ORIGINAL ARTICLE

Different receptor models show differences in ligand binding strength and location: a computational drug screening for the tick-borne encephalitis virus

Felicitas Finke[1] · Jonathan Hungerland[1] · Ilia A. Solov'yov[1,2,3] · Fabian Schuhmann[4]

Abstract

The tick-borne encephalitis virus (TBE) is a neurotrophic disease that has spread more rapidly throughout Europe and Asia in the past few years. At the same time, no cure or specific therapy is known to battle the illness apart from vaccination. To find a pharmacologically relevant drug, a computer-aided drug screening was initiated. Such a procedure probes a possible binding of a drug to the RNA Polymerase of TBE. The crystal structure of the receptor, however, includes missing and partially modeled regions, which rendered the structure incomplete and of questionable use for a thorough drug screening procedure. The quality of the receptor model was addressed by studying three putative structures created. We show that the choice of receptor models greatly influences the binding affinity of potential drug molecules and that the binding location could also be significantly impacted. We demonstrate that some drug candidates are unsuitable for one model but show decent results for another. Without any prejudice on the three employed receptor models, the study reveals the imperative need to investigate the receptor structure before drug binding is probed whether experimentally or computationally.

Keywords Molecular dynamics · Drug discovery · Receptor models · AlphaFold

Introduction

The tick-borne encephalitis (TBE) virus is a neurotropical disease with flu-like symptoms that may evolve to a severe neurological illness. The virus can be subdivided into three main types, namely the European TBE (TBE-Eu), the Siberian TBE (TBE-Sib), and the far eastern variant (TBE-FE) [1], that differ by their death toll. Primarily, TBE-FE is associated with a severe course of the illness with long-lasting neurological effects [2] and a death toll of 20–60% of the infected. The other variants of the virus, TBE-Eu and TBE-Sib, exhibit a death rate of 1–2% and 6–8%, respectively [3]. However, the virus subtype alone is not the only criterion for a mild or severe case of the disease, as for each subtype any scenario is possible, and the implications are mainly influenced by the age or genetic predisposition of the patient [1].

The different TBE virus subclasses can be found in Europe and Asia and are classified as an epidemic in numerous countries. [2, 4, 5]. The disease associated with the TBE virus is notifiable in the European Union since 2012 [1]. According to the Annual Epidemiological Report of the ECDC (European Centre for Disease Prevention and Control), 24 countries within the European Economic

✉ Ilia A. Solov'yov
  ilia.solovyov@uni-oldenburg.de

✉ Fabian Schuhmann
  fabian.schuhmann@nbi.ku.dk

  Felicitas Finke
  felicitas.finke@uni-oldenburg.de

  Jonathan Hungerland
  jonathan.hungerland@uni-oldenburg.de

1   Institute of Physics, Carl von Ossietzky Universität,
    Carl-von-Ossietzky-Str. 9-11, 26129 Oldenburg, Germany

2   Research Centre for Neurosensory Science, Carl von
    Ossietzky Universität, Carl-von-Ossietzky-Str. 9-11,
    26129 Oldenburg, Germany

3   Center for Nanoscale Dynamics (CENAD), Carl von
    Ossietzky Universität, Ammerländer Heerstr. 114-118,
    26129 Oldenburg, Germany

4   Niels Bohr International Academy, Niels Bohr
    Institute, University of Copenhagen, Blegdamsvej 17,
    2100 Copenhagen, Denmark

Area reported a total of 3,734 cases of TBE in 2020 and an increase in the more recent years [2]. For instance, the reported cases increased by 30% in Germany in 2023 compared to 2022 [6].

The disease is mainly transmitted through tick bites [7]. In 1% of the cases, it is transmitted through consumption of products stemming from infected animals, like unpasteurized milk from cows, sheep, or goats. A case of a direct infection transmitted from human to human is not known [8]; however, a case has been reported in which the virus was transmitted through the transplantation of infected organs [9].

Even though the TBE virus is widely spread in Europe and Asia and studies have been performed to inhibit the similar Japanese encephalitis virus or the Zika virus [10–13], no medication is known to treat it, which reduces the options for supporting or symptomatic treatments, while effective vaccinations are available [1].

With rising global temperatures, the odds of ticks surviving winter increases, thereby boosting the chances of infection. The rising numbers highlight the need for specific medications that can be employed post-infection. Fortunately, nowadays suggestions for potentially pharmaceutically relevant drug candidates can be established computationally to screen a large number of potential drugs and find well-binding candidates prior to the individual experimental tests [14, 15]. The present investigation focuses on such a virtual drug screening procedure for the RNA-dependent polymerase (RdRp) protein of the TBE virus with the aim to find putative medicinally relevant drug ligands. A collection of 2,000 ligand molecules were investigated, similar to the frequently studied drug molecule that has the potency to inhibit the RdRp protein, namely the 7-Deaza-2'C-methyladenosine (7DMA); the 7DMA molecule has been documented to effect some viruses from the flavivirus genus [7].

