# Bioinformatics Analysis of SARS Proteins and Molecular Dynamics Simulated Structure of an Alpha-helix Motif

**Amarender D Reddy, Seung Bum Suh, Reza Ghaffari, N. Jiten Singh, Dae-Jin Kim,† Joon Hee Han,† and Kwang S. Kim***

*National Creative Research Initiative Center for Superfunctional Materials, Department of Chemistry, Division of Molecular and Life Sciences, Pohang University of Science and Technology, San 31, Hyojadong, Namgu, Pohang 790-784, Korea*
*†Bioinformatics S/W Research Center, Department of Computer Sciences and Engineering, Pohang University of Science and Technology, San 31, Hyojadong, Namgu, Pohang 790-784, Korea*
*Received June 28, 2003*

**Key Words :** SARS, Bioinformatics, Homology, Alpha-helix, Simulation

The complete sequencing of the severe acute respiratory syndrome (SARS) coronavirus (CoV) within a short span of time is a remarkable achievement, which has given the scope for extensive research on the virus.[1] Sequence analysis reveals the phylogeny of SARS-CoV that the genome has most of the characteristic features of a coronavirus, but it belongs to a new group which is sufficiently different from all known coronaviruses.

All the coronaviruses consist of four important structural proteins: spike glycoprotein (S), small envelope protein (E), matrix glycoprotein (M) and nucleocapsid protein (N). The E protein comprised of at least 76 amino acids (aa) is a pivotal player in the morphogenesis of the virion envelope by interacting with the M protein.

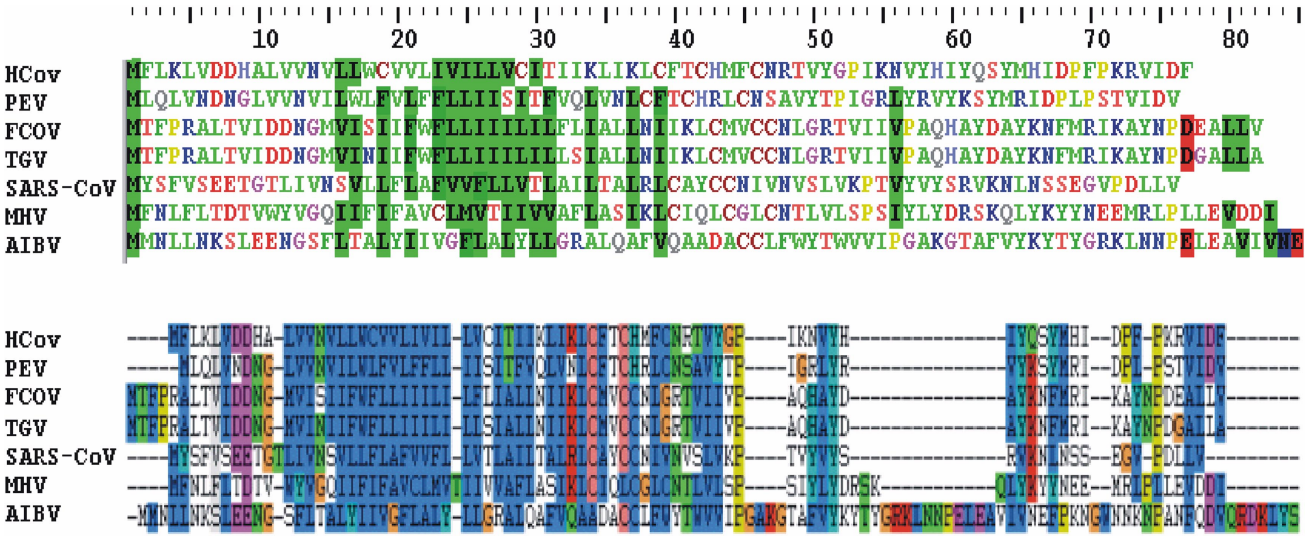


**Figure 1**. Sequence alignment of all the E proteins among coronaviruses using Bioedit (top) and ClustalX (below); SARS-CoV [NP_828854], HCoV [Human CoV 229E, NP_073554], FCoV [Feline CoV CAA74228.1], TGV [Transmissible gastroenteritis virus, NP_058426], PEDV [Porcine epidemic diarrhea virus, NP_598312], MHV [Murine hepatitis virus, NP_068673], AIBV [avian infectious bronchitis virus, AAO33465].


