An all-atom, active site exploration of antiviral drugs that target Flaviviridae polymerases

James J. Valdés,¹,² Victor A. Gil,³ Philip T. Butterill⁴ and Daniel Růžek¹,²,⁴

¹Institute of Parasitology, Czech Academy of Sciences, Branišovská 31, CZ-37005 České Budějovice, Czech Republic
²Department of Virology, Veterinary Research Institute, Hudcova 70, CZ-62100 Brno, Czech Republic
³Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Jordi Girona 29, 08034 Barcelona, Spain
⁴Biology Center, Czech Academy of Sciences, University of South Bohemia, Branišovská 31, CZ-37005 České Budějovice, Czech Republic

Natural 2'-modified nucleosides are the most widely used antiviral therapy. In their triphosphorylated form, also known as nucleotide analogues, they target the active site of viral polymerases. Viral polymerases have an overall right-handed structure that includes the palm, fingers and thumb domains. These domains are further subdivided into structurally conserved motifs A–G, common to all viral polymerases. The structural motifs encapsulate the allosteric/catalytic (N2) nucleotide-binding sites. The current study investigated how nucleotide analogues explore the N2 site of viral polymerases from three genera of the family Flaviviridae using a stochastic, biophysical, Metropolis Monte Carlo-based software. The biophysical simulations showed a statistical distinction in nucleotide-binding energy and exploration between phylogenetically related viral polymerases. This distinction is clearly demonstrated by the respective analogue contacts made with conserved viral polymerase residues, the heterogeneous dynamics of structural motifs, and the orientation of the nucleotide analogues within the N2 site. Being able to simulate what occurs within viral-polymerase-binding sites can prove useful in rational drug designs against viruses.

INTRODUCTION

Nucleosides are generally hydrophilic and require membrane-bound proteins for intracellular transport. Once inside the cell, nucleosides are phosphorylated into their active metabolite as nucleotides; however, nucleotides can be dephosphorylated by 5'-nucleodases. Nucleotide analogues are incorporated in the newly synthesized product causing premature termination. These are factors that determine the efficacy of nucleoside analogues as antiviral drugs (Galmarini et al., 2001; Graci & Cameron, 2004). There are many nucleotide-bound viral polymerases in the Protein Data Bank (PDB) (Berman et al., 2000) that are static representations of interacting residues. In nature, however, enzyme–substrate interactions are not static. To visualize this, dynamic interaction may aid in drug design by identifying unreported residues that make contact with nucleotides.

Nucleoside analogue inhibitors were first used in the late 1980s as a palliative treatment for human immunodeficiency virus (HIV)-infected patients (Lin et al., 1987). Since then, rationally designing inhibitors that target viral polymerases is a major directive in drug discovery. The nucleoside analogue 2'-C-methyladenosine (2'CMA) was first used against hepatitis C virus (HCV) (Carroll et al., 2003), but Migliaccio and colleagues showed the effectiveness of 2'CMA against several phylogenetically related viral polymerases (Migliaccio et al., 2003). Migliaccio and colleagues also tested 2'O-methylcytidine (2'OMC) and 2'C-methylguanosine against HCV (Migliaccio et al., 2003). Eyer and colleagues found that the nucleoside analogue 7-deaza-2'C-methyladenosine (7-DMNA) possesses high antiviral activity against tick-borne encephalitis virus (TBEV), while maintaining low cytotoxicity for infected porcine kidney cells and human neuroblastoma cells (Eyer et al., 2015). The 7-DMNA has also been shown to attenuate HCV polymerase activity (Carroll et al., 2003). The 7-DMNA (Zmurko et al., 2016) and other 2'-modified nucleoside inhibitors (Eyer et al., 2016) have recently shown to be effective against the Zika virus. Natural,
2’-modified nucleoside substrates are a promising direction in rational design for viral polymerase inhibitors (Carroll et al., 2003), but a deeper understanding is needed on the molecular mechanisms of viral polymerase–substrate interactions.

The family Flaviviridae of viruses infects mammals, including humans and domestic animals; ticks and mosquitoes are known to transmit some members of the family Flavi-

virusidae. There are four main genera in the family Flavi-
virusidae that include Flavivirus, Hepacivirus, Pestivirus, and Pegivirus. The West Nile virus (WNV), Japanese encephali-
tis virus (JEV) and TBEV are examples of the genus Flavivi-

rus. As the name suggest, HCV is a Hepacivirus and the

immunosuppressant bovine viral diarrhea virus (BVDV) is a Pestivirus. Most of these viral species (except for TBEV) currently have resolved crystal structures of their RNA-
dependent RNA polymerases in the PDB and most (except for TBEV and JEV) were among the viral polymerases

Migliaccio and colleagues examined using 2’-CMA

(Migliaccio et al., 2003).

Members of the family Flaviviridae are all Group IV, posi-
tive-sense ssRNA viruses that use host ribosomes for the translation of, among other proteins, RNA-dependent RNA polymerase. Viral RNA-dependent RNA polymerases have a right-hand overall structure composed of the fingers, palm and thumb domains (N terminus to C terminus, respec-
tively). These domains are further subdivided into structural

motifs (A–G) (Poch et al., 1989) that possess highly con-
served amino acid residues common to all viral polymerases

(Koonin, 1991). In 1991, Koonin described five conserved

amino acid residues in all viral polymerases, namely the Lys

residue in motif F, two Asp residues in motif A, and two

Asp residues (of the GDD moiety) in motif C. The Arg resi-
due in motif F is also a key conserved residue for viral poly-

merase superfamilies I and II (Koonin, 1991).

