Journal of Biomolecular NMR **29:** 535–536, 2004.

Letter to the Editor: Backbone and side chain resonance assignments of domain III of the tick-borne Langat flavivirus envelope protein

Munia Mukherjee[a,b], Kaushik Dutta[c], Steven M. Pascal[d] & Robert O. Fox[a,b,*]

[a]*Department of Human Biological Chemistry & Genetics, Sealy Center for Structural Biology, University of Texas Medical Branch, Galveston, TX 77555-0647, U.S.A.;* [b]*W. M. Keck Center for Computational and Structural Biology, Rice University, Houston, TX 77005, U.S.A.;* [c]*New York Structural Biology Center, 89 Convent Avenue, Park Building at 133rd St., New York, NY 10027, U.S.A.;* [d]*Institute of Fundamental Sciences, PN 461 Massey University, Private Bag, 11 222, Palmerston North, New Zealand*

Received 18 December 2003; Accepted 8 March 2004

Key words: domain III, envelope protein, flavivirus, NMR assignments

Biological context

Flaviviruses have single stranded, positive sense RNA genomes that are 40 to 60 nanometers in diameter, 11 kilobase pairs in length, and the virions are icosahedral and enveloped. Flaviviruses cause many human diseases including dengue fever, yellow fever, West Nile virus encephalitis, hemorrhagic fevers, and are transmitted to their vertebrate hosts by infected mosquitoes and ticks (Boyle et al., 1987). Several flaviviruses are also recognized as potential bioweapons of mass destruction, including members of the tick borne encephalitis (TBE) serocomplex such as Russian spring-summer Encephalitis, Kyasanur forest disease and Omsk hemorrhagic fever. The flavivirus envelope protein E (50–60 kDa) is believed to play a dominant role in flavivirus entry and membrane fusion. The E-protein contains three distinct domains that have been identified both immunologically (Mackenzie, 1999) and via X-ray crystallography (Ray et al., 1995). Domain III (E-D3), the putative receptor binding domain is approximately 100 amino acids in length (depending on the flavivirus) and has been identified as the potential site for host-cell interaction (Ray et al., 1995). Domain III has a single disulfide bond and no glycosylation sites that makes the biologically relevant stable structure easily expressed in bacteria by recombinant technology. The solution structure of domain III of mosquito borne Japanese encephalitis E-protein, has been solved recently by NMR spectroscopy (Wu et al., 2003). We have initiated structural investigations of domain III of tick-borne Langat flavivirus E-protein (LGT-E-D3) in our efforts to design inhibitors to Langat and other pathogens of the TBE serogroup. Langat (LGT) virus is a naturally attenuated BSL-2 TBE virus and is a model for the pathogenic BSL-3 and BSL-4 viruses in the serogroup. Langat bears only 37.9% sequence identity (ExPasy: ALIGN) to the mosquito-borne JE flavivirus. As a first step towards obtaining the solution structure of LGT-E-D3, the sequential backbone and side chain hetero-nuclear chemical shift assignments of LGT-E-D3 are presented.

*To whom correspondence should be addressed. E-mail: fox@xray.utmb.edu

Materials and methods

Domain III of Langat flavivirus E-protein (LGT-E-D3) was expressed in *E. coli* strain BL21(DE3) Codon Plus (Stratagene) as a C-terminal fusion to an N-terminal six histidine-MBP (maltose binding protein) tag, with a thromibin cleavage site using plasmid H-MBP-T that encodes for residues 300-401 of LGT-E-D3. Details for the H-MBP-T vector are reported elsewhere (Alexandrov et al., 2001). Cells were grown at 37 °C in M9 minimal media with ^{15}N labeled NH_4Cl (1 g/L) and uniformly labeled ^{13}C glucose (3 g l^{-1}) producing $^{13}C/^{15}N$ labeled LGT-E-D3 sample. Cell lysis was achieved by sonication and the clear supernatent (per liter of growth) was incubated with 5ml Talon cobalt affinity resin (Clonetech, Palo Alto, CA) for 2 h. Extensive washing of the Talon resin by 10 mM Tris, 10 mM NaCl, pH 7.5 was followed by overnight

