A molecular modelling approach to understand the effect of co-evolutionary mutations (V344M, I354L) identified in the PB2 subunit of influenza A 2009 pandemic H1N1 virus on $m^7$GTP ligand binding

Dipali Bhoye, Abhisek Kumar Behera and Sarah S. Cherian

Correspondence
Sarah S. Cherian
sarahcherian@icmr.org.in

Bioinformatics and Data Management Group, National Institute of Virology, Pune 411001, Maharashtra, India

The cap binding domain of the polymerase basic 2 (PB2) subunit of influenza polymerases plays a critical role in mediating the ‘cap-snatching’ mechanism by binding the 5’ cap of host pre-mRNAs during viral mRNA transcription. Monitoring variations in the PB2 protein is thus vital for evaluating the pathogenic potential of the virus. Based on selection pressure analysis of PB2 gene sequences of the pandemic H1N1 (pH1N1) viruses of the period 2009–2014, we identified a site, 344V/M, in the vicinity of the cap binding pocket showing evidence of adaptive evolution and another co-evolving residue, 354I/L, in close vicinity. Modelling of the three-dimensional structure of the pH1N1 PB2 cap binding domain, docking of the pre-mRNA cap analogue $m^7$GTP and molecular dynamics simulation studies of the docked complexes performed for four PB2 variants observed showed that the complex possessing V344M with I354L possessed better ligand binding affinity due to additional hydrogen bond contacts between $m^7$GTP and the key residues His432 and Arg355 that was attributed to a displacement of the 424 loop and a flip of the side chain of Arg355, respectively. The co-evolutionary mutations identified (V344M, I354L) were found to be established in the PB2 gene of the pH1N1 viral population over the period 2010–2014. The study demonstrates the molecular basis for the enhanced $m^7$GTP ligand binding affinity with the 344M–354L synergistic combination in PB2. Furthermore, the insight gained into understanding the molecular mechanism of cap binding in pH1N1 viruses may be useful for designing novel drugs targeting the PB2 cap binding domain.

INTRODUCTION

The pandemic influenza A H1N1 (pH1N1) virus that emerged in 2009 subsequently showed multiple waves of activity globally (Ellis et al., 2011; Keramarou et al., 2011; Truelove et al., 2011; Yang et al., 2011; Elderfield et al., 2015). Among various factors that could contribute to enhanced virulence and viral fitness of the virus, the acquisition of mutations in crucial functional sites of the viral proteins plays a significant role. The influenza virus RNA-dependent RNA polymerase (RdRp) is required for both transcription and replication of influenza viral RNA occurring in the nucleus of infected cells. The RdRp is a heterotrimer composed of polymerase acidic (PA), polymerase basic 1 (PB1) and polymerase basic 2 (PB2) subunits (Li et al., 2001; Resa-Infanate et al., 2011). The cap binding domain of the PB2 subunit plays a critical role as it mediates the ‘cap-snatching’ mechanism by binding the 5’ cap of host pre-mRNAs, in the initial step of viral mRNA transcription (Plotch et al., 1981; Li et al., 2001; Guilligay et al., 2008; Sugiyama et al., 2009). Several studies have identified host signatures and putative host-adaptive mutations in the PB2 subunit (Shinya et al., 2004; Mehle & Doudna, 2009; Yamada et al., 2010). Experimental studies have shown that mutations including T271A, H357N, T339K and R477G in PB2 enhanced polymerase activity in mammalian cells (Li et al., 2009; Bussey et al., 2010; Zhu et al., 2012). A single mutation in PB2, E627K, was found to contribute to high polymerase activity, enhanced replication, virulence in mice and adaptation of avian influenza H5N1 viruses to humans (Shinya et al., 2004).

The pH1N1 2009 virus is known to be a triple reassortant and of swine origin (Gibbs et al., 2009). The virus was found to possess PB2 and PA segments from North
American avian viruses and the PB1 segment of the human influenza A (H3N2) viruses (Garten et al., 2009; Medina & Garcia-Sastre, 2011). PB2 mutations referred to as the SR polymorphism (590S and 591R) helped the swine-origin influenza virus H1N1, overcome host restriction by enhancing viral replication (Mehle & Doudna, 2009). The pH1N1 virus was also reported to possess increased pathogenicity upon the acquisition of PB2-T388I mutation (Zhao et al., 2014). Other mutations have been noted in the PB2 protein of pH1N1 viruses that may have consequences for altered pathogenicity (Maurer-Stroh et al., 2010). The mutation K340N was found to frequently co-occur with the haemagglutinin (HA) D222G mutation associated with enhanced virulence in the pH1N1 viruses (Kilander et al., 2010; Moussi et al., 2013). Continuous tracking of single and co-evolutionary variations in the PB2 protein is thus vital in surveillance of the evolution and understanding of the pathogenic potential of the influenza virus. Further, co-evolutionary mutational positions gain significance if they are in close proximity, especially in the tertiary fold of the protein (Simonetti et al., 2013). Being associated with functional sites, the nature of residue co-evolution is such that it should maintain the overall functionality and structural integrity of the protein (Chakrabarti et al., 2010).