The three binding sites encountered by a nucleotide as it enters the polymerase tunnel between the palm domain and motif F are the interrogating, priming and catalytic sites

(Bressanelli et al., 2002). The catalytic site is also referred to

as the N2 or active site (Dutartre et al., 2005). Within the

binding cavity, there is also the allosteric position known as

the initiation or N1 site (Dutartre et al., 2005). Viral poly-

merases also possess several allosteric sites for non-nucleo-
side inhibitors (Brown & Thorpe, 2015). Non-nucleoside

inhibitors maintain viral polymerases in an inactive conformation mainly by binding to the surface of the thumb domain (De Francesco & Migliaccio, 2005). However, natural 2’-modified nucleoside analogues that target the N2 site of viral polymerases are the most widely used antiviral therapy.

The advent of computational methods has granted the scientific community with a deeper insight on the all-
atom interactions occurring during enzyme–substrate binding. Using molecular dynamics simulations, Brown and Thorpe were able to assess the enhanced inhibition of HCV polymerase caused by dual allosteric binding of non-nucleotide inhibitors that showed novel structural conformations (Brown & Thorpe, 2015). These structural conformations either constrain the HCV polymerase preventing functional conformations or destabilize the polymerase preventing the transition of functional conformations (Davis et al., 2015). Here we employ a Metropolis Monte Carlo-based software to explore the all-atom interaction between the triphosphorylated (TP) nucleoside analogues 2’-CMA-TP, 2’-CMC-TP and 7-
DMA-TP within the orthosteric (N2) catalytic site of sev-
ral phylogenetically related viral polymerases. We hypothesize that phylogenetically related viral polymerases react differently to nucleotide analogues during the N2 site exploration by the diverse residue contacts formed within structural motifs, the dynamics of these motifs, and the analogue orientation within the N2 site.

RESULTS

The binding sites of phylogenetically related viral polymerases

The choice of crystal structures for the current study was guided by a reduction of extraneous variables due to experimental methods. Therefore, the crystal structures used throughout this study, of bound and unbound (apostructure) viral polymerases, were, respectively, from the same authors, namely BVDV (Choi et al., 2004), HCV (O’Farrell et al., 2003) and poliovirus (Thompson & Peersen, 2004). The tertiary TBEV-modelled structure constructed by Eyer and colleagues, which was chosen based on structural qualification algorithms (see Eyer et al., 2015), was used since there are no crystal structures available for this viral polymerase. The TBEV model was based on the JEV crystal structure (Lu & Gong, 2013) that was also included in this study. The methyltransferase domain of JEV, however, was not used in this study since we were only focusing on the nucleotide-binding sites of viral polymerases.

The phylogenetic relationship among the viral polymer-

ases from this study is shown in Fig. 1. The tree splits into three highly supported bootstrap values with poliovir-

us as an outgroup, since the poliovirus structure globally deviates from the other five viral polymerases. Poliovirus belongs to the viral family Picornaviridae and the effects of 2’-CMA have been previously documented (Migliaccio et al., 2003). The Flavivirus polymerases WNV and JEV cluster together with the more distant TBEV polymerase. The polymerase tertiary structures are also shown in Fig. 1 and are orientated 180° from the conventional right-handed positioning with clear depiction of the thumb, fingers and palm domains as outlined in HCV of Fig. 1. The structural motifs A–G (Poch et al., 1989) are colour coded for each viral polymerase depicting a similar conformation. An observable deviation, however, is motif F (red) of the Flavivirus polymerases compared with the other polymerases. The extremity is shown by WNV, as in the Dengue virus polymerase apostructure (Iglesias...
et al., 2011), with motif F orientated perpendicular to motifs A–E (i.e. the palm domain); the other polymerases are parallel to the palm domain. Due to its perpendicular orientation of motif F (Malet et al., 2007), the WNV polymerase crystal structure can act as a negative control for subsequent, biophysical analysis since the parallel orientation is closer to the bound conformation (as in the other polymerases in Fig. 1). A large section of motif F in WNV is missing (or disordered) in its crystal structure; however, this missing loop region was added using the Modeller (Eswar et al., 2006) loop-refinement application.

The structures in Fig. 1 depict the experimentally determined nucleotide positions for poliovirus, HCV, BVDV and JEV and two main sites for nucleotides within the binding cavity: the allosteric (N1) site and the orthosteric (N2) active site. The nucleotide position of the HCV–template–primer ternary complex (Appleby et al., 2015) is in close proximity (~2 Å) to the nucleotide-bound N2 (i.e. without template-primer). The nucleotide position for HCV and poliovirus is within the N2 site, closer to the finger and palm domains; for BVDV, the nucleotide is within the N1 site, closer to the thumb domain. The nucleotide position for JEV lies in between the N1 and N2 binding sites, closer to the position of BVDV. There are no resolved nucleotide-bound structures for the Flavivirus polymerases TBEV and WNV. Henceforth, we will only refer to the nucleotide-binding sites N1 and N2 according to the positions for HCV polymerase (Dutartre et al., 2005), since this is the most widely studied viral polymerase.