The TBE virus contains an RdRp protein, which is essential for the replication of the virus inside the host system. The RdRp protein is highly conserved through various viruses of the flavivirus genus, which also includes the TBE virus. Such conservation might indicate the specific importance and sensitivity of the protein and thus reduces the chance that the virus can successfully undergo a sustainable mutation to gain immunity against the putative medications. Furthermore, it was shown in mice that the protein inhibits the growth of the neurites, which directly favors the onset of the neurological phase of the disease [16]. Lastly, the protein does not have a eucaryotic homologous counterpart. Therefore, a successful drug might have little to no secondary effects on human cells [7].

The RdRp protein structure is only partially available in its crystalline form with 98 out of 638 (partially) missing residues. The present study attempts to reconstruct the missing parts of the RdRp receptor through computational modeling. Three structural models were established and their differences and similarities rationalized. The first model employs the homology modeling approach relying on the SWISS-MODEL platform [17–21]. The second model employs AlphaFold [14, 22, 23] to reconstruct the structure of the entire receptor. The third model uses AlphaFold to replace only the partially modeled or missing residues, while keeping the rest of the crystal structure unchanged. The multiple models were considered to address the low-confidence regions in the receptor structure, such that the structure obtained using SWISS-MODEL could be compared to the two other models from the same virus genus. As computer-generated protein structures have warring confidence levels in different regions, it is natural that no structure can be proven to be the only correct one.

The 2,000 potential drug ligands were then tested for all three models individually using the Automated Ligand Searcher (ALISE) [26], which is incorporated into the browser-based computational platform VIKING (Scandinavian online kit for nanoscale modeling) [27].

The most-promising ligands for all three models were compared in their binding energy and binding location. We have found that the atomistic details in the receptor structure severely influence the potential drug binding affinity and that it is unpractical to unquestioningly rely on a receptor structure obtained using any modeling tool. Three ligands were identified as the best-performing ligands in two of the three receptor models, making them the most robust independent of the receptor. While the results should be considered with care, an experimental study of these ligands might significantly further the cause in finding a medically relevant drug against the TBE virus.

Background and methods

The receptor structure

A potentially viable receptor for drug targeting and the inhibition of the TBE virus is its RNA-dependent RNA-polymerase (RdRp), which is essential for the self-replication of the virus. The RdRp protein is at the C-terminal of the so-called NS5 non-structural protein inside the virus, which consists of the methyltransferase domain followed by the polymerase domain [7, 16].

Figure 1 shows the major motifs in the structure of the RdRp protein of the TBE virus and its so-called right-hand architecture, which forms a catalytic unit. The structure is subdivided into three regions, surrounding the active site of the RdRp receptor. The thumb domain is shown in gray-blue, the finger domain in blue, and the palm domain in teal. The priming loop (red) is part of the thumb domain. Overall, the right-hand architecture can be further classified

Fig. 1 A Right-hand-architecture of the RdRp Protein of the TBE virus. The structure is subdivided into three regions surrounding the active site in its center. The so-called thumb domain is shown in gray-blue, the finger domain in blue, and the palm domain in teal colored. The priming loop (red) is part of the thumb domain. **B** Colored representation of the different motifs in the RdRp protein structure. Motifs A, B, C, and F interact directly with the so-called NTP. The other motifs are endowed with structural responsibilities

○ Thumb ● Priming Loop
● Fingers
● Palm

○ Motif A ○ Motif B ● Motif C
○ Motif D ● Motif E ● Motif F
● Motif G
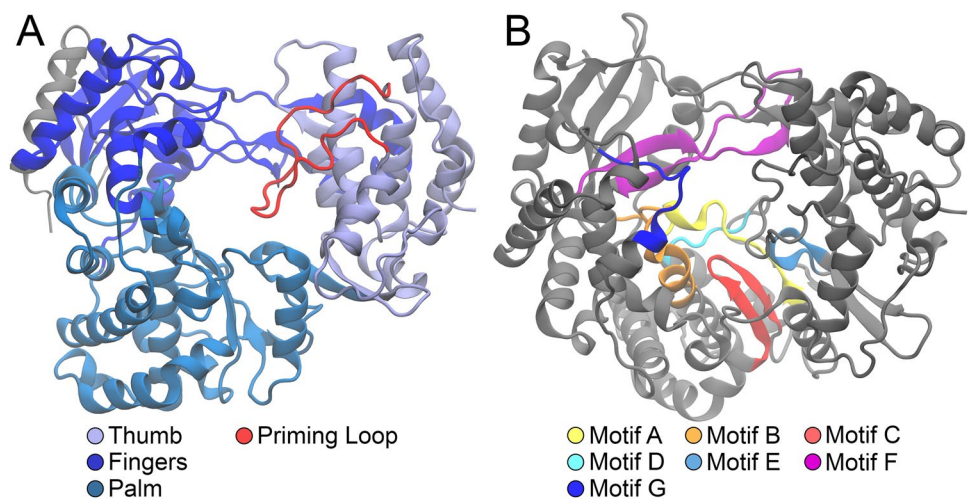
through the abundance of functional motifs, denoted as A–G in Fig. 1B. The motifs A, B, C, and F interact directly with the nucleoside triphosphate (NTP), which is the precursor for RNA synthesis.

