Biochimica et Biophysica Acta 1848 (2015) 1383–1392

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbamem

Genotype-specific differences in structural features of hepatitis C virus (HCV) p7 membrane protein

Monoj Mon Kalita [a], Stephen Griffin [b], James J. Chou [c], Wolfgang B. Fischer [a,*]

[a] Institute of Biophotonics, School of Biomedical Science and Engineering, Biophotonics & Molecular Imaging Research Center (BMIRC), National Yang-Ming University, Taipei, TW
[b] Leeds Institute of Cancer & Pathology, Faculty of Medicine and Health, St James' University Hospital, University of Leeds, Beckett St., Leeds, West Yorkshire LS9 7TF, UK
[c] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 29 November 2014
Received in revised form 12 February 2015
Accepted 6 March 2015
Available online 13 March 2015

Keywords:
p7 of HCV
Viral channel protein
Genotypes
Channel gating
Molecular dynamics simulation
NMR structure

ABSTRACT

The 63 amino acid polytopic membrane protein, p7, encoded by hepatitis C virus (HCV) is involved in the modulation of electrochemical gradients across membranes within infected cells. Structural information relating to p7 from multiple genotypes has been generated *in silico* (e.g. genotype (GT) 1a), as well as obtained from experiments in form of monomeric and hexameric structures (GTs 1b and 5a, respectively). However, sequence diversity and structural differences mean that comparison of their channel gating behaviour has not thus far been simulated. Here, a molecular model of the monomeric GT 1a protein is optimized and assembled into a hexameric bundle for comparison with both the 5a hexamer structure and another hexameric bundle generated using the GT 1b monomer structure. All bundles tend to turn into a compact structure during molecular dynamics (MD) simulations (Gromos96 (ffG45a3)) in hydrated lipid bilayers, as well as when simulated at 'low pH', which may trigger channel opening according to some functional studies. Both GT 1a and 1b channel models are gated *via* movement of the parallel aligned helices, yet the scenario for the GT 5a protein is more complex, with a short N-terminal helix being involved. However, all bundles display pulsatile dynamics identified by monitoring water dynamics within the pore.

1. Introduction

A growing number of viral genomes are being identified as encoding a specialised type of membrane protein which form oligomeric ion channels [1–6]. These viral channel forming proteins (VCPs), also called viroporins, pose an interesting challenge for the understanding of ion channel mechanics in general, due to their low molecular weight.

Similar to larger host-cell ion channels, the function of the VCPs is to modulate electrochemical gradients in various cellular subcompartments or at the site of the plasma membrane. This function is essential to the life cycle of many viruses. Some viruses encode bitopic membrane proteins, such as M2 of influenza A [7], Vpu of human immunodeficiency virus type-1 (HIV-1) [8], or 8a of severe acute respiratory syndrome-coronavirus (SARS-CoV) [9]. Other viral genomes harbour the code for polytopic membrane proteins such as p7 of hepatitis C virus (HCV) [10–12] and 2B of polio virus [13] with 2 transmembrane domains (TMDs), or 3a of SARS-CoV [14] and E5 of human Papillomavirus type 16 (HPV-16) [15] with 3 TMDs. A common feature of all these proteins is that they have to oligomerise to function as a channel.

Protein p7 of HCV is a 63 amino acid protein [16,17] essential for virion production ([18–21], reviewed in [22]). It is a cleavage product from a polyprotein precursor. Its location in the polyprotein follows the two structural proteins E1 and E2 (reviewed in [23]) and is succeeded by non-structural (NS) proteins. Protein p7, together with subsequent protein NS2, is found to act at an early stage of virion morphogenesis [21]. Its function is to support, possibly together with the crosslinked heterodimer E1 and E2 glycoproteins as well as NS2, capsid assembly at the plasma membrane [24]. Experimental results indicate that these proteins initiate the release of RNA-containing core proteins from lipid droplets.

As another function, p7 dissipates proton gradients which has been shown using pH sensitive dyes in cell based fluorescence spectroscopic experiments as well as liposome efflux essays [25]. This function as well as the capability of the protein to render lipid membranes permeable to ions is thought to be due to the protein forming channels, especially when expressed in cells, purified and reconstituted into artificial lipid bilayers [11,26]. Also, peptides derived from biochemical sources, generate adequate conductance data [10,12]. Overall, the channel shows weak cation selectivity [11,12].