As a well-studied nucleotide analogue viral inhibitor, 2′-CMA-TP was initially used to run the Metropolis Monte Carlo exploration simulations with the phylogenetically related polymerases in Fig. 1 to depict how 2′-CMA-TP approaches the N2 site. The analogue starting position was ~40 Å away from the N2 site. The conserved Ser moiety within motif B was used to plot the Cartesian analogue approach, since this residue is in proximity to the N2 site and is important in mutational studies (Migliaccio et al., 2003). The scatter plots in Fig. S1 are from these simulations showing that 2′-CMA-TP approaches the N2 site of viral polymerases (<2 Å) except for BVDV. The 2′-CMA-TP exploration for BVDV approaches the N1 site – 15 Å–25 Å from the N2 site. The scatter plots show that the other viral polymerases display slightly different geometric modes of 2′-CMA-TP exploration within 10 Å of the N2 site and 10 Å of the Ser moiety (Fig. S1). For BVDV, is the N1 site more favourable for nucleotides than the N2 site, as in the other polymerases?
The nucleotide analogue 2'-CMA-TP for BVDV migrates from the N1 toward the N2 site

The structural, sequence alignment-based line graph in Fig. 2(a) shows the all-residue conformations between the apostructures and bound structures of BVDV (red), HCV (green) and poliovirus (purple). The BVDV polymerase shows the highest root mean square deviation (RMSD) for the C terminus (~4 Å) and the non-conserved N terminus (~1.5 Å). Compared with BVDV, both HCV and poliovirus showed lower RMSD (<1 Å) throughout the sequence. The BVDV also has higher RMSDs within the conserved structural motifs compared with the other two viral polymerases. These structural motifs are boxed in the line graph of Fig. 2(a) and indicated in the multiple sequence alignment in Fig. S2. The residues that have an RMSD >1 Å are Gly318 (between motifs F and A), Glu492/Cys497/Ser498 (motif E), and residues Thr510, 531–537 between motifs E and the conserved inter-helical priming loop (PL).

The JEV polymerase also has resolved apostructures and bound crystal structures. Surana and colleagues described the importance of motif F in nucleotide binding for the JEV polymerase. In the nucleotide-bound JEV polymerase, the fingertip of motif F forms a beta-hairpin structure forming contacts with motif D. The inactive JEV apostructure, however, has a disordered motif F and lacks contact with the motif D (Surana et al., 2014). We did not include JEV in the structural, sequence alignment-based line graph due to

---

**Fig. 2.** Structural conformations and BVDV-binding site exploration. (a) Line graph based on the multiple structural sequence alignment in Fig. S2(b). The y-axis is the side chain RMSD (Å) for each residue (x-axis) between the bound structures and the unbound, apostructures of BVDV (red), HCV (green) and poliovirus (purple). (b) 3D plot of the 2'-CMA-TP-binding energy (z-axis) coloured from favourable (purple–blue) to unfavourable (yellow–red). The 2'-CMA-TP exploration of BVDV started from the allosteric (N1)-binding site (x-axis; distance in Å). The distance (Å) from the orthosteric (N2) active site is also indicated (y-axis). The two line graphs in (c) depict the residue conformational shifts (left y-axis; red) and the distance of 2'-CMA-TP exploration from the orthosteric active site (right y-axis; black). The frames selected (x-axis) are within the energy-favourable ensemble for 2'-CMA-TP exploration in the 3D plot.
the extreme differences (>4.5 Å) in motif F between apostructures and bound structures.

Compared with both HCV and poliovirus polymerase, the increased residue conformational shifts of BVDV may be due to the N1 positioning of its nucleotide substrate (Fig. 1). To test this hypothesis, we started an exploration simulation with 2′-CMA-TP at the N1 position of BVDV. The 3D plot in Fig. 2(b) represents the analogue exploration of BVDV from the N1 position toward the N2, active site. The energy mapping demonstrates that the majority of the N1 exploration is unfavourable until it approaches near the N2, active site (within 1.5–5.5 Å), thus indicating a binding preference.

The two line graphs in Fig. 2(c) show the conformational shifts for the BVDV motif residues with high RMSD in Fig. 2(a). As 2′-CMA-TP approaches ~2 Å of the N2 site, the α-carbon backbone reverts closer to its apostructure state (<1 Å RMSD). This also occurs for most of the BVDV residues with high RMSD values (>1 Å) in Fig. 2(a). The C-terminus still maintains a higher RMSD (>1 Å) compared with the other residues. The C-terminal region of BVDV occludes the N2 site (Choi et al., 2004) and the 2′-CMA-TP migration toward the N2 site may cause higher conformational shifts within this region (as specifically seen in residues 531–537 of Fig. 2c).

Differential response by phylogenetically related viral polymerases to nucleotide analogue exploration in proximity to the N2 site

Due to the results in Figs S1 and 2(b), Metropolis Monte Carlo exploration simulations were initiated from the N2 site using the analogues 2′-CMA-TP, 2′-CMC-TP and 7-DMA-TP and the viral polymerases in Fig. 1. The HCV-UTP bound structure (PDB: 1NB6) has two Mn2+ ions as cofactors with several water molecules surrounding the nucleotide at the N2 site (within 5 Å). These were added to all structures since most nucleotide-bound viral polymerases in the PDB have at least one metal ion within the binding cavity and nucleotides are hydrophilic. The effects from these simulations were all highly significant (Table 1) for each polymerase, nucleotide analogue and their interactions on the dependent variables (distance from the N1 and N2 sites and the binding energy). This significance is perhaps not surprising due to the large sample sizes generated (>10 000 frames per simulation). Guided by Cohen’s rule of thumb (Cohen, 1988) for the magnitude of effect sizes (small=0.01, medium=0.059, large=0.138), the ANOVA effect sizes, represented by eta squared (η²) in Table 1, reveal large to substantially large effects for the phylogenetically related viral polymerases on all dependent variables. The effects of the interaction between viral polymerase and nucleotide analogue were small (0.015–0.03). The nucleotide analogue alone had insignificant effects (<0.01).