The potential drug binding site of the RdRp protein is located in its active center, where the RNA transcription takes place. The priming loop limits the accessibility to the active site, so only ssRNA may enter the center during the initialization of the transcription [28, 29]. A well-binding drug ligand might, therefore, prohibit initialization of the RNA synthesis.

A crystal structure for the RdRp protein with the PDB ID 7D6N [28] is available in the PDB database [30]. The protein sequence comprises 638 amino acid residues, of which, in total, 19 residues at the N- and C-terminal have not been reconstructed. Furthermore, the crystal structure does not contain the well-defined positions of 50 amino acid residues and exhibits unmodeled side chains for 48 amino acid residues (see Table A1 in the Supplementary Material (SM)). As the partially modeled or missing residues are expected to be close to the active center, modeling methods were employed to obtain a more complete receptor structure.

Three different modeling methods were employed to generate the receptor structure. The first method uses the SWISS-MODEL [31] homology modeling, and the second utilizes the artificial intelligence workflow provided by AlphaFold [14, 22, 23]. Lastly, a hybrid model was considered, where the structural information obtained from AlphaFold was merged with the crystal structure, effectively replacing the missing and partially modeled regions.

The receptor obtained from SWISS-MODEL [17–21] requires the RdRp protein's sequence and is denoted as Swiss. A visual inspection of the resulting receptor model showed some differences in crystal structures of similar viruses: the Japanese encephalitis virus (PDB ID 4K6M) [10] or the Zika virus (PDB ID 5TFR) [11], which belong

to the same virus genus as the TBE virus. The Japanese encephalitis virus structure and the Zika virus exhibit less unmodeled residues and have been subjected to drug docking procedures in the past [12, 13].

A visual representation of the differences is shown in Fig. 2. With this in mind, different receptor models were employed to compare the drug ligand binding results. No judgement is made, however, on the realism of any of the resulting receptor models.

The second receptor was generated using AlphaFold [14, 22, 23]. AlphaFold only requires the amino acid sequence to generate a three-dimensional protein structure, but the generated structure can be refined by selected homologous crystal structures. In this case, the algorithm used four crystal structures for refinement (pdb IDs: 4C11 [24], 5U0C [25], 4K6M [10], 5U0B [25]). AlphaFold is an artificial intelligence tool trained on the existing crystal structures, which are, for instance, located in the PDB database. AlphaFold also generates the N- and C- terminal residues, which were completely unmodeled in the crystal structure. To match all studied models, these 19 amino acid residues were subsequently removed. Alpha denotes the receptor model obtained from AlphaFold.

Finally, because AlphaFold relies on the predictions from all available crystal structures, and not necessarily on the ones matching the RdRp protein of the TBE virus, a hybrid model was constructed, denoted by Hybrid. The model was obtained by aligning the crystal structure and Alpha model. The partially modeled or missing amino acid residues in the crystal structure were then inserted into the crystal structure from the Alpha model. The hybrid model, therefore, directly contains information from the crystal structure, supplemented by the information obtained with AlphaFold.

In the last step, the protonation states of the different amino acid residues had to be taken into account as, for instance, a wrongly protonated histidine residue can have

Fig. 2 A The structure obtained from the Swiss model is colored according to the certainty estimated by the modeling process (arb. units). The certainty is measured on a scale from 0 to 100, with 0 describing an uncertain region and 100 denoting a high confidence. **B** The uncertain region obtained from the Swiss model is highlighted in red, while the same region is overlaid with the structure of the Japanese encephalitis virus (yellow) and the Zika virus (cyan)
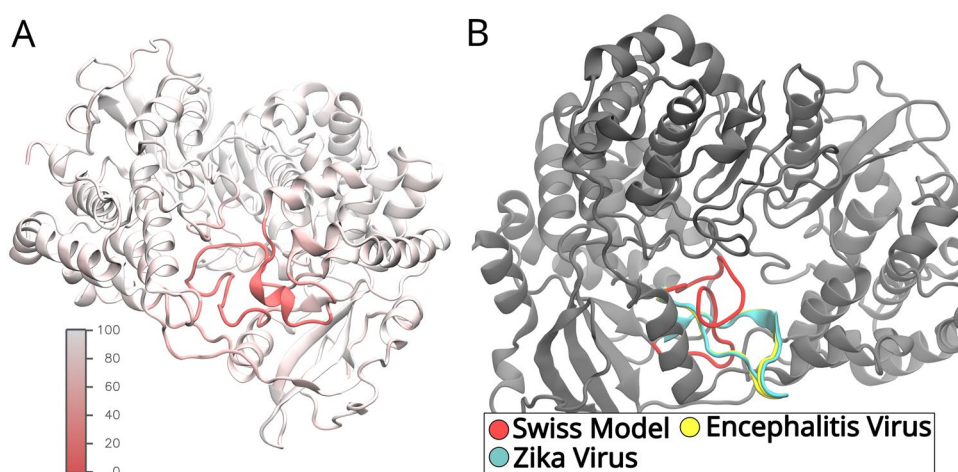
significant effects on the binding affinity of molecules [32]. Considering a physiological pH-value of 7.4 [33], the so-called pK_a values can be calculated, which was done employing the PROKKA3 program [34, 35] for the Swiss model. The resulting pK_a values, see Table B1 in the Supplementary Materials, combined with a visual inspection, led to changes in the protonation of histidines. By default, all histidines were considered to be δ-protonated, with only the His180 being ε-protonated. Furthermore, the Asp402 residue was additionally protonated. To keep the comparability of receptor models consistent, the protonation states from the Swiss model were also used for the Alpha and Hybrid models.