The middle domain of the PB2 subunit (residues 318–483) is identified as the cap binding domain (Tarendeu et al., 2007; Guilligay et al., 2008). The crystal structure of the cap binding domain of the pH1N1 virus is not yet available in the protein data bank (PDB). However, the cap binding domain of the PB2 protein of avian influenza A H5N1 (A/Duck/Shantou/4610/2003), human influenza A H3N2 (A/Victoria/3/1975) and human seasonal influenza A H1N1 (AH1N1) (A/Peruto Rico/8/1934) either in unbound form or co-crystallized with pre-mRNA cap analogue m<sup>7</sup>GTP sepharose are available. The structure of the PB2 cap binding domain exhibits a new fold, consistent with the absence of homology to other proteins (Guilligay et al., 2008; Tsurumura et al., 2013). Despite this, m<sup>7</sup>GTP bound in a mode that is commonly observed in other cap binding proteins, whereby aromatic side chains of the active site residues in the cap binding pocket sandwich the positively charged aromatic guanine ring of the m<sup>7</sup>GTP (Ishida et al., 1988; Darzynkiewicz & Lonberg, 1989; Ueda et al., 1991; Stolarski et al., 1996). The interacting residues are completely conserved in all influenza strains, and their mutation abolished cap-dependent transcription but not cap-independent replication in recombinant mini-ribonucleoproteins (Guilligay et al., 2008).

In the present study, to identify possible sites of adaptive evolution in the PB2 gene of the pH1N1 viruses, we carried out selection pressure and co-evolutionary mutation analyses of PB2 gene sequences of influenza viruses of the period 2009–2014. In the absence of detailed structural information on the pH1N1 cap binding domain, three-dimensional (3D) structure modelling of the pH1N1 PB2 cap binding domain using homology modelling, subsequent docking of m<sup>7</sup>GTP and molecular dynamics (MD) simulation studies of docked complexes were performed to understand the role of amino acid mutational sites specifically in the PB2 cap binding domain, which has been shown recently as a promising target for the discovery of novel drugs against the influenza A viruses (Clark et al., 2014; Byrn et al., 2015). MD simulations have been used in several recent studies to analyse the extent of motions and conformational changes of functional sites to understand the flexibility of ligand binding cavities and specifically loop regions, both of which are known to govern the biological function of the viral protein (Buragohain et al., 2008; Purohit et al., 2011; Pan et al., 2012; Behera et al., 2015a, b).

### RESULTS

#### Phylogeny, selection pressure and co-evolution analyses

The neighbour-joining (NJ) based phylogenetic tree obtained from the aligned PB2 gene sequences (n=642), showed a prominent group (n=225) encompassing the pH1N1 strains (Fig. 1) with a strong bootstrap of 100% (Fig. S1, available in the online Supplementary Material). This group was noted to include additional sequences from swine H3N2 (6 isolates), swine H1N1 (28 isolates), swine H1N2 (5 isolates), swine H3N1 (1 isolate), avian H3N2 (2 isolates), avian H1N1 (2 isolates), canine H3N1 (1 isolate), cheetah H1N1 (1 isolate) and sloth bear H1N1 (1 isolate) along with the human pH1N1 isolates (n=178). An inspection of the GenBank entries of these virus isolates revealed that they were reported as reassortments possessing the PB2 segment of the pH1N1 viruses (Pasqua et al., 2013). The percentage nucleotide identity within the group was found to be 99.93%.