The relative effect sizes in Fig. 3 show the analogue exploration from the allosteric (N1) and orthosteric (N2) sites since the resolved structures of the related viral polymerases possess different geometric nucleotide positions within these two binding sites (Fig. 1). The distances from the N1 site were quite large for each analogue and all related polymerases (8–11.5 Å), thus indicating a preference toward the N2 site. For the genus Flavivirus (TBEV, JEV and WNV), the distances from the N2 site are similar to HCV for 2′-CMA-TP, but with varying binding energies. For 2′-CMC-TP and 7-DMA-TP, Flavivirus members differ in both distance from the N2 site and binding energies; however, JEV maintains

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F value</th>
<th>P value</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance from N1 site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polymerase</td>
<td>5</td>
<td>151 458</td>
<td>30 292</td>
<td>12 256.3</td>
<td>&lt;2e-16***</td>
<td>0.211</td>
</tr>
<tr>
<td>NTP*</td>
<td>2</td>
<td>3238</td>
<td>1619</td>
<td>655.0</td>
<td>&lt;2e-16***</td>
<td>0.005</td>
</tr>
<tr>
<td>Polymerase NTP*</td>
<td>10</td>
<td>10805</td>
<td>1080</td>
<td>437.2</td>
<td>&lt;2e-16***</td>
<td>0.015</td>
</tr>
<tr>
<td>Error</td>
<td>223 123</td>
<td>551 453</td>
<td>2</td>
<td></td>
<td></td>
<td>0.769</td>
</tr>
<tr>
<td>Distance from N2 site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polymerase</td>
<td>5</td>
<td>74 290</td>
<td>14 858</td>
<td>6808.6</td>
<td>&lt;2e-16***</td>
<td>0.128</td>
</tr>
<tr>
<td>NTP</td>
<td>2</td>
<td>1359</td>
<td>679</td>
<td>311.3</td>
<td>&lt;2e-16***</td>
<td>0.002</td>
</tr>
<tr>
<td>Polymerase NTP*</td>
<td>10</td>
<td>18 441</td>
<td>1844</td>
<td>845.0</td>
<td>&lt;2e-16***</td>
<td>0.032</td>
</tr>
<tr>
<td>Error</td>
<td>223 123</td>
<td>486 910</td>
<td>2</td>
<td></td>
<td></td>
<td>0.838</td>
</tr>
<tr>
<td>Binding energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polymerase</td>
<td>5</td>
<td>190 616 441</td>
<td>38 123 288</td>
<td>13 485.8</td>
<td>&lt;2e-16***</td>
<td>0.227</td>
</tr>
<tr>
<td>NTP</td>
<td>2</td>
<td>3 031 278</td>
<td>1 515 639</td>
<td>536.1</td>
<td>&lt;2e-16***</td>
<td>0.003</td>
</tr>
<tr>
<td>Polymerase NTP*</td>
<td>10</td>
<td>15 847 821</td>
<td>1 584 782</td>
<td>560.6</td>
<td>&lt;2e-16***</td>
<td>0.019</td>
</tr>
<tr>
<td>Error</td>
<td>223 123</td>
<td>630 750 333</td>
<td>2827</td>
<td></td>
<td></td>
<td>0.751</td>
</tr>
</tbody>
</table>

*NTP, Nucleoside triphosphate (i.e. the nucleotide analogues used in this study).
***Highly significant effect.
similar distances as HCV. The drastic differences are seen in 7-DMA-TP, with TBEV having the farthest distance from the N2 site, compared, respectively, with JEV and WNV. The binding energy for TBEV 7-DMA-TP, however, is similar to WNV and more favourable than JEV. In general, members of the Flavivirus genus have similar effect patterns (distance from the N2 site) to Hepacivirus, HCV polymerase than to the Pestivirus, BVDV and poliovirus polymerases (of the viral family Picornaviridae).

Variations in residue contacts and structural conformations during nucleotide analogue active site exploration

The histograms in Fig. 4 are the key conserved residues from the viral polymerase motifs (namely, A–F) that form contacts with the nucleotide analogues 2’-CMA-TP, 2’-CMC-TP and 7-DMA-TP. Motif G (Shen et al., 2012) was not included since no contacts were observed. All viral polymerases formed contacts with at least the invariant Asn
residue of motif B (Koonin, 1991), but vary in contacts with other motifs. Except for WNV, all polymerases show that the first Arg residue of motif F had the highest mean contacts with the nucleotide analogues. The perpendicular orientation of motif F to the palm domain in WNV is the reason that no contacts were made with motif F. The PL is absent in the poliovirus polymerase, but poliovirus also showed no contacts for motifs D and E. Analogue contacts were also not made with the PL and motif D of HCV.

