynamics trajectory to low frequency normal modes from a coarse-grained elastic network model. Finally, we investigated low frequency normal modes obtained with elastic network model in detail and we elaborated on their biological implications.

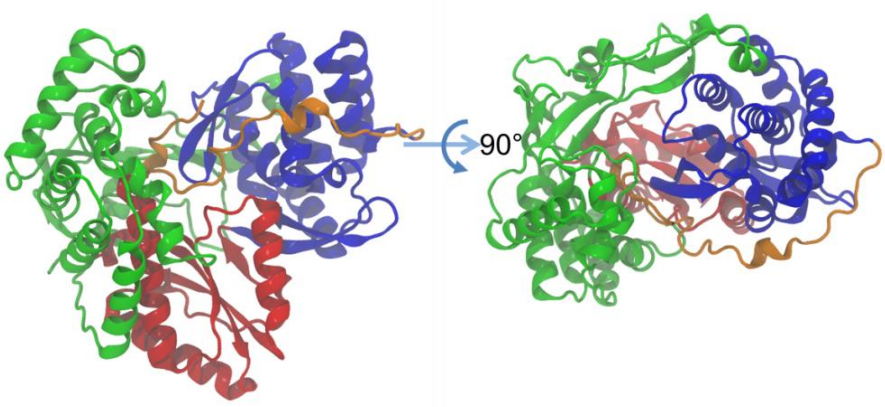

Figure 1. Side (left panel) and top (right panel) views of NS5B protein. Fingers domain is green (left group), palm domain is red (middle group) and thumb domain is blue (right group) new cartoon representation. Please check online version for colour figure.

2. Materials and methods

We utilized Visual Molecular Dynamics (VMD) program, Bio3D R package, and Prody package in addition to many custom made python scripts for visualization and analysis throughout the text [3-5].

2.1. Initial structure and its domains

We used 4aep crystal structure of HCV NS5B deposited in Protein Data Bank for all calculations [6]. It has been suggested that HCV NS5B resembles to a right hand [7]. Therefore, this structure contains three domains (Figure 1). Residues 1 to 191 and 230 to 184 comprise fingers domain. Palm domain is made up of residues from 192 to 229 and residues 285 to 367. Finally, thumb domain is from residue 368 to 563. Thumb and fingers domains open or close during RNA replication process.

2.2. Molecular dynamics (MD) simulation and principle component analysis

Only protein atoms and crystal waters of the crystal structure have been retained for simulations. 19607 solvent molecules of TIP3P model have been added to solvate the protein in a dodecahedral box[8]. 63 Na and 79 Cl ions have been added to the system to neutralize it. Particle Mesh Ewald method has been used for electrostatic energy calculations [9]. 50 picoseconds (ps) of constrained NVT simulations preceded 100 ps of unconstrained NPT simulations. All of the simulations have been performed under 300 K and 1 bar pressure. Gromacs simulation package with Charmm27 force field has been employed for the simulations [10, 11]. Production run was 100 nanoseconds with 2 femtoseconds time step. The trajectory frames were dumped out every 40 ps.

Normalized cross correlations (C_{ij}^{n}) can be calculated from an MD trajectory by using

$$C_{ij}^{n} = \frac{\left\langle \left(R_i - \left\langle R_i \right\rangle\right)\left(R_j - \left\langle R_j \right\rangle\right) \right\rangle}{\sqrt{\left\langle R_i - \left\langle R_i \right\rangle \right\rangle^2 \left\langle R_j - \left\langle R_j \right\rangle \right\rangle^2}} \quad (1)$$

where $\langle\rangle$ denotes averaging over trajectory frames and R_i(or R_j) denotes position of Cα atom i (or j). This matrix can be diagonalized and eigenvectors of it are principal components of a trajectory. Low frequency principal components (PC) can indicate collective motions observed trajectories. Moreover, when MD trajectory frames are projected onto the first few PCs, major conformational changes can be visualized through clusters observed in 2D plots.