The group (n=225) mainly comprising the pH1N1 gene sequences was analysed for detection of positive selection pressure. Site-specific selection pressure was examined by using likelihood procedures, single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and internal fixed effects likelihood (IFEL). High statistical significance was defined to be P=0.1 and sites were considered to be under positive selection if at least two of the methods indicated selection satisfying this criterion. Accordingly, four codon sites at positions 227, 255, 344 and 461 were determined to show evidence of evolving under positive selection pressure (Table 1). Among these four amino acid residue positions, 227 and 255 fall in the N-terminal region of PB2 while residues 344 and 461 fall specifically in the PB2 cap binding domain (residues 318–483). Two crystal structures [2vqz.pdb (strain: A/Victoria/3/1975/H3N2; resolution: 2.30 Å) and 4cb4.pdb (strain: A/duck/Shantou/4610/2003/H5N1; resolution: 1.60 Å)] were available in PDB for WT PB2 cap binding domain bound with m<sup>7</sup>GTP. Upon search of the query pH1N1 PB2 sequence using PDB BLAST, the cap binding domain of the PB2 protein sequence of pH1N1 showed maximum identity with 4cb4.pdb (99%) compared
with 2vqz.pdb (89 %), and the former was thus selected as the template structure for further studies.

Co-evolutionary analysis using the mutual information server to infer co-evolution (MISTIC), inputting both the PB2 alignment (n=225) and the PDB structure file (4cb4.pdb), revealed that in the top 5 % rank, only a single amino acid pair, 344–354, was detected (Fig. S2). On mapping of sites 344 and 354 on the crystal structure (4cb4.pdb), it was found that residue position 344 was in the vicinity of the cap binding groove (~8 Å from the active site residues) while the atomic distance between the two residues was ~4.69 Å (Fig. 2a). Moreover, the position 354 was within the cap binding pocket.

Observing the conservation of amino acid residues in the alignment of the PB2 sequences (n=225) at the 344 position revealed that the majority of the sequences had Val (n=156) at this position, followed by Met (n=68), and one sequence had Leu at this position. Mutated residues at position 354...

**Fig. 1.** Representation of the phylogenetic tree based on the PB2 gene sequences of the 2009 pandemic H1N1 sequences of the period 2009–2014 (n=225) reconstructed using the NJ method. Bar represents 0.005 substitutions per site. All strains in the sub-cluster, with bootstrap support 91, possess the co-evolved mutations V344M and I354L. Except for few strains (n=6) all other strains in the tree possess V344 and I354 (n=173).
included Iso (n=159) and Leu (n=61). Only three strains were found to possess Met and two strains had Val at position 354. Notably, a robust sub-cluster with bootstrap support 91% could be identified in the phylogenetic tree (Fig. 1), with all the constituent strains possessing the mutations V344M and I354L.

### Table 1. Codon sites identified to be under positive selection pressure (shown in bold type) in the PB2 gene sequences of the 2009 pandemic influenza A/pH1N1 at a significance level (P<0.1) by at least two of the methods, SLAC, FEL and IFEL, available at the Datamonkey server

<table>
<thead>
<tr>
<th>Codon Site</th>
<th>SLAC</th>
<th>FEL</th>
<th>IFEL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P value</td>
<td>Normalized</td>
<td>P value</td>
</tr>
<tr>
<td>227(V/I/S/A)</td>
<td>0.132</td>
<td>0.473</td>
<td>0.049</td>
</tr>
<tr>
<td>255(V/I/F)</td>
<td>0.199</td>
<td>0.378</td>
<td>0.079</td>
</tr>
<tr>
<td>344(V/M/L)</td>
<td>0.134</td>
<td>0.589</td>
<td>0.037</td>
</tr>
<tr>
<td>461(I/V)</td>
<td>0.226</td>
<td>0.366</td>
<td>0.092</td>
</tr>
</tbody>
</table>

\(d_S\), Synonymous site; \(d_N\), non-synonymous site.

### Homology modelling and docking

Alignment of mutant sequences (indicated as CBD1-4) that were selected based on residues at positions 344 and 354 along with the sequence of the template selected for homology modelling in the region corresponding to the PB2 cap binding domain is shown in (Fig. 2b). Validation of the

---

**Fig. 2.** (a) Mapping of site at residue position 344 under positive selection pressure in the PB2 protein along with the co-evolving residue at position 354 on the crystallographic structure of the PB2 cap binding domain with bound m^7^GTP (4cb4.pdb). The atomic distance between the co-evolving residue pair is 4.69 Å. The m^7^GTP ligand is labelled in the figure alongside. (b) Alignment of the PB2 protein cap binding domain (residues 318–483) of representative strains of pH1N1 (CBD1: A/California/07/2009/USA/H1N1; CBD2: A/Ankara/WRAIR1435T/2009/Turkey/H1N1; CBD3: A/Mexico/INER0958/2012/Mexico/H1N1 and CBD4: A/Chile/6/2010/Chile/H1N1) with the sequence of 4cb4.pdb (A/Duck/Shantou/4610/2003/HSN1). The region corresponding to the 424 loop in the PB2 structure is represented in violet ribbon colour in (a) and a box in (b).
modelled 3D structures using the SAVES server showed 100 % residues in the favourable region (93.1 % in the core region and 6.9 % in the generously allowed region) of the Ramachandran plot. The root mean square deviation (RMSD) between the template and homology models was 0.06 Å in all cases.