Both HCV and poliovirus also showed differential contacts with specific conserved motif residues. For instance, each polymerase shows that the conserved Asp residue of motif A forms analogue contacts, but only with adjacent residues Thr/Gly/Tyr in poliovirus. The Lys of motif F and at least the second Asp residue of the conserved GDD moiety (motif C) formed contacts with the analogues, except for HCV. The dynamics for these motifs should differ given the various analogue contacts formed. These distinct dynamics are shown in Fig. 5. The analogue 7-DMA-TP also showed high RMSD values for the motifs of WNV, except for motif B. Motif F of WNV and the PL of TBEV showed the highest RMSD for the analogues compared with the other polymerases; however, the PL of TBEV showed lower dynamics for 7-DMA-TP. The remaining viral polymerases maintained 1.0–1.5 Å or less for each nucleotide analogue during the exploration. The polymerase BVDV showed the lowest RMSD, maintaining ~0.5 Å or less for each nucleotide analogue N2 site exploration (Fig. 5).

Motif F and the inter-helical PL are structurally heterogeneous among viral polymerases, specifically of the genus Flavivirus. The PL is disordered in its apostructure state when motif F (also disordered) is perpendicular to the palm domain (as in WNV, Fig. 1). The PL shifts when motif F is parallel to the palm domain, then resolves to a beta-hairpin when the template, NTP and the F is parallel to the palm domain, then resolves to a beta-palm domain (as in WNV, Fig. 1). The PL shifts when motif F (also disordered) is perpendicular to the palm domain (as in WNV, Fig. 1).

The orientation of nucleotide motif F to the palm domain in WNV is the reason that no contacts were made with motif F. The PL is absent in the poliovirus polymerase, but poliovirus also showed no contacts for motifs D and E. Analogue contacts were also not made with the PL and motif D of HCV.

**Orientation clusters of nucleotide analogues during active site exploration**

The orientations of the nucleotide analogues during the active site exploration were determined. The pyProCT algorithm choice was, in all cases, k-medoids, which was always able to classify all of the elements of the data (0% noise). The mean value of the Silhouette index, 0.281±0.034, and the number of clusters was only 5 of 18 times in the lower extreme (never in the higher extreme). The choice of the algorithm, the number of clusters, and the value of the Silhouette index indicate that the clustering algorithm was able to perform a quality partition of the space. The number of contacts, however, depends slightly on the analogue orientation. Figs S3 and S4 show the contacts with the proportional part of the cluster populations in different colours. Few clusters concentrate contacts, e.g. those with values >0.006 for residues and >0.015. The cluster colour legends are not shown since the ordering of the cluster is irrelevant. The analogues form defined aggregates in their exploration when the warm colours of the histograms prevail the number of clusters (Figs S3 and S4).

Closely examining the clusters using the Visual Molecular Dynamics (VMD) software (Humphrey et al., 1996) reveals that the cluster analysis was successful in capturing the translational and rotational differences during nucleotide analogue exploration. The representations in Fig. 6 show these differences within the structural motifs – orientated 90° counter-clockwise from Fig. 1. (Note: the PL is absent in the poliovirus polymerase.) The 9–10 clusters formed within the N2 site of WNV polymerase were caused by the perpendicular conformation of motif F, thereby indicating a poor binding of the analogues. The clusters depicted for WNV can, therefore, be used as controls. The other polymerases that formed 9–10 clusters within the N2 site were for 2′-CMA-TP, JEV, HCV and poliovirus polymerases, and for 7-DMA-TP, JEV. The orientation clusters for the other polymerases present asymmetries, forming ≤4 clusters depending on the nucleotide analogue; these include for 2′-CMA-TP, JEV, BVDV, HCV and poliovirus polymerases, and for 7-DMA-TP, JEV. The orientation clusters for the other polymerases present asymmetries, forming ≤4 clusters depending on the nucleotide analogue; these include for 2′-CMA-TP, JEV, BVDV, HCV and poliovirus polymerases, and for 7-DMA-TP, JEV.

The efficacy of the analogues against the viral polymerases used in this study has been demonstrated by experiments with varying potencies (cited throughout this paper). The resolved crystal structures of HCV, poliovirus, BVDV and JEV also vary in nucleotide position (Fig. 1); however, these are static representations and not the most energy-favourable positions, as seen in BVDV (Fig. 2b). We show that nucleotide analogues explore within 2.5–5.5 Å of the N2 site, depending on the viral polymerase (Fig. 3). The residue contacts made with the analogues (Fig. 4) are depicted in the PDB as a ligand diagram between nucleotide–viral polymerase within the binding cavity site. This serves as proof of principle for the biophysical simulations (see Fig. S2 for these specific residues in BVDV, HCV and poliovirus). There are, however, several other residues forming contact with the analogues not reported in the PDB static structures. These residues can act as markers for further lethal