After the structures were modeled and the protonation states were set, a molecular dynamics (MD) simulation was conducted in multiple steps to equilibrate the structures. For the MD simulation, the VIKING platform was employed [27], which relies on the simulation software NAMD [36, 37] and VMD [38] for file preparation. All equilibration stages employed the CHARMM force field with CMAP corrections [39–41] with a temperature of 310 K. The systems were neutralized, and sodium chloride ions were added to achieve a salt concentration of 0.15 mol/L. In the first two stages, the MD simulation was conducted in an NPT (constant number of atoms, pressure, and temperature) ensemble with a timestep of 1 fs and a pressure of 1.01325 bar. Everything but the solvent was restrained in the first stage, while in the second stage, the amino acid residue's side chains were allowed to move. The first stage was initiated with 1000 conjugate gradient minimization steps. Stages three and four were conducted in the NVT (constant number of atoms, volume, and temperature) ensemble. In the third stage, no restraints on atomic positions were imposed. Finally, in stage 4, the integrator time step was set to 2 fs, and the covalent bond to hydrogen atoms were considered rigid. The receptor protein structures were simulated for 30 ns in stage 4. For the Hybrid equilibration, a second minimization step was introduced to account for the artificially over-long bonds generated through the manual combination of Alpha and the crystal structures.

The final snapshots from the equilibration simulation were further used for the drug screening workflow.

Automated ligand searcher

All three receptor models of the RdRp protein were used in the ALISE program [26] to predict possible drug-like ligands that could bind inside the receptor. ALISE is an automated workflow incorporated into the VIKING online platform [27], which sorts ligands based on their binding affinity to a provided receptor model. A search space to limit the docking to physiologically sensible binding sites can also be defined which was chosen to be the active center in the middle of the receptor structures. Different ligand molecules were generated by ALISE that was provided with the SMILES string of 7DMA, a known drug for viruses similar to the TBE virus. The search space for the drug docking was chosen to contain the RNA and the NTP binding location to block the most important functions of the receptor. With this information, ALISE automatically found 2000 similar ligands according to the Tanimoto score in the PubChem database [26]. The Tanimoto score K is given for two molecular fingerprints by [42]

$$K = \frac{|A \cap B|}{|A \cup B|}. \tag{1}$$

ALISE relies on a three-step process in which, gradually, potential ligands are filtered out, while the computational accuracy of the binding prediction increases. Drug docking is performed in the first step relying on the docking program Vina (AutoDock Vina 1.1.2)[43] , with ligands being prepared by the AutoDockTools [44] and Openbabel [45] programs. The docked ligands were then sorted according to their binding scores. The highest-ranked 30 ligands of the

docking phase were then considered in the second step of ALISE relying on atomistic MD simulations.

The MD simulation phase includes simulations of the receptor, the ligand, and a complex consisting of the receptor and ligand in an implicit solvent [26, 46]. While the docking procedure in the first step is static, the MD simulations allow motions in the receptor and ligand. These motions, for instance, might lead to a repulsion and force ligands to leave the binding pocket, which is a clear indication of a low binding affinity. The binding energy of the ligands was estimated from the three different simulations per ligand (receptor only, ligand only, and complex simulations), and the ligands were sorted by their binding affinity. ALISE employs the molecular mechanics-generalized Born and surface area continuum solvation (MM/GBSA) method [46]. ALISE's MD step employs NAMD [36, 37], while the simulation files were prepared using VMD [38]. The Charmm General Force Field (cgenFF) [47] was employed for the ligands, while the receptor was modeled using the Charmm Force Field with CMAP corrections [40, 48–54].

The five highest-scoring ligands from the MD step were subjected to a free energy perturbation (FEP) stage. In the FEP stage, windowed MD simulations were conducted in which the interactions between the receptor and the ligand were gradually decoupled and recoupled. Conformational restraints are used and later subtracted to avoid a diffusion of the ligand when the interactions are completely decoupled. Such an approach allows the most precise estimate for the binding free energy and establishes the final ranking of the remaining ligands [26]. A total of 50 simulation windows were employed, each including an MD equilibration. The FEP step employs NAMD [36, 37] with an explicit solvent and the same force fields and parameters as used in the MD step described above.