Docking of m7GTP with the four models of the cap binding domain showed the highest Goldscore (91.42) for m7GTP docked to CBD3 and lowest score (78.37) for m7GTP docked to CBD1. The Goldscores obtained for m7GTP–CBD4 and m7GTP–CBD2 were 90.41 and 86.98, respectively. The interaction energy analysis of the four docked complexes showed that the m7GTP–CBD2 complex had the lowest free energy (−10.28 Kcal mol−1) compared with the other protein–ligand complexes [m7GTP–CBD1 (−5.68 Kcal mol−1), m7GTP–CBD3 (−9.36 Kcal mol−1), m7GTP–CBD4 (−8.15 Kcal mol−1)].

**Molecular dynamics simulations**

The structural stability of the four simulated systems (m7GTP–CBD1–4) analysed by the RMSD of the backbone Cα atoms over the 10 ns time period, showed that the structures attained stability after about 2 ns, after which the fluctuations were within 0.5 Å (Fig. 3a). Interaction energy values of simulated complexes at 10 ns showed that the CBD3–m7GTP complex was energetically the most stable complex (−14.79 Kcal mol−1). Residues forming intermolecular hydrogen bonds (H-bonds) and hydrophobic contacts at the two time points (0 ns and 10 ns) have been listed in Table S1 while Fig. 4 shows the interacting residues of the ligand in the cap binding domain pocket at the end of the 10 ns simulation. 3D surface visualizations of all four complexes, were also generated (Fig. 5), in terms of the m7GTP pose in the binding pocket.

**DISCUSSION**

An abundance of host-adaptive mutations in the PB2 subunit has been predicted previously, though for most of these mutations, little is known about their mode of action (Kilander et al., 2010). Although the E627K mutation was proved experimentally to be a major virulence marker, results of replication assays of non-pathogenic and mouse-adapted variants have helped identify other substitutions in PB2 that can increase virulence by enhancing the polymerase activity in mouse cells in spite of the presence of a PB2 K627E reverse mutation. This also supports studies suggesting that the presence of either Glu or Lys at PB2 residue 627 could retain the virulence of viruses with certain genetic background in mammalian models (Naffakh et al., 2000; Labadie et al., 2007). In this light, the monitoring of compensatory or co-evolutionary mutations becomes essential. In the present study, we attempted to identify functional co-occurring mutational sites in the PB2 protein of pH1N1 viruses and further adopted molecular modelling and simulation approaches to understand the mechanism by which mutating residues in the PB2 may exert their effect.

Phylogenetic analysis of the PB2 gene sequences of the N1 and N2 viruses of the period 2009–2014 helped identify the sub-group corresponding to the pH1N1 viruses while, based on selection pressure analysis, we identified an amino acid site 344V/M in the vicinity of the cap binding pocket showing evidence of adaptive evolution in the pH1N1 cluster. Mutual information co-evolution analysis could detect only one co-evolving residue pair (344, 354) in the PB2 protein dataset with high statistical significance. Further, 3D-structure modelling of the pH1N1 PB2 cap binding domain, docking of the pre-mRNA cap analogue m7GTP and MD simulation studies of the docked complexes were performed for the four dominantly observed PB2 variants (CBD1: Val344, Ile354; CBD2: Val344, Leu354; CBD3: Met344, Leu354; CBD4: Met344, Ile354).

The final conformation of the CBD1–m7GTP complex indicated only three H-bonds with key residues (two with Arg355 and one with Phe404) and eight hydrophobic contacts with residues Ile354, Met431, Pro430, Asn429,
Gln406, Phe363, Phe323 and His357 (Fig. 4a). In two other complexes, it was noted that among critical H-bond contacts, only two of the key residues were involved (two H-bonds with Arg355 and one H-bond with Asn429 in the CBD2–m7GTP complex and one H-bond with Glu361 and three H-bonds with Asn429 in the CBD4–m7GTP complex). On the other hand, the final conformation of the m7GTP in the CBD3–m7GTP complex possessed nine intermolecular H-bond contacts and eight hydrophobic contacts (Table S1). N2 and O6 of the 7-methyl guanine ring formed H-bonds with key residues (Guilligay et al., 2008) Glu361 and Lys376, respectively (Fig. 4c). The hydroxyl group of the ribose sugar also formed an H-bond contact with another key residue, His357. The triphosphate chain formed a total of six H-bonds (γ-phosphate: three H bonds with key residue Arg355 and one H-bond with Met431; β-phosphate: one H-bond with Arg355; α-phosphate: one H-bond with His432). Residues Phe323, Phe363 and Ser337 formed hydrophobic contacts with the 7-methyl guanine ring; residues Phe404 and Phe325 formed hydrophobic contacts with ribose sugar, and residues Pro430, Leu354 and Gln406 formed hydrophobic contacts with the triphosphate chain. Thus, it can be noted that among the four complexes, the CBD3–m7GTP complex, which was found to be energetically more favourable, may possess better ligand binding affinity due to additional hydrogen bond contacts between m7GTP and key residues.