**Figure 2**. Secondary structure prediction of the SARS-CoV E protein by PredictProtein (PP).

*Corresponding author. E-mail: kim@postech.ac.kr

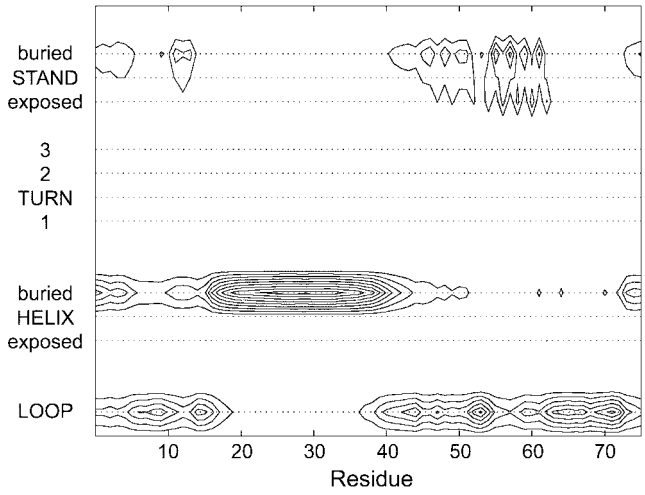


**Figure 3**. Secondary structure probabilities for SARS-CoV E protein from PSA Protein structure prediction.

**Table 1**. Sequence-structure homology recognition based on FUGUE

| Profile Hit | #aa | Zscore | ZSCORE: |
|---|---|---|---|
| hs1jdma | 31 | 3.85 | >6.0 (CERTAIN 99% confidence) |
| hs1khva | 493 | 3.60 | >4.0 (LIKELY 95% confidence) |
| hs1ce4a | 100 | 2.78 | >3.5 (MARGINAL 90% confidence) |
| hsd2occm1 | 43 | 2.69 | >2.0 (GUESS 50%) |
| hsd1cwva5 | 100 | 2.63 | <2.0 (UNCERTAIN) |

When the E protein is compared with corona-viruses, the pairwise sequence identity is less than 40%. The sequence alignment of all the E proteins among coronaviruses shows a clear feature that the aa are well conserved. This is more significant in the region of residues 12-39 (based on Bioedit) or 12-52 (based on ClustalX) (Figure 1).

The bioinformatics analysis of SARS-CoV E protein was done using public databases and the bioinformatics tools.[2] The BLAST and FASTA comparisons show significant matching. The 3D PSSM prediction for the protein structure reveals that E protein belongs to a class of membrane protein with high reliability of helix structure in the conserved region. The 3D PSSM results indicate that the region between 17 and 42 is a hydrophobic region of alpha helix type buried in the membrane. TMHMM predicts that the protein is a transmembrane type (aa: 12-34) with most of the hydrophilic domain consisting of the carboxy terminal tail inside the cell and a short hydrophilic amino terminus (aa: 1-11).

As for the secondary structure, PredictProtein (PP) predicts an $\alpha$-helix for aa: 14-43 and two extended $\beta$-sheets for aa: 47-52 and 56-63 (Figure 2). The secondary structure probabilities obtained from PSA Protein structure prediction (Figure 3) are $\alpha$-helix (aa: 19-38) and $\beta$-sheets (aa: 45-52 and 55-63) (Figure 3). Other programs predict these $\beta$-sheets as $\alpha$-helices or loops. Thus, this assignment is not clear yet, so it is being studied using simulation in our laboratory.

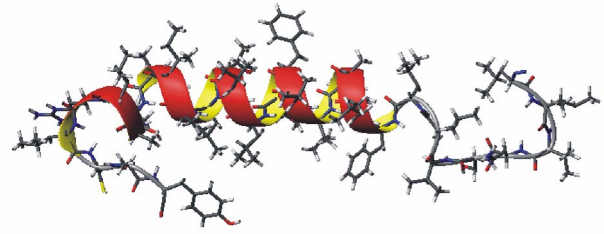

**Figure 4**. Simulated structure of $\alpha$-helix (aa: 12-40) in SARS. Note that aa:19-38 is an $\alpha$-helix.

The part of a structure that showed high conservation and transmembrane $\alpha$-helix type is taken and more systematically exploited by the homology recognition program FUGUE (Table 1). This part is taken to model the protein motif.

The molecular modeling of the $\alpha$-helix region has been investigated using molecular dynamics (MD) simulation.[3] The initial geometry was used with the NMR structure of hs1jdma which showes the highest Zscore. Taking aa: 11-42 as the starting structure, we have performed simulated annealing optimization[4] (Figure 4). The aa: 19-32 was calculated to be an $\alpha$-helix.

**Acknowledgment**. This research was supported by KISTEP(CRI) and BK21.

## References and Notes

1. (a) Rota, P. A.; Oberste, M. S.; Monroe, S. S.; Nix, W. A.; Campagnoli, R. *et al.*, *Science* **2003**, *300*, 1394. (b) Marra, M. A.; Jones, S. J. M.; Astell, C. R.; Holt, R. A.; Wilson, A. B. *et al.*, *Science* **2003**, *300*, 1399. (c) Anand, K.; Zeibuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. *Science* **2003**, *300*, 1763.
2. (a) BLAST and FASTA: *http://www.ncbi.nih.gov/* (b) FUGUE: *http://www-cryst.bioc.cam.ac.uk/fugue/* (c) 3D PSSM: *http://www.sbg.bio.ic.ac.uk/servers/3dpssm/* (d) TMHMM: *http://www.cbs.dtm.dk/services/TMHMM* (e) PP: *http://cubic.bioc.columbia.edu/predictprotein/* (f) PSA: *http://bmerc-www.bu.edu/psa/* (g) BioEdit: *http://www.mbio.ncsu.edu/BioEdit/bioedit.html* (h) ClustalX: *http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/*
3. (a) Son, H. S.; Hong, B. H.; Lee, C.-W.; Yun, S.; Kim, K. S. *J. Am. Chem. Soc.* **2001**, *123*, 514. (b) Kim, K. S. *Bull. Korean Chem. Soc.* **2003**, *24*, 757.
4. (a) Kim, K. S.; Dupuis, M.; Lie, G. C.; Clementi, E. *Chem. Phys. Lett.* **1986**, *131*, 451. (b) Kim, K. S. *Bull. Korean Chem. Soc.* **1993**, *14*, 18.