2.3. Elastic network model

In an elastic network model, generally all Cα atoms of a protein within a cutoff radius (15 Å), which are connected with springs of uniform stiffness, interact through a simple Hookean potential given below [12, 13]:

$$V = \frac{\gamma}{2} \sum_{j(j \neq i)} \Gamma_{ij} \left(R_{ij} - R_{ij}^{0} \right)^2 \tag{2}$$

γ denotes uniform spring constant, R^{0}_{ij} is equilibrium distance between Cα atoms of residue i and j, while R_{ij} is instantaneous distance between Cα atoms of residue i and j. Γ_{ij} is ijth element of Kirchoff matrix Γ and it is zero for interactions beyond the cutoff radius. A hessian matrix can be constructed from this potential. The hessian matrix has to be diagonalized and low frequency eigenvalues of the diagonalized matrix correspond to frequencies of collective motions. It has been shown that normal modes obtained with this type of potential can be used to predict observed conformational changes of proteins as well as fluctuations of proteins around native conformations [14, 15].

Cross-correlations (C_{ij}) can be calculated from a selected subset of L non-zero normal modes using

$$C_{ij} = \left\langle \Delta R_i \cdot \Delta R_j \right\rangle = \frac{3k_B T}{\gamma} \sum_{k}^{L} \left[\frac{u_k u_k^T}{\lambda_k} \right]_{ij} \tag{3}$$

where k_B is Boltzmann's constant, T is temperature, and γ is uniform spring force constant. λ_k denotes eigenvalue and u_k is the corresponding eigenvector.

One can calculate normalized cross-correlations (C_{ij}^{n}) from the equation above as follows:

$$C_{ij}^{n} = \frac{\left\langle \Delta R_i \cdot \Delta R_j \right\rangle}{\sqrt{\left\langle \Delta R_i \cdot \Delta R_i \right\rangle \left\langle \Delta R_j \cdot \Delta R_j \right\rangle}} \tag{4}$$

3. Results and discussion

3.1. B-factor calculations with molecular dynamics and normal mode analysis

We calculated theoretical B-factors using molecular dynamics trajectory and elastic network model. Correlation cosine (cc) agreement between experimental B-factors and the ones calculated with molecular dynamics is 0.812. B-factors calculated with elastic network model show a better overlap (0.954) with experimental data.

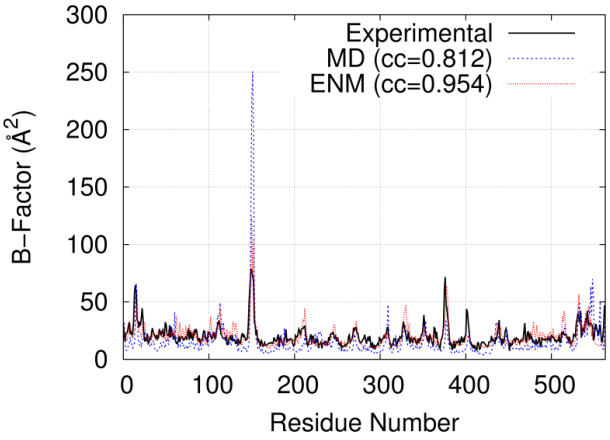

Figure 2. B-factors calculated with elastic network model (dotted) and molecular dynamics simulations (dashed) have been superimposed onto the experimental data (solid).

We can see that both methods overestimate fluctuation for residues around residue 150. However, degree of overestimation is better for elastic network model compared to MD. This excessively floppy region is located in fingers domain. Despite this overestimation, both methods can predict B-factors with a reasonable accuracy overall.

3.2. Principal components of the molecular dynamics trajectory and their agreement with normal modes.

We calculated the principal components (PCs) of our molecular dynamics trajectory. The first 5 principal components comprises of about ~35% of the motion observed in the trajectory. PC1 contributes 18% to the sum, PC2 contributes ~8%, and PC3 contributes ~6%. Contribution of the other principal components to the total motion decreases with principal component index. Contribution of the first 20 principal components reaches to 61%.