Results

Stability

Following the initial equilibration MD simulations, the stabilities of the structures were validated by employing the Root Mean Square Deviation (RMSD) analysis. The RMSD values for the last snapshot of the equilibration simulations are compiled in Table 1 to compare the different structures with each other. The RMSD was additionally calculated for the regions with high and low certainties, following the structure prediction based on the Swiss model. A visual representation of the confidence is shown in Fig. 2A. The resulting values suggest that the three receptor structures are relatively similar with each other in the high-certainty regions, while Swiss and Alpha models differ significantly in the uncertainty regions. Additionally, the RMSD time evolution

Table 1 The root mean square deviation (RMSD) between the final configurations of the different receptor models

Reference simulation	Swiss compared to hybrid (Å)	Swiss compared to Alpha (Å)	Hybrid compared to Alpha (Å)
Whole protein	3.93	4.24	3.66
Low certainty	12.19	12.78	4.81
High certainty	3.04	3.35	3.60

The RMSD is calculated for the whole protein, high-certainty, and low-certainty regions as estimated by the SWISS model platform [17] during receptor generation. In this table, the high-certainty regions have a confidence score above 50; low certainty are the regions below 50 as seen in Fig. 2A. All receptor models are similar in the high-certainty regions, while significant differences are found primarily between Swiss and Alpha models in the low-certainty regions

for the entire protein is shown in Fig. 3. The overall low RMSD below 4 Å for the protein consisting of 619 residues indicates fairly stable structures [55–59]. Additionally, even though the magnitude of the fluctuations is higher in the Hybrid and Swiss model compared to the Alpha model, they seem to evolve around a flat plateau after 10 ns or 20 ns, respectively.

Drug screening

Following the stability analysis of the different receptor models, the virtual screening program ALISE was used to study the binding affinity of a variety of ligands. The three steps of ALISE were conducted automatically, starting with the drug docking stage. All stages were completed for the three studied receptor models. After each step, ALISE ranks the ligands; the best-performing ligands are then advanced to the next stage:

The ranking for the best 50 docked ligands in their five best binding positions is provided in Tables C2–C4 in the Supplementary Material. The 30 best-performing ligands were considered for the MD stage. During the MD stage, all atoms in the system are free to move and exhibit attraction and repulsion interactions. These interactions might result in the ligand to leave the originally determined binding pocket. Therefore, additionally to a more rigorous estimate of the binding energy, ligands that propelled out of their binding pocket can be excluded from further testing. The different locations of the potential ligands obtained from the docking stage provide different starting configurations for the MD simulations.

The binding free energy estimated from the MD simulations for all receptors and all ligands is summarized in the Supplementary Material (see Tables D5–D7 in Supplementary Material). A visual inspection of the bound ligands in the different receptors reveals three noteworthy results.

Fig. 3 The three panels show the time evolution of the RMSD for the whole protein during the 30 ns equilibration phase. For each panel, structures were aligned to the respective reference structures prior to RMSD calculation, i.e., in the first panel, all three simulation trajectories are aligned to the first snapshot of the Alpha model simulation. The dependencies show an overall stable behavior of the different receptor models with most fluctuations in the Hybrid model. Throughout the simulation, one also notices that the Swiss model differs more from the Alpha model than the Hybrid model and that the Hybrid model and the Alpha model are similarly dissimilar to the Swiss model, concurring with the expectations
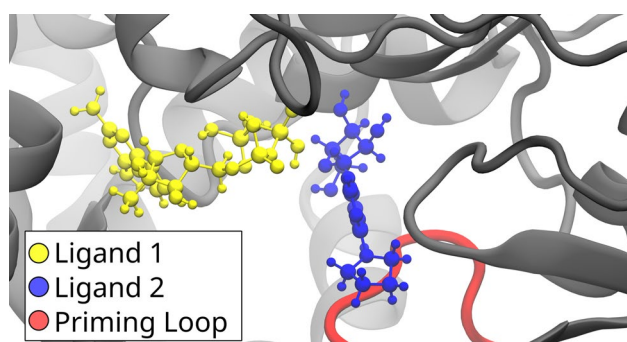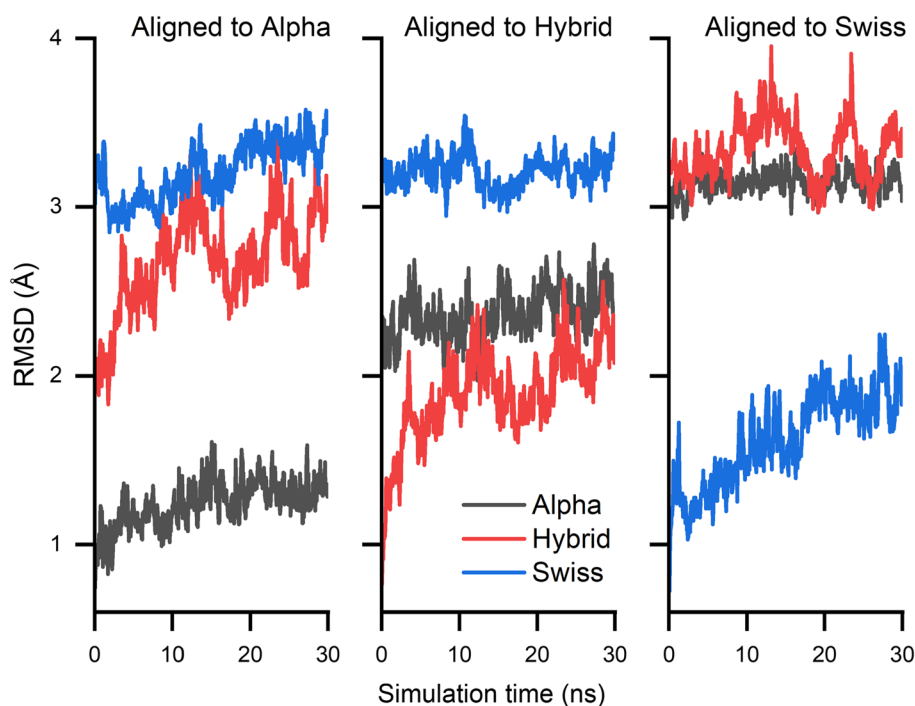
Fig. 4 Two ligands are shown within the Hybrid model receptor, with the priming loop in red. Ligand 1 (yellow) is the only ligand that obtained a negative binding energy in the MD stage, while the other 10 best-performing ligands, represented by the exemplary ligand 2 (blue), perform more poorly