Upon 3D surface visualization of all four complexes, somewhat different conformations of m7GTP in the cap binding pocket were observed (Fig. 5). The conformations were also compared with that in the native/template structure 4cb4.pdb (Fig. S3a). Additionally, in silico docking of the PB2 cap binding domain extracted from the template crystal structure (4cb4.pdb) with m7GTP followed by MD simulation showed that the conformation at the 10 ns time period possessed comparable H-bond and hydrophobic contacts (Fig. S3b). The RMSD value between the crystal structure of m7GTP bound in the cap-binding domain of influenza A/ H5N1 (4cb4.pdb) and the in silico modelled structure was also less than 1 Å, validating the reliability of the model calculated.

As reported previously (Guilligay et al., 2008; Pautus et al., 2013), in all cases, 7-methyl guanine (m7G) is sandwiched

---

**Fig. 4.** Residues forming H-bonds and hydrophobic contacts in the different PB2 cap binding domain variants [(a) V344, I354 (CBD1), (b) V344, L354 (CBD2), (c) M344, L354 (CBD3), (d) M344, I354 (CBD4)] docked with m7GTP. Residues forming the H-bonds are in cyan colour while the H-bonds are depicted with yellow lines and residues forming hydrophobic contacts are in grey colour.
between the two aromatic residues Phe363 and His357, while basic residues (Arg355 and/or His432) are involved in triphosphate stabilization. In most cases, m\textsuperscript{7}G binds in the pocket formed by His357, Phe404, Glu361 and Lys376; triphosphate is bent around the base with the α-phosphate interacting with His432/Asn429 and the β/γ-phosphate interacting with Arg355 (Fig. 4). Lys339, which forms another significant H-bond contact with ribose in the crystal structure (Fig. S3a) was found to be absent in all the complexes modelled.

A study of the docking poses of m\textsuperscript{7}GTP in the four complexes (Fig. 5) revealed that the pose of the ligand in CBD3 is noticeably similar to that seen in the crystal structure complex, 4cb4.pdb (Fig. S3a). We analysed the root mean square fluctuation (RMSF) for each residue to identify the regions of high fluctuations. The highest fluctuations were observed at three residue positions of the CBD3–m\textsuperscript{7}GTP complex, Arg423 (2.48 Å), Arg427 (2.15 Å) and Val421 (2.22 Å), and in the CBD4–m\textsuperscript{7}GTP complex, at the Arg368 (1.96 Å) position (Fig. 3b). Notably, the residues 420–427 form part of the 424 loop in the cap binding domain structure, the integrity of which is required for cap-dependent transcription. The 424 loop is also reported to have an allosteric role in regulating the activity of the PB1 subunit (Guilligay et al., 2008). As a consequence of the displacement of the 424 loop outwards, significant H-bonds are formed with Met431 and His432 in CBD3 (M-L combination) with m\textsuperscript{7}GTP. Uniquely, the M-I case (CBD4) possesses a contact with His432 in the form of a hydrophobic interaction as in the crystal structure, while in the V-L (CBD2) case an H-bond exists with Met431. The M-I case is also unique in that Arg355 does not form any contact with m\textsuperscript{7}GTP. The acidic residue Glu361, which is typically involved in base recognition and charge neutralization was noted to form H-bonds in CBD3 and CBD4 only, as also in the crystal structure (4cb4.pdb). The residue at position 354 did not form any contact in the crystal structure. However, except in the M-I case, it was found to form hydrophobic interactions with the triphosphate group. Interestingly, the side chain of Arg355 was found to be flipped inwards towards the triphosphate in the case of CBD3 (Fig. 4c), as a result of which it formed four H-bonds with m\textsuperscript{7}GTP. The side chain of Arg355 was found to face outwards in the other cases, including in the crystal structure (4cb4.pdb). H-bond contacts with ribose were noted only in CBD1 (V-I) and CBD3 (M-L), with residues Phe404 and His357, respectively. Overall, the H-bonds in the cases of CBD3 and CBD1 included bonds with triphosphate and