In as much both functions as mentioned come along with different structural features are still under debate. NMR spectroscopic measurements have identified the monomeric protein with two TMDs in a hairpin motif with each of the helices being segmented into two smaller units [27,28] as well as the protein in an oligomeric state forming a

* Corresponding author at: Institute of Biophotonics, School of Biomedical Science and Engineering, National Yang-Ming University, 155, Sec. 2, Li-Nong St., Taipei 112, TW. Tel.: +88622826 7394; fax: +886 2 2823 5460.
E-mail address: wfischer@ym.edu.tw (W.B. Fischer).

http://dx.doi.org/10.1016/j.bbamem.2015.03.006
0005-2736/

hexameric bundle [29]. A feature of the bundle models built from the hairpin monomers is a ring of histidines (His-17s) pointing into the putative lumen of the pore [30]. The oligomeric architecture suggests an 'i + 3' motif in which the C terminal helix of one monomer interacts/links with the N terminal helix of the third monomer. There is currently no information of how the different architectures are related to each other. Negative stain transmission EM data mostly identify a hexameric assembly for the protein [11,31] as well as GT specific heptameric bundle [26].

The motif of the TMDs of p7, and all of today's known VCPs, is identified to be helical ([4] and references therein). Thus, secondary structure prediction programmes can be applied to identify the length of a putative TMD and the respective amino acids can be translated into a helical conformer. In case of polytopic proteins TMDs need to be assembled to form a monomer using a docking approach [32,33]. These structures form the basis for simulating functional data, such as ion conductance, which can be compared with those from experiments [34–39].

The different overall structural features, hairpin motif versus i + 3 motif, sparked the investigations reported in this paper. Bundle models of p7 of GT 1a were generated using the abovementioned protocols [32,33]. The hexameric bundle of GT 1b was generated using a monomeric NMR structure [40] and positioning six copies around a pseudo six-fold symmetry axis. The bundle of GT 5a derives from NMR spectroscopic data with the protein in an oligomerized (hexamer) form [29]. The three bundles were subjected to multi-nanosecond molecular dynamics (MD) simulations to elucidate structural and dynamic features at both, neutral and acidic pH, with the latter mimicked via protonation of His17 during simulation processing.

2. Materials and methods

The sequence of the TMDs of p7 was taken from the HCV GT 1a, H77 strain [10]: $ALENLVILNA^{10}$ $ASLAGTHGLV^{20}$ $SFLVFFCFAW^{30}$ $YLKGRWVPGA^{40}$ $VYAFYGMWPL^{50}$ $LLLLLALPQR^{60}$ AYA according to Patargias et al. [30]:

TMD1: $ALENLVILNA^{10}$ $ASLAGTHGLV^{20}$ $SFLVFFCFAW^{30}$ YL
TMD2: $WVPGA^{40}$ $VYAFYGMWPL^{50}$ $LLLLLALPQR^{60}$ AYA.
Loop (as predicted): KGR

The model according to GT 1b-J4 uses the following sequence (PDB ID: 3ZD0) [41,42]:

1b (J4): $ALENLVVLNA^{10}$ $ASVAGAHGIL^{20}$ $SFLVFFCAAW^{30}$ $YIKGRLAPGA^{40}$ $AYAFYGVWPL^{50}$ $LLLLLALPPR^{60}$ AYA.

The following amino acids were found in a helical motif corresponding to the respective TMDs (non-helical amino acids shown in italics):

TMD1: $ALENLVVLNA^{10}$ $ASVAGAHGIL^{20}$ $SFLVFF$*CA*AW^{30} YIK
TMD2: *GR*$LAPGA^{40}$ $AYAFYGVWPL^{50}$ $LLLL$*LALP*PR^{60} AYA.

GT 5a model uses (EUH1480; PDB ID: 2M6X) [29]: $GAKNVIVLNA^{10}$ $ASAAGNHGFF^{20}$ $WGLLVVTLAW^{30}$ $HVKGRLVPGA^{40}$ $TYLSLGVWPL^{50}$ $LLVRLLRPHR^{60}$ ALA.