In the course of completed MD simulations, most ligands were able to stay in their originally placed binding locations as determined by the docking stage. The original binding locations may differ, however, even within a single receptor model. Figure 4 shows the positioning of the second-ranked ligand (yellow) compared to an exemplary ligand (blue), which represents the orientations of the majority of the 10 best-performing ligands bound in the Hybrid model. During the MD stage, it is revealed that only the yellow ligand results in a favorable binding energy. Thus, the ALISE's MD step corrects unphysical placements of ligands during the docking stage.

Another example is provided by the ligand with the PubChem ID 189237 which was ranked among the best 30 performing ligands in the docking stage for all three receptors; however, its placement inside the receptors turned out to be different. The ligand was thus subjected to MD simulations three times, once for each receptor model. Figure 5 shows the ligand's positioning in the different receptor models. The Swiss and Hybrid model are consistent with each other in respect to the ligand binding location, but not orientation. The Alpha model features the ligand on the opposite side of the priming loop.

Another result followed from the visual inspection of conformations of the final snapshot of the different MD simulations. Here, the Trp533 residue in the Swiss and Alpha model receptor was studied, which turned out to assume two different distinct orientations as visualized in Fig. 6. It turns out that the side chain orientation, as seen in the Alpha model, actively perturbs the potential drug binding location, as has been identified, for example, for the ligand with the PubChem ID 189237, seen in Fig. 6B. Figure 6A shows the orientation of the tryptophan in the Swiss model with the bound ligand. The slight change in side chain orientation might, therefore, significantly affect the binding site.

Five ligands with the best binding energy estimates from the MD simulations were subjected to the FEP calculations. Remarkably, not a single ligand had simultaneously a favorable (negative) binding free energy for all the three receptor models. However, three ligands still showed binding affinity for two receptor structures. The chemical structure of these three ligands is illustrated in Fig. 7. The
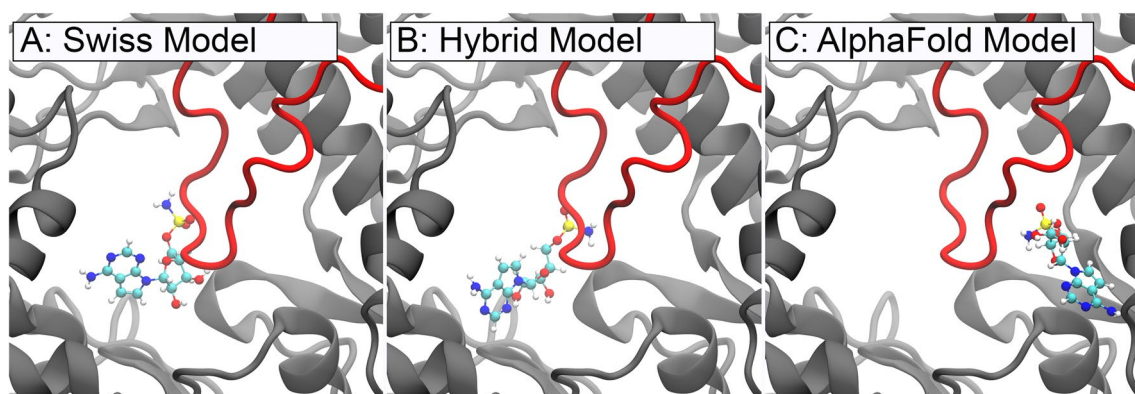
Fig. 5 Binding locations for the ligand with the PubChem ID 189237 within the different receptor models. In the Alpha model, the same ligand exhibits a different binding site on the other side of the priming loop compared to the Swiss and Hybrid Models