digestion of the His-MBP tag from LGT-E-D3 by adding 300 units of Thrombin and subsequent elution of the cleaved protein. Eluted protein was then passed through a HiTrap Q sepharose column (Amersham Biosciences) to remove fusion tag and other impurities. The Cys307-Cys338 disulfide bond was formed by air oxidation. The protein was found to be > 99% pure by SDS-PAGE and mass spectroscopy. The final $^{13}C/^{15}N$ LGT-E-D3 NMR sample was prepared by dialyzing against 10 mM Tris, 20 mM Bis-Tris, 10 mM NaCl, pH 6.2 and concentrated to ~0.7 mM. Similarly a uniformly labeled ^{15}N LGT-E-D3 (~0.8 mM) was also made. All NMR experiments were recorded at 25 °C on a Varian Inova 600 MHz spectrometer. Spectral data was processed using NMRpipe viewed by NMRDraw (Delaglio et al., 1995) and analyzed using NMRView (Johnson and Blevins, 1994). The ^{1}H chemical shifts were referenced to water signal at 4.75 ppm (at 25 °C) and the ^{13}C and ^{15}N chemical shifts were referenced indirectly using the $\gamma^{13}C/\gamma^{1}H$ and $\gamma^{15}N/\gamma^{1}H$ ratios respectively. The backbone sequential assignment were done using ^{1}H-^{15}N-HSQC, ^{1}H-^{13}C-HSQC, 3D^{-15}N-NOESY-HSQC, HNCO, HNCACB, CBCA(CO)NH (Sattler et al., 1999). The side chain carbon and proton assignments were obtained using C(CO)NH, HC(CO)NH, HBHA(CO)NH and HCCH-TOCSY experiments. Assignments of the aromatic resonance were obtained by analyzing ^{1}H-^{13}C HSQC, 3D ^{15}N TOCSY-HSQC and 3D ^{15}N NOESY-HSQC.

Extent of assignments and data deposition

In all, 99% of the backbone resonances (HN, N, Cα, Cβ and CO) were assigned with the exception of the carbonyl carbons of the residues immediately preceding the prolines in the sequence and the extreme C-terminal carboxyl carbon (Figure 1). We could not unambiguously assign the CO chemical shifts for S333 and ^{15}N and NH proton chemical shifts for G334. None of the resonances could be assigned for P377 that precedes P378 in the sequence. In total 1038 chemical shifts are reported, representing 89% of all resonances and 96% of resonances excluding the aromatic and guanidino side chain resonances. Chemical shifts have been deposited in the BioMagRes Bank Database (accession code BMRB-5971). The chemical shift deviation from random coil indicates that the secondary structure of LGT-E-D3 is composed of predominantly β-strand (Wishart and Sykes, 1994). This pattern of secondary structure is consistent with previously reported structures of E protein domain III from JE and TBE flaviviruses.

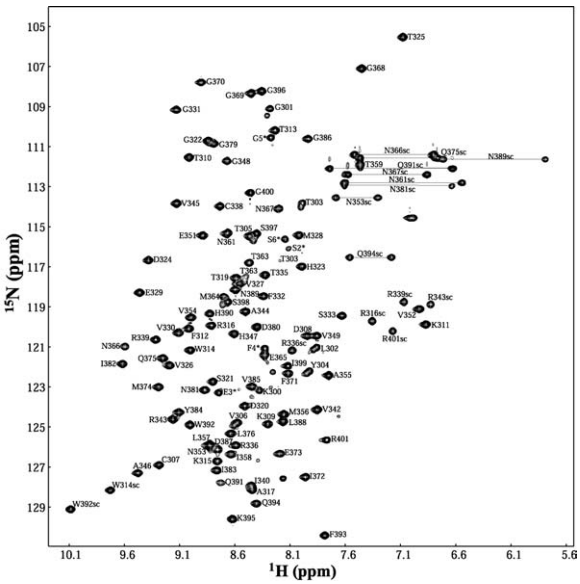

Figure 1. 2D ^{15}N-^{1}H HSQC of LGT-E-D3 at 25 °C and pH 6.2. Residues obtained from the cloning vector are indicated with an asterisk.

Acknowledgements

This research was supported by grants from DARPA and CDC to R.E. Shope, The Robert A. Welch Foundation (H-1345 to R.O.F) and the NIH (AI056326 to R.O.F). M.M. was supported by a training fellowship from the W.M. Keck Foundation of the Gulf Coast Consortia through the Keck Center for Computational and Structural Biology.

References

Alexandrov, A., Dutta, K. and Pascal, S.M. (2001) *Biotechniques*, **30**, 1194–1198.

Boyle, J.F., Weismiller, D.G. and Holmes, K.V. (1987) *J. Virol.*, **61**, 185–189.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Johnson, B.A. and Blevins, R.A. (1994) *J. Biomol. NMR*, **4**, 603-614.

Mackenzie, J.S. (1999) *Emerg. Infect. Dis.*, **5**, 1–8.

Rey, F.A., Heinz, F.X., Mandl, C.W., Kunz, C. and Harrison, S.C. (1995) *Nature*, **375**, 291–298.

Sattler, M., Schleucher, J. and Griesinger, C. (1999) *Prog. Nucl. Magn. Reson. Spectrosc.*, **34**, 93–158.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.

Wu, K.P., Wu, C.W., Tsao, Y.P., Kuo, T.W., Lou, Y.C., Lin, C.W., Wu, S.C., Cheng, J.W. (2003) *J. Biol. Chem.*, **278**, 46007–46013.