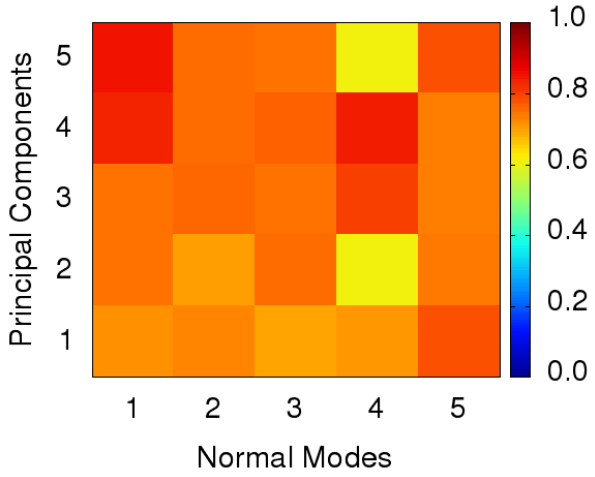

Figure 3. Overlaps of the first five principal components and normal modes. 1.0 indicates a perfect directional agreement while 0.0 denotes that the vectors are perpendicular or unrelated. Please check online version for colour figure.

Moreover, we wanted to observe if there is an agreement between principal components and normal modes calculated from elastic network model. So, we calculated the cosine angle between 3N dimensional principal components and normal modes. If angle is 0°, it means that two vectors are in same direction. If angle is 90°, it means that the vectors are perpendicular to each other and so on. Our data shows that the cosine correlation is greater than 0.6 and this indicates a good directional agreement between results of the two different methods. Finally, we should note that the order of low frequency motions between principal components and normal modes is not preserved.

3.3. Normalized cross-correlations obtained with molecular dynamics and elastic network model.

Normalized cross-correlations can give information about positive and negative cooperation within a protein. Normalized cross correlations have been calculated as explained above in Materials and Methods section (Figure 4). For normal modes, we considered only the first 100 low frequency modes. The first 10 residues have a high correlation with residue 280, 400 and 500. This pattern can be observed both in MD and normal mode cross correlations. Overall, we observe that positive correlation predictions are in better agreement between two methods. C terminal residues show a negative correlation with fingers domain residues (around residue 100). Despite methodical differences and some minor disagreements in the plots, general similarity of normalized cross-correlations is striking.

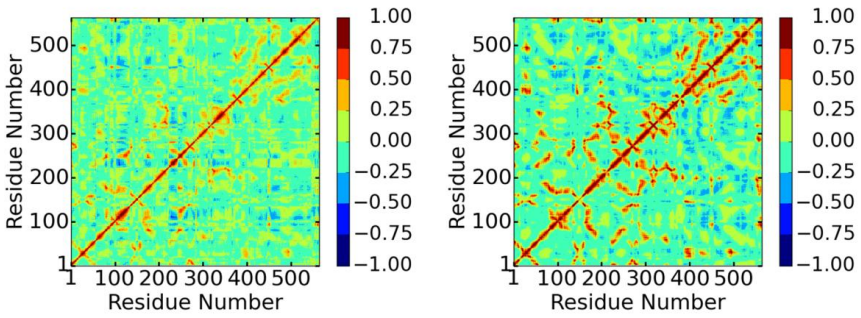

Figure 4. Normalized cross correlations obtained from molecular dynamics trajectory (left) and elastic network model (right). Please check online version for colour figure.

3.4. Low frequency normal modes and their biological meanings

We visualized low frequency normal modes calculated with coarse-grained elastic network model (Figure 5). We observed that most of the low frequency modes are involved in fingers and thumb domain motions. The first mode is mainly about rotation of thumb domain and fingers in reverse directions. The width of RNA cavity is approximately same in this mode. However, the cavity opens up and gets closed in mode 2 and 3. It is highly possible that these two modes are functional modes corresponding to conformational states in RNA replication. Mode 4 is mostly an out of plane motion of a loop located in fingers domain. This motion is coupled with thumb domain. However, motion of the same loop is coupled with motions of all domains including the palm domain. Considering that major experimental conformational states are involved in changes of fingers-thumb domain distance variations, we conclude that low frequency normal modes can sample conformational states of NS5B observed in experimental studies [16].