DISCUSSION

The efficacy of the analogues against the viral polymerases used in this study has been demonstrated by experiments with varying potencies (cited throughout this paper). The resolved crystal structures of HCV, poliovirus, BVDV and JEV also vary in nucleotide position (Fig. 1); however, these are static representations and not the most energy-favourable positions, as seen in BVDV (Fig. 2b). We show that nucleotide analogues explore within 2.5–5.5 Å of the N2 site, depending on the viral polymerase (Fig. 3). The residue contacts made with the analogues (Fig. 4) are depicted in the PDB as a ligand diagram between nucleotide–viral polymerase within the binding cavity site. This serves as proof of principle for the biophysical simulations (see Fig. S2 for these specific residues in BVDV, HCV and poliovirus). There are, however, several other residues forming contact with the analogues not reported in the PDB static structures. These residues can act as markers for further lethal
Fig. 4. Mean number of contacts between the nucleotide analogue and specific residues of the viral polymerase. The number of contacts was weighted by the number of atoms and number of frames of the simulation (y-axis) to compare the different antiviral drugs and simulations. This measurement allows a comparison between the different nucleotide analogues and the simulations in the same reference frame. We have removed all residues (x-axis) with a non-significant number of contacts (~8% of the total) for the sake of readability.
mutagenesis studies. Lethal mutagenesis is a recent antiviral concept, which uses nucleoside analogues that cause mutations past a tolerable threshold to reduce viral fitness that ultimately leads to its extinction (Graci & Cameron, 2008). The authors are aware that there may be other factors affecting the residue contacts made by antiviral drugs, such as the whole ternary complex (polymerase–template–nucleotide–primer). During the N2 site exploration, there were also variations in motif dynamics (Fig. 5) and analogue orientation (Fig. 6) among the related polymerases.

**Fig. 5.** The structural motif dynamics of viral polymerases. The histograms are the mean RMSD in Å (y-axis) and structural motifs (x-axis) for each frame throughout the simulations. The RMSD has been calculated using pyRMSD, superposing all frames with the initial conformation.
The function of motif D has been linked to the fidelity of viral polymerases. If an incorrect nucleotide is incorporated during synthesis, then motif D cannot reach its optimal, active conformation (Yang et al., 2012). The Lys of motif D forms contacts with the triphosphate moiety of nucleotides (Shen et al., 2012) and determines the fidelity of viral polymerases (Yang et al., 2012). The collective dynamics between motifs A and D is also critical for viral polymerase function (Shen et al., 2012). Although the actual active conformation of viral polymerases is dependent on its ternary structure (i.e. polymerase–RNA–NTP), we show distinct contacts made by the nucleotide analogues (Fig. 4). In our study, the BVDV and Flavivirus polymerases showed contacts with the nucleotide analogues and the conserved Lys residue of motif D. Both HCV and poliovirus polymerases did not show contacts that may be due to the phosphorylation state of the analogues and/or the protonation of the conserved Lys of motif D (Yang et al., 2012). Additionally, HCV possesses an Arg, not Lys (Fig. S2), and mutation of the Arg to a Lys increases polymerase activity (Lohmann et al., 1997). Regardless, all polymerases had similar dynamics between motifs A and D, though distinct for each nucleotide analogue (Fig. 5). Further investigating the dynamics of motif D is highly regarded as a step forward in identifying preventative measures against viruses (Cameron et al., 2009).

Little is known of the conserved Lys of motif F, however, Koonin suggested it as a candidate for RNA binding (Koonin, 1991). Fig. 4 shows that the related polymerases all have this motif F Lys interacting with the analogues. Further investigations are needed for this conserved Lys residue. Motif F binds incoming nucleotides and RNA, while coordinating efforts with motifs A and C of the palm domain (Curti & Jaeger, 2013). Curti and Jaeger (2013) found that mutating the conserved Arg in motif F of BVDV resulted in an inability for the polymerase to incorporate nucleotides. This conserved Arg of motif F had the highest mean number of contacts between the nucleotide analogues and viral polymerases (Fig. 4), further indicating this conserved Arg as a therapeutic target. Relative to the palm domain (i.e. motifs A–E), the fingers domain (i.e. motif F) possesses substantial conformational changes in molecular dynamics simulations (Moustafa et al., 2011), which concur with the results in Fig. 5. During N2 exploration, all polymerases showed that motif F had the highest RMSD compared with motifs A–E.

As part of the palm domain, motif B is associated with the nucleotide-binding site for many viral polymerases (Argos, 1988) and forms the base for the template-entry channel (Bruenn, 2003). Our results show that the dynamics of motif B are lower compared with the other motifs for all polymerases (Fig. 5) that concur with its stability during molecular dynamics simulations (Moustafa et al., 2011). Several studies (e.g. Migliaccio et al., 2003) have focused on a single mutation of the conserved Ser residue in motif B that causes viral polymerase resistance to nucleoside analogues. Only the polymerases JEV, TBEV and poliovirus had analogue contact with this conserved Ser (Fig. 4). However, other residues (or at least one) in motif B also make contact with the analogues for all polymerases. How these
additional residues affect viral polymerase resistance may be worth investigating.

Both in vitro and in vivo mutational studies of GDD, in motif C, have shown its importance in enzymatic activity leading Jablonski and Morrow (1995) to conclude that enzymatic function requires the first Asp residue of the GDD, but substitution is inflexible for the second Asp. The Asp residues are also linked with coordinating efforts with residues of motif E that help positioning of the primer nucleotide (Ferrer-Orta et al., 2007). The polymerases in our study, except for HCV, form analogue contacts with at least one Asp residue of the GDD moiety (Fig. 4). Motif E is less conserved in viral polymerases but possesses a Cys-Ser moiety, and mutations of either of these residues significantly reduces de novo synthesis (Lai et al., 1999). The Ser residue of motif E for the polymerases interacts with the analogues except for the less-conserved poliovirus (Fig. 4), which may suggest some evolutionary aspect of incorporating incorrect nucleotides.