![Fig. 5. Conformation of m\textsuperscript{7}GTP in the cap binding pocket of the four different PB2 cap binding domain variants, (a) CBD1, (b) CBD2, (c) CBD3 and (d) CBD4, of pH1N1 viruses at the end of the 10 ns MD simulation.](http://jgv.microbiologyresearch.org)
methyl guanine as well as ribose of m^7GTP, as in the crystal structure. In the case of CBD2, H-bonds were only formed with the triphosphate group while in the case of CBD4 three H-bonds were with triphosphate and methyl guanine. Thus it can be noted that the combination of mutations at position 344 (V→M) and 354 (I→L) have a synergistic effect in terms of contributing to increased interactions with the ligand as well as the shape of the binding pocket resulting in better space-occupying effects, markedly as in the native structure. Analysis of the ligand interactions revealed that although the residue at position 344 is not directly involved in the m^7GTP binding, the presence of the bulkier side chain of Met at this position in comparison to Val results in the displacement of the 424 loop favouring interactions of the ligand with His432. On the other hand, the replacement of Ile with Leu at position 354 most likely results in reduced steric hindrance enabling vital interactions with Arg355. This residue position was also found to contribute to a hydrophobic interaction with m^7GTP. Several studies have shown the significance of hydrophobic interactions in the formation of stable protein–ligand complexes (Pyrkov et al., 2007; Patil et al., 2010). The influence of sites in the vicinity of the ligand binding active sites has also been reported to play a vital role (Yen et al., 2006; Bradley et al., 2011). Overall, the co-evolved mutations at positions 344 and 354 resulted in the energetically more favourable binding energy and thus enhanced stability. The docking and simulation studies for the V-L and M-I variants revealed that they possessed higher m^7GTP binding energy compared with the other cases (Figs 5a, c and S3a).

The PB2 cap binding domain of influenza/A/H5N1 is, overall, conserved when compared with that of the pH1N1 virus, with residues at positions 344 and 354 being Val and Ile, respectively. The P453S and V478I mutations noted in the H5N1 PB2 cap binding domain as in the template selected for modelling in this study (Fig. 2b) are away from the cap binding pocket and hence do not have a major influence on the conformation of the m^7GTP binding pocket. However, observations of the contact residues of CBD1–4 with m^7GTP showed that Lys339 does not feature in the contact residues of the m^7GTP binding in the pH1N1 viruses, as was noted in the H5N1 complex. Thus, in spite of the similarity between the conformation of the m^7GTP in the H5N1 cap binding pocket and the case of CBD3, there are some differences. The best posed m^7GTP conformation in CBD3 compared with the four different cap binding domain PB2 variants presented in this study has the following features: (1) the side chain of Arg355 flipped to hold the triphosphate; (2) the bend of γ-phosphate to interact with His357 was not observed, instead interaction of ribose with His357 was noted; and (3) no interaction of m^7GTP with Lys339 in terms of either hydrophobic interaction or H-bond.

A scrutiny of all the PB2 sequences of pH1N1 viruses from the human host as available in GenBank revealed that strains possessing residues V344 and I354 were rapidly replaced by the M344/L354 combination over the period 2010–2012 and fixed in the population by 2014 (Fig. 6). Geographical distributions of the co-evolving residue positions (344/354), over the period 2009–2014, were also studied by scrutinizing representative PB2 sequences of pH1N1 viruses from the human host obtained by removing identical sequences and considering a country-wise representation for the different regions (Fig. 7). Although at any time point there is predominance of either of the major variants (V344/I354 and M344/L354), globally, it can be noted that during the years 2010 and 2011 there is a certain amount of co-circulation, with Europe showing almost equal proportion of both variants during 2010. Further, more than 95% occupancy of virus strains possessing the M344/L354 combination is noted since the year 2012 at a global level. The V344M mutation was reported (Zamboni et al., 2012) in a post-pandemic period wave (Yang et al., 2011) of the 2009 pH1N1 influenza in Mexico and Central America during 2012, as was also found to occur in the sequences of isolates from the 2010–2011 influenza season in Asia and Europe (Fig. 7b–d). Strains possessing the M344/L354 mutations were found to be delineated phylogenetically into a well-defined cluster with strong bootstrap support (Fig. 1). Notably, the proportion of strains possessing the other two combinations of residues at positions 344 and 354 was less than 1% in the case of 2009 pH1N1, which corroborates the results of this study with regard to the fitness of the M344/L354 variant. No such delineation can be observed in other common human influenza subtypes such as sH1N1, H3N2 and H5N1. In the case of representative sequences of PB2 for the subtypes H3N2 and H5N1 (from human isolates), the V344/I354 combination was found to be predominant (Fig. S4b, c). In the case of influenza sH1N1 isolates over the period 1998–2011, the dominance of V344/L354 was noted until the year 2009 with exceptions in 2001 and 2002. Notably, during 2001–2002 and 2010–2011, more than 50% strains possessing the M344/L354 combination were in
circulation (Fig. S4a), implying that this combination may provide a synergistic advantage to the virus. Although the complete biological relevance of these and other novel mutations being noted is yet unknown; none the less, recent reports do mention the V344M and I354L mutations in the PB2 gene suggesting that the said combination could be beneficial for global evolution of influenza A (H1N1) pdm09 viruses either through enhancement of viral fitness or virulence (Belanov et al., 2015; Mishel et al., 2015). Further, strains possessing these mutations were shown to enhance virus replication in human cells and mice and contribute to an increased severity of pH1N1 virus infections in patients admitted to hospitals in