Fig. 6 Orientation of the tryptophan 533 residue in the Swiss model (**A**) and the Alpha model (**B**). The panels also show the placement of the ligand with the PubChem ID 189237 in the Swiss model. The difference in orientation of the tryptophan side chain perturbs the potential binding site of the ligand, making it impossible to bind to that particular location. The slight change in side chain orientation might, therefore, significantly affect the ligand binding site
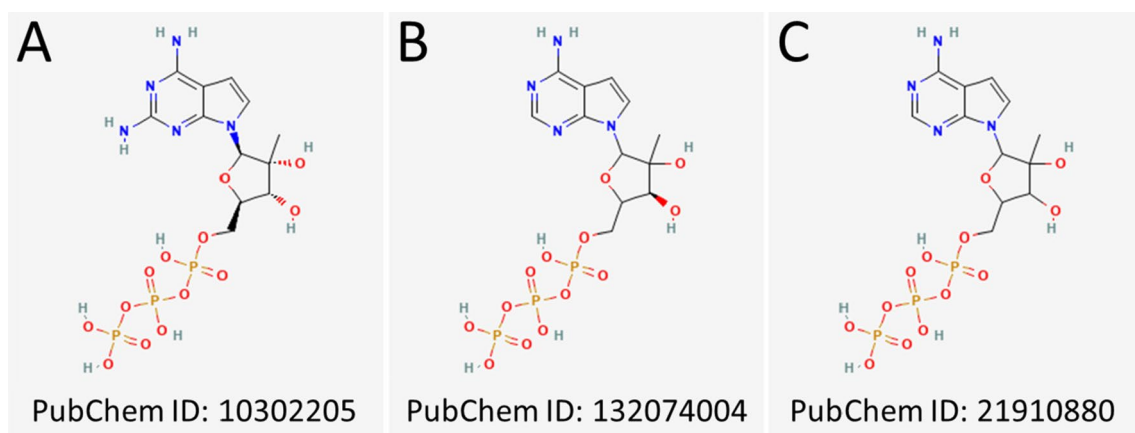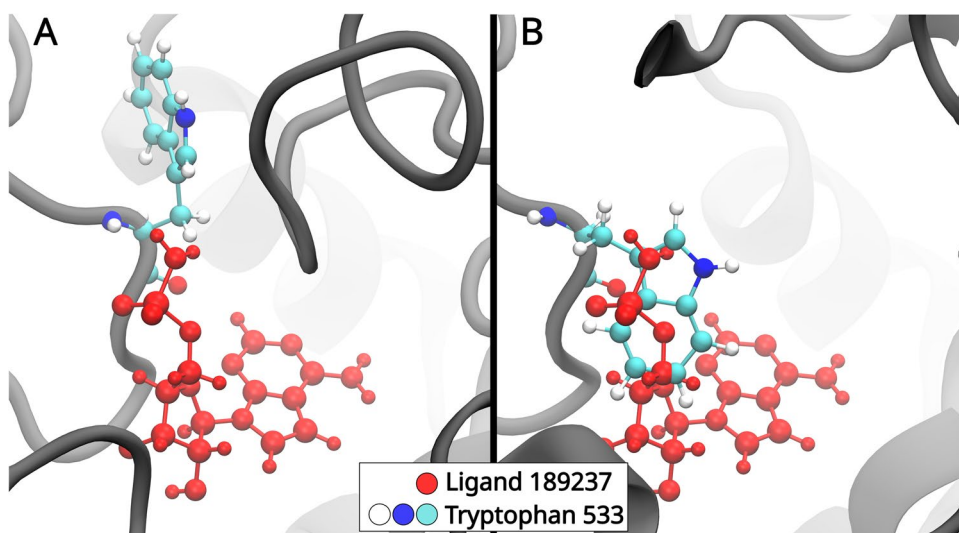
Fig. 7 Structural formulas for the three most robust ligands that show favorable binding across the different receptor models. The corresponding PubChem IDs are indicated. The ligands in **B** and **C** are structurally identical; however, the ligand in (**B**) is an enantiomer of the ligand in (**C**)

binding free energies estimated by the FEP stage are given in Table 2. The ligands occurring in two receptor models are colored according to their pair. Complete tables containing individual contributions can be found in the Supplementary Material (see Tables E8–E10 in Supplementary Material). To validate the binding locations of the three reoccurring ligands, additional MD simulations were performed for the highlighted receptor-ligand complexes in three replicas each. Based on an RMSD analysis, the ligands maintain their location in the receptor structure for at least another 100 ns. A time evolution of the RMSD for the validation simulations is found in Fig. F11.