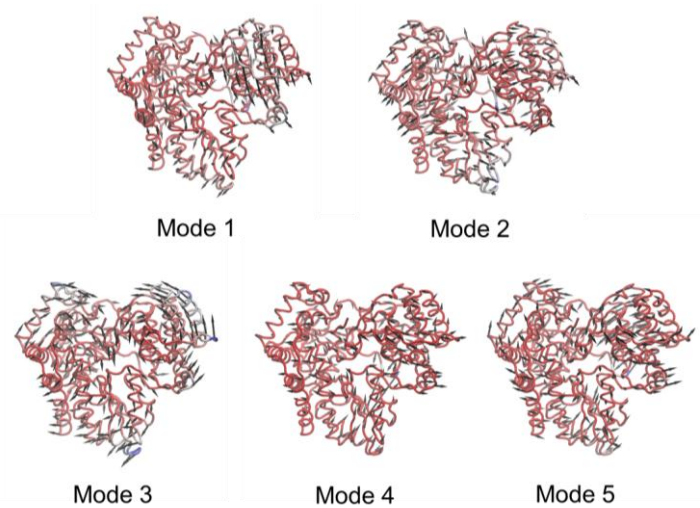

Figure 5. Side view of the first 5 low frequency normal modes. Colour indicates fluctuations. Blue means high fluctuations and red means low. Black arrows denote direction of motion and length of the arrow is proportional to amplitude of the motion. Please check online version for colour figure.

4. Conclusions

Molecular dynamics is an invaluable tool for understanding protein dynamics in atomistic detail. However, long computational times and big parallel computational resources are needed to perform MD simulations. On the other hand, elastic network models can provide a sufficiently accurate dynamics

information at a coarse-grained level. Our study also proves this point by showing the agreement between experimental, MD and coarse-grained results of B-factors for HCV NS5B. Moreover, when both methods are employed we obtain a similar picture for normalized cross-correlations of NS5B.

Low frequency motions of NS5B indicate that thumb and finger domains open up and get closed. Experimental studies show that these conformational states are required for RNA replication by HCV NS5B. Therefore, elastic network model can predict functional conformational states of NS5B protein. Moreover, inter-domain correlations may indicate possible allosteric interactions. These interactions can guide drug design efforts, which aim to stop transitions between conformational states of NS5B. Based on these results, we conclude that elastic network modes can provide invaluable information about dynamics of HCV NS5B protein without using lengthy MD simulations.

References

[1] Davis BC, Thorpe IF 2013 Molecular simulations illuminate the role of regulatory components of the RNA polymerase from the hepatitis C virus in influencing protein structure and dynamics. Biochemistry.52(26):4541-52.
[2] Brown JA, Thorpe IF 2015 Dual Allosteric Inhibitors Jointly Modulate Protein Structure and Dynamics in the Hepatitis C Virus Polymerase. Biochemistry.54(26):4131-41.
[3] Humphrey W, Dalke A, Schulten K 1996 VMD: visual molecular dynamics. J Mol Graph.14(1):33-8, 27-8.
[4] Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS 2006 Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics.22(21):2695-6.
[5] Bakan A, Meireles LM, Bahar I 2011 ProDy: Protein Dynamics Inferred from Theory and Experiments. Bioinformatics.27(11):1575-7.
[6] Scrima N, Caillet-Saguy C, Ventura M, Harrus D, Astier-Gin T, Bressanelli S 2012 Two crucial early steps in RNA synthesis by the hepatitis C virus polymerase involve a dual role of residue 405. J Virol.86(13):7107-17.
[7] Lesburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF, Weber PC 1999 Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. Nat Struct Biol.6(10):937-43.
[8] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML 1983 Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics.79(2):926-35.
[9] Darden T, York D, Pedersen L 1993 Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. Journal of Chemical Physics. 98(12):10089-92.
[10] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ 2005 GROMACS: fast, flexible, and free. J Comput Chem. 26(16):1701-18.
[11] MacKerell AD, Brooks CL, Nilsson L, Roux B, Won Y, Karplus M 1998 CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In: Schleyer, editor. 1: John Wiley & Sons: Chichester. p. 271-7.
[12] Bahar I, Atilgan AR, Erman B 1997 Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des.2(3):173-81.
[13] Tirion MM 1996 Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. Phys Rev Lett.77(9):1905-8.
[14] Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I 2001 Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J. 80(1):505-15.
[15] Tama F, Sanejouand YH 2001 Conformational change of proteins arising from normal mode calculations. Protein Eng.14(1):1-6.
[16] Appleby TC, Perry JK, Murakami E, Barauskas O, Feng J, Cho A, et al 2015 Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. Science.347(6223):771-5.