One distinct evolutionary characteristic is the missing interhelical PL in the poliovirus polymerase, which has a larger template–primer channel and is primer dependent compared with the other viral polymerases in this study. The PL acts as a stacking platform for nucleotides that ensures terminal de novo initiation during replication (te Velthuis et al., 2016). Fig. 4 shows that PL-analogue contacts were specific for the Flavivirus polymerases only. The lack of PL-analogue contacts with HCV and BVDV polymerases may be due to the PL conformational changes that occur during replication. The PL changes from a loop to a beta-hairpin as displayed in Fig. 1 when comparing the Flavivirus with HCV and BVDV polymerases. The Flavivirus polymerases also have the PL extending further into the palm domain compared with the other two polymerases. Interestingly, though, deletion of the PL in HCV polymerases has shown >10-fold increase in de novo RNA synthesis (Mosley et al., 2012).

Viruses of the family Flaviviridae infect millions of humans worldwide. The efficiency of antiviral therapy varies depending on the type of virus targeted. The herein analyses demonstrate how currently available antiviral agents interact within the active site of related viruses. We specifically identify additional residues (from those documented by crystal structures) that form contact with nucleotide analogues. These residues are within the conserved motifs of viral polymerases. Experimentations, however, are needed to explore the possibility that these residues may affect viral resistance against antiviral drugs or act as markers for lethal mutagenesis. Furthermore, understanding the molecular nuances of how antiviral agents affect various viruses will provide a better insight for future drug design.

**METHODS**

**Tertiary structures of viral polymerases.** All crystal structures were downloaded from the PDB (Berman et al., 2000). The viral polymerase crystal structures used in this study were BVDV (PDB: 1S49), HCV (PDB: 1NB6), poliovirus (PDB: 1RA7), WNV (PDB: 2HFZ), IEV (PDB: 4K6M) and TBEV. All structures were minimized to remove any steric clashes and optimized using the Schrödinger’s Maestro Protein Preparation Wizard (Li et al., 2007). Essentially, the Protein Preparation Wizard analyses the structure and clusters hydrogen bonds with the highest degree of sampling. Afterwards, the algorithm performs 100 000 Monte Carlo orientations for each cluster. Based on electrostatic and geometric scoring functions, the algorithm then determines an optimized structure. After protein preparation, an energy minimization – default conditions in the Schrödinger’s Maestro package (Schrödinger, 2010) – was performed for each crystal structure (including the modelled TBEV polymerase structure) to alter the initial conformation and to remove any steric clashes prior to the biophysical simulations.

**Multiple sequence alignment and phylogeny.** A multiple sequence alignment was performed using the MAFFT alignment server (Katoh & Toh, 2008) with the strategy E-INS-i (for sequences with multiple conserved domains and long gaps), a gap opening penalty of 1.5 and an offset value of 0.1. To choose an appropriate substitution model for maximum likelihood (ML) phylogenetic analysis, the alignment was submitted to the ProTest server (Abascal et al., 2005). We used both selection criteria implemented in the ProTest server, namely Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The substitution models selected were LG+I+G+F model for AIC and LG+G model for BIC. We then performed a phylogenetic analysis using the RAxML server (Stamatakis et al., 2008) with both substitution models. The LG+G ML tree was the only topology with the highest bootstrap values (or confidence).

**Residue conformation analysis.** The multiple sequence alignment was used to align the unbound viral polymerase apостuctures of BVDV, (PDB: 1S4F), HCV (PDB: 1NB4) and poliovirus (PDB: 1RA6) with the bound structures. The conformational change per residue was performed by importing the alignment to the MultiSeq plug-in implemented in the VMD programme (Humphrey et al., 1996). The conformational changes were plotted between the bound and apostructures using GRUNPLOT (http://grunplot.sourceforge.net).

**Biophysical simulations.** For the nucleotide analogue exploration simulations, we used the Metropolis Monte Carlo-based Protein Energy Landscape Exploration server (PELE) (Madadkar-Sobhani & Guallar, 2013). The PELE server provides ready-made scripts for unconstrained substrate-binding site search (exploration) that can be accessed at https://pele.bsc.es/. The PELE algorithm and its many applications are thoroughly explained elsewhere (Borrelli et al., 2005; Madadkar-Sobhani & Guallar, 2013). Briefly, the PELE algorithm performs in three stages. First, localized substrate perturbations are performed and protein perturbations of the α-carbon backbone are also performed using an ANM (Atigilan et al., 2001). Second, residue side chains in proximity to the substrate are optimized with steric filters and a rotamer library (Jackson et al., 2002). To reach a local minimum after the initial perturbation, the third and final stages of PELE perform a minimization using a truncated Newton minimizer within a surface-generalized Born implicit solvent (Still et al., 1990; Tamar & Aaron, 1992).

The three stages are performed for a desired number of steps running in parallel with several computer-processing units. The results are an ensemble that represents protein α-carbon backbone perturbations, side-chain conformational changes and substrate exploration. A Metropolis Monte Carlo criterion implemented in PELE either accepts a step if it is equal to and/or less than the initial energy or rejects the step if it is greater than the initial energy (Borrelli et al., 2005). A standard force field known as the optimized potentials for liquid simulations (OPLS-2005) (Jorgensen & Tirado-Rives, 1988) calculates the energy per step. The binding energy and position of the respective nucleotide from all frames (>10 000) produced by the PELE algorithm during
nucleotide–polymerase exploration were plotted using GNU PLOT (http://gnuplot.sourceforge.net).