Fig. 7. Graph showing the geographical distribution of the co-occurring mutations: V344/I354 (blue); V344/L354 (purple); M344/L354 (red); M344/354I (green); L/V/M344/354I/V/M (grey) based on representative sequences of the PB2 gene of pandemic influenza A H1N1 during the years (a) 2009, (b) 2010, (c) 2011, (d) 2012, (e) 2013 and (f) 2014.
England during 2009–2011 (Elderfield et al., 2015). Thus, our work, through the MD simulation studies and in-depth analyses to monitor the dynamic behaviour of the PB2–mGTP molecular complexes and comparison with the experimental crystal structure complex, further substantiates these observations by providing a molecular basis for the enhanced fitness.

To conclude, this study has identified co-evolutionary mutations in the PB2 cap binding domain of pH1N1 viruses with I354L having a synergistic effect on the V344M substitution. Our results are indicative of enhanced mRNA cap-binding affinity facilitated by the co-occurring mutations that might have contributed to increased viral fitness in subsidiary waves of the pandemic by mediating increased efficiency of cap binding and thus the replication competence of these viruses. Although experimental evidence of the direct consequences of these mutations would be important to fully establish their role, the insight gained into understanding of the molecular mechanism of cap binding in pH1N1 viruses as obtained from this study may be useful for designing novel antivirals targeting the cap binding domain.

**METHODS**

**Phylogenetic analysis.** PB2 gene sequences were retrieved from the influenza virus resource database at NCBI (https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database). Sequences of influenza A (N1 and N2 subtypes) of the time period 2009–2014 that totaled 7154 were included as the starting dataset in this study. Identical sequences were removed and representative sequences were selected based on the host, country and year of isolation. The reduced dataset of 642 nucleotide sequences was aligned using ClustalW and phylogenetic analysis was conducted using the NJ method in MEGA version 5.1 (Tamura et al., 2011) with the Kimura two-parameter model of nucleotide substitution. Tree reliability was checked by using the bootstrapping method with 1000 replicates. Genetic distances were also calculated in MEGA version 5.1 using the Kimura two-parameter model to identify the definite cluster for the pH1N1 sequences. This dataset consisted of 225 sequences including a majority of pandemic human H1N1 isolates as well as few sequences of other subtypes and other hosts.

**Selection pressure and co-evolutionary mutation analyses.** Selective pressure on individual sites of codon alignments of the sequences in the dataset as derived above, (n=225), was detected based on the ratio of non-synonymous (da) to synonymous (ds) substitution rates. The gene is said to be undergoing positive selection when the rate of non-synonymous substitutions per potential non-synonymous site (da) is greater than the rate of synonymous substitutions per potential synonymous site (ds) (Yang, 2001).

SLAC, FEL and IFEL tests, three likelihood procedures, were performed using Datamonkey, a web-based interface of the HyPhy software package (http://www.datamonkey.org). The random effects likelihood (REL) test was not used in this study due to the alignment data set size restrictions (Sergiev et al., 2005, 2006). In all cases, da/ds estimates were based on the nucleotide substitution model string (010230), the best fitting nucleotide substitution model and P value set at 0.1.

Co-evolutionary mutational positions in PB2 sequences were also identified using a web-based server, MISTIC (http://mistic.ledoir.org). The results from the server include the information of the multiple sequence alignment (MSA) condensed into a circos representation and a mutual information (MI) network, a distance network based on MI score if a reference protein structure is provided (Simonetti et al., 2013).