This sequence contains 5 mutations of unconserved amino acids (T1G, C2A, A12S, C27T, C44S). The three helices extend from

H1: $VIVLNA^{10}$ ASAAGN
H2: F^{20} $WGLLVVTLAW^{30}$ $HVKGRLVPGA^{40}$ T
H3: WPL^{50} $LLVRLLRP$.

2.1. Generation of the protein models

The computational p7 bundles were generated in a stepwise manner according to protocols described previously [32,33]. Briefly, the two GT 1a TMDs are assembled into a monomer (monomer-1a) first, followed by assembling six copies of monomer-1a around a pseudo six-fold symmetry axis into a bundle (bundle-1a). The conformational space for both of the assemblies, the monomer and the bundle, is screened by varying the degrees of distance, rotational angle and tilt stepwise as follows: (i) inter helical distance in steps of 0.25 Å covering 7.5 Å to 13.0 Å for monomers and 11.75 Å to 20 Å for hexameric bundles; (ii) rotational angles around the helical axis in steps of 5° covering 360°; and (iii) tilt in steps of 2° covering −36° to +36°. At each position, the conformation of the side chains was introduced choosing the most likely conformation from an internal library of the MOE software suit (www.chemcomp.com). At this stage, 15 steps of steepest descent minimization followed to remove 'bad' (overlaps) van-der-Waals contacts. Any extensive minimization is compensated by a fine grained, in terms of step width of the three degrees of freedom such as distance, rotational and tilt angle, search of the conformational space. Potential energy of each conformer was evaluated, using AMBER94. The docking was done using a dielectric constant of $\varepsilon = 2$ to mimic the environment of the bilayer.

The structure of the bundle according to GT 1b (bundle-1b) was generated from a FLAG-tagged monomer (PDB ID: 3ZD0) [40]. The FLAG-tag was removed and the respective monomer used for pore assembly as described for bundle-1a. Bundle of the geneotype 5a (PDB ID: 2M6X, bundle-5a) was taken without modification for the MD simulations.

2.2. MD simulations

The monomers and the bundles were individually inserted into pre-equilibrated patches of POPC lipids (16:1–18:1 diester PC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) for which parameters of Chandrasekhar et al. [43] were adopted. Lipids which overlapped with the helices were removed. The protein–lipid system was hydrated and, after steps of minimization (2000 steps of steepest decent and 5000 steps of conjugated gradient), it was equilibrated for a total of 1.9 ns. Equilibration was achieved by gradually increasing the temperature from 100 K to 200 K and 310 K, whilst keeping the peptide fully restrained with $k = 1000$ kJ mol^{-1} nm^{-2}. The first two simulations (100 K and 200 K) were run for 200 ps each, the third simulation (310 K) was run for 500 ps. Holding the system at 310 K, the restraints, imposed by a force constant k on the peptide, were released in 2 steps ($k = 500$ kJ mol^{-1} nm^{-2}, $k = 250$ kJ mol^{-1} nm^{-2}), running each of the steps for 500 ps. MD simulation of the system is carried out using Gromos96 (ffG45a3) force field. The temperature of the peptide, lipid, and the water molecules were separately coupled to a velocity rescaling thermostat [44] with a coupling time of 0.1 ps and a Berendsen barostat with a coupling time of 2.0 ps during the early equilibration at temperatures 100 K, 200 K and 310 K. During the production run, the system was coupled to a Nosé–Hoover thermostat with a coupling time of 0.1 ps [45,46].

The simulation systems of the individual TMDs consisted of 122 lipids (6344 atoms) and 3655 water molecules (10,965 atoms), for simulating monomers 276 lipids (14,352 atoms) and 8746 water molecules (26,238 atoms), the bundle systems consisted of 488 lipids (25,376 atoms) and 14,640 water molecules (43,920 atoms).

2.3. Data analysis and hardware equipment

The software 'Fluxer' developed and supplied by Dr. Manuel N. Melo, University of Groningen, NL, was used to count the number of water molecules passing through the pore (http://cgmartini.nl/cgmartini/index.php/tools2/proteins-and-bilayers/223). The central membrane embedded region of the bundle was chosen from residues 14 to 24 of each of the individual monomers.

The simulations were run on an Acer i7-2600 workstation with 8 cores and submitted to the ALPS-Acer AR585 F1 Cluster in the National Center for High-Performance Computing (NCHC), Hsinchu, TW. Plots and pictures were made with Origin 8.5.1 and VMD 1.9, respectively