To facilitate sampling, several parameters were altered from the ready-made PELE scripts. For the unconstrained substrate-binding site exploration, (i) short substrate translation (tra_r 0.45) and rotation (rot_r 0.15) were used to focus within the N2, active site of viral polymerases. (ii) Steric tries (steric_tr) were deleted and the number of substrate perturbations per step (tries) was lowered to 40, since substrates were at the N2 site. (iii) An ANM type (anm_altm_type) of 4, which maintains the ANM mode, was added since the default ANM mode produced favourable α-carbon backbone perturbations. The number of steps (iv) was increased to 3000 for sufficient sampling. Finally (v), the substrate constraint (iocom_con) was deleted for a freely perturbed substrate. All simulations were initiated using the a posteriori configuration of the viral polymerases.

Statistical analysis of biophysical simulations. A two-way ANOVA was employed to test for differences in the response variables of the distance (in Å) to the allosteric (N1) position, the orthosteric (N2) position and the binding energy (kcal mol\(^{-1}\)), as calculated by PELE, and for the categorical explanatory variables ‘viral polymerase’ and ‘nucleotide analogue’. With large sample sizes, most tests that assume a normal data distribution, such as ANOVA, are robust to non-normality. In extremely large sample sizes, as in our data set, the central limit theorem (CLT) validates normal-based methods (Láirá, 2009; Zuur et al., 2010). All tests and corresponding plots were carried out in the \(^{\text{R}}\) statistical package (R Core Team, 2013).

Residue contact analysis between nucleotide analogues and viral polymerases. We investigated the atom contacts between nucleotide analogue and viral polymerase by means of the Prody Python package (Bakan et al., 2011). We defined a contact as any couple of atoms (one from the analogue and one from the polymerase) that are separated by no more than 2.0 Å. All contacts were classified according to the clusters of their specific frame and the viral polymerase, and the results were weighted by the number of atoms and number of frames of the simulation so that the data could be easily compared. A Python script was written to perform these tasks that is publicly accessible at http://git.io/vEtEY.

Cluster analysis of biophysical simulations. We performed a cluster analysis of nucleotide analogue conformations using pyProCT (Gil & Guallar, 2014) (Python Protein Clustering Toolkit) to understand how these analogues explore the N2 site of viral polymerases. Using pyProCT offers two main advantages. First, it has been designed to take full advantage of parallel architectures. For instance, the huge amount of supercomputing positions needed to calculate an RMSD distance matrix is accelerated using the pyRMSD (Gil & Guallar, 2013) Python package. All clustering and analysis jobs are also distributed among all the available calculation nodes by means of a parallel scheduler. Second, pyProCT allows the user to cluster analysis techniques without the need of expertise in the cluster analysis field. Usually, the best clustering result is the one that fulfils the cluster analysis techniques without the need of expertise in the cluster analysis field. Usually, the best clustering result is the one that fulfils the user’s expectations, i.e. the results are correct and constructive. Ligand RMSD was chosen as the distance measure. When RMSD is used in parallel architectures. For instance, the huge amount of supercomputing positions needed to calculate an RMSD distance matrix is accelerated using the pyRMSD (Gil & Guallar, 2013) Python package. All clustering and analysis jobs are also distributed among all the available calculation nodes by means of a parallel scheduler. Second, pyProCT allows the user to cluster analysis techniques without the need of expertise in the cluster analysis field. Usually, the best clustering result is the one that fulfils the user’s expectations, i.e. the results are correct and constructive. Ligand RMSD was chosen as the distance measure. When RMSD is used in the present polymerase–analogue system, it captures the translational and rotational differences. As the movement of the ligand in the binding site is quite restricted, we expect the clusters to capture mainly the rotational differences.

To successfully execute pyProCT, users must define a clustering hypothesis expressing their needs in terms of their domain knowledge or preferences. Then pyProCT explores the clustering solution space by using different algorithms and sets of parameters and scores the results with the user’s hypothesis. We defined pyProCT options, the clustering hypothesis and used all available algorithms (namely, k-medoids, gromos, spectral clustering, DBCAN and a single-linkage hierarchical clustering). We expected a small number of clusters (3–15) due to limited mobility given the size of the nucleotide analogues and the site of exploration. The scoring function selected was the Silhouette index (Rousseeuw, 1987) that is defined as an internal cluster validation index that simultaneously scores intra-cluster cohesion and inter-cluster separation. Finally, we chose to discard as noise clusters that are not significantly populated, allowing a maximum of 20% noise per clustering.

ACKNOWLEDGEMENTS

J. J. V. was supported by the European Union Seventh Framework Programme (FP7/2077-2013) under grant agreement no. 316364, by the CZ.1.07/2.3.00/30.0032 project, co-financed by the European Social Fund and the state budget of the Czech Republic, and project LO1218, with financial support from the MEYS of the Czech Republic under the NPU I program. P. T. B. was supported by the Grant Agency of the Czech Republic (grant no. 14-36098G). D. R. was supported by the Czech Science Foundation (grant no. GA14-29256S) and by the Ministry of Health of the Czech Republic (grant no. 16-34238A). The authors declare no competing financial interests.

REFERENCES


Choi, K. H., Groarke, J. M., Young, D. C., Kuhn, R. J., Smith, J. L., Pevear, D. C. & Rossman, M. G. (2004). The structure of the RNA-