**Homology modelling and structure validation.** Four PB2 variants of pH1N1 viruses observed over the period 2009–2012 were selected for homology modelling studies, based on the selection pressure result in the PB2 cap binding domain, visual inspection of the PB2 alignment in the cap binding domain region and identification of co-evolutionary positions linked to positive selection pressure sites. The swiss-model web server accessed from the ExPasy site (http://swissmodel.expasy.org) was used for homology modelling, a tertiary structure-prediction method. It works at its best when the protein with unknown structure (target) shares substantial sequence homology with a protein of known structure (the template). As per rule of thumb, in order for homology modelling to be successful, >35% sequence identity between the template and target proteins is desired so that the root mean square error will be <2Å (Sanchez & Sali, 1997). The crystal structure of A/Duck/Shantou/4610/2003 (4cb4.pdb) showed 98% sequence identity with the target sequence and was used as the template in the present study. The strains selected to build the 3D model of the PB2 cap binding domain possessed mutations as follows: V344 and I354 in A/California/07/2009/USA/H1N1; V344 and L354 in A/Ankara/WRAIR1435T/2009/Turkey/H1N1; M344 and L354 in A/Mexico/INEB0958/2012/Mexico/H1N1; and M344 and I354 in A/Chile/6/2010/Chile/H1N1; and 3D models were named as CB1, CB2, CB3 and CB4, respectively. The modelled structures were validated in the Structure Analysis and Verification Server (SAVES) using PROCHECK (http://services.mbi.uc.edu/SAVES/). PROCHECK determines the Ramachandran plot by calculating the main-chain torsional angles for the predicted structures (Laskowski et al., 1993).

**Docking of mGTP to the cap binding domain and MD simulations.** GOLD suite version 5.2 (Jones et al., 1997) was used for rigid body docking of mGTP with the modelled structures of the cap binding domain of the PB2 protein for the four selected strains. The central residue in the cap binding domain (Phe325) with atoms in a 10 Å radius was selected as the reference for the docking protocol. Water molecules were removed from the protein and both the molecules were hydrogenated. GOLD docking is based on a genetic algorithm in which the parameters optimized are (a) dihedrals of rotatable bonds and ring geometries of the ligand dihedrals of protein OH groups and NH3 groups, and (b) mappings of the position of the ligand in the binding site. GOLD generates 10 best low-energy docking solutions within the binding site for the ligand and the Goldcore ranks the docked complexes based on the binding mode and affinity of the ligands (Verdonk et al., 2003). The binding energy of the top-ranking complex was studied using the Program of Enhetic and Receptor Ligand System (PEARLS) server (Falkov et al., 2000; Wang et al., 2004).

The best-docked complex of PB2–mGTP was further considered for MD simulations using the Yasara package version 13.6.16 (YASARA Biosciences). MD simulations are used to sample the conformation space of ligand–protein interaction systems by solving equations of motion (van Gunsteren & Berendsen, 1990). The TIP3P water model was used to solvate the simulation system in a periodic simulation cell with boundaries extending 20 Å from the surface of the complex simulation box. To neutralize the solvated systems, counter ions (Cl– and Na+) with a physiological strength of 0.15 M were added in the simulation box. Further, the structure was energy-minimized using the YAMBER3 (Yet Another Model Building and Energy Refinement force field 3) force field (Krieger et al., 2004) using a 7.86 Å cut-off and the particle mesh ewald (PME) method to treat long-range electrostatic interactions. Conformational stress was removed by following a short steepest descent minimization before initiation of simulation and the procedure continued by simulated annealing (integration time steps were set to 2 and 1.25 fs for intra and intermolecular forces, respectively, atom velocities scaled down by 0.9 every tenth step) until convergence was reached, i.e. the energy improved by less than 0.03 kJ mol–1 per atom during 200 steps. Minimization was followed by an equilibration procedure using the NVT ensemble to 298K. The resulting minimized and equilibrated
models were used for MD simulations. The simulations were run at constant pressure (NPT ensemble); temperature was controlled by rescaling atom velocities using a Berendsen thermostat (Berendsen et al., 1984). The production simulations were carried out for 10 ns and snapshots of trajectories were captured at each 1 ps time interval. Analyses of the MD trajectories for the four complexes in terms of RMSD and RMSF calculations were performed using YASARA and the Grace plotting tool (http://plasma-gate.weizmann.ac.il/Grace/).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Dr D. T. Mourya, Director, National Institute of Virology (NIV), for the facilities and support provided. D. B. was supported by Junior Research Fellowship from the University Grants Commission/Rajiv Gandhi National Fellowship (UGC/RGNF).

REFERENCES