Comparison of high-ranking ligands

Differences in the starting configuration for the FEP calculations carry over from the MD simulations. As such, the binding location may differ within the three ligands. For instance, the ligand with the PubChem ID 10302205 (rank four in the Swiss model, rank two in the Hybrid model) exhibits a slightly different orientation in the Swiss and Hybrid models, respectively; a visual inspection shows that the ligand might prohibit the ssRNA from entering the

structure of the receptor in the Swiss model. At the same time, the ligand's location might indicate an inhibition of the NTP interaction in the Hybrid model. As NTP is a precursor to synthesizing RNA, blocking its interaction site makes the ligand a likely inhibitor for the virus. More precisely, in the Swiss model, the ligand is exposed to a large surface area of the motifs B and F and parts of the motifs C and E of the receptor, see Fig. 1B. In the Hybrid model, the ligand is close to the A, B, C, and F motifs. Figure 8 shows a visualization of the drug binding location and orientation at the beginning of the FEP calculations. The A, B, C, and F motifs interact directly with the NTP. Therefore, one can suspect that the ligand may block the binding site for the NTP and significantly influence the functionality of the RdRp. Furthermore, the ligand binds fairly robustly with a binding free energy of negative 13–16 kcal/mol to both receptor models (see Table 2).

The ligand with the Pubchem ID 132074004 is ranked first in the Hybrid model and fourth in the Alpha model. The ligand has a similar position in both receptor models, but a different orientation. While the triphosphate region is oriented in different directions, the ligand's aromatic rings share a similar location. The ligand is close to the A, B, C,

Fig. 8 Positioning of the ligand with the PubChem ID 10302205 inside the Swiss model (**A**) and the Hybrid model (**B**). In the Swiss model, the ligand is located in the center of the active site, potentially interacting with the B, C, E, and F motifs. In the Hybrid model, the ligand is close to the A, B, C, and F motifs
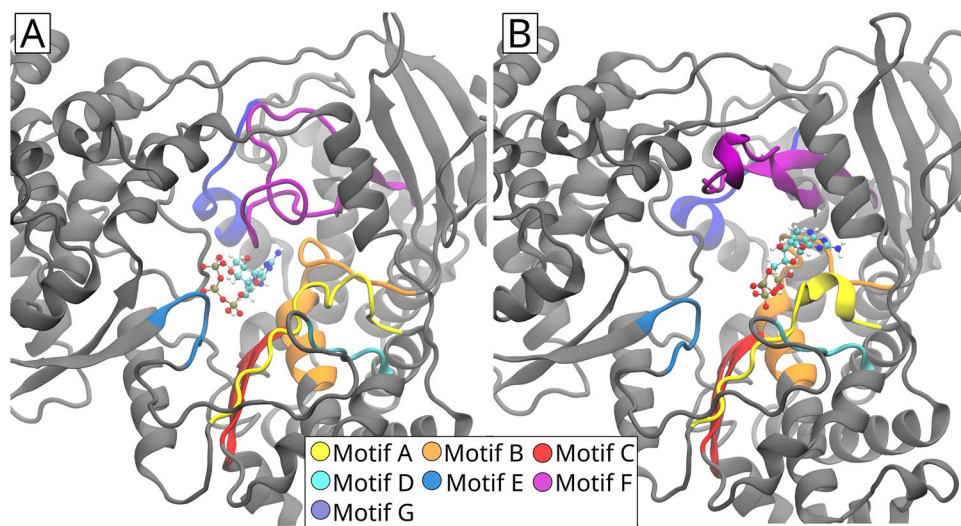
Table 2 The binding free energies obtained from the FEP stage in the ALISE workflow are shown for the different receptor models

Hybrid		Swiss		Alpha	
PubChem ID	ΔG_0 (kcal/mol)	PubChem ID	ΔG_0 (kcal/mol)	PubChem ID	ΔG_0 (kcal/mol)
132074004	−24.27	21910880	−23.65	275781	−24.86
10302205	−15.72	163033685	−21.02	189237	−19.06
154307377	−3.63	88903284	−16.14	21910890	−11.83
168270532	−2.42	10302205	−13.78	132074004	−6.66
21910880	−0.39	77983075	−4.48	89442359	−3.50

The ligands are indicated by their PubChem ID. The highlighted ligands ranked among the best-performing five ligands in the MD stage for two different receptor models. A detailed account of binding free energy estimates is given in the Supplementary Material (Tables E8–E10 in Supplementary Material)

and F motifs, similar to the binding site exhibited in Fig. 8B. Hence, the ligand might also inhibit NTP interactions.

Lastly, the ligand with the PubChem ID 21910880, which ranks highest in the Swiss model, was the least performing ligand in the Hybrid model with a binding free energy estimate of − 0.39 kcal/mol. The binding energy value close to zero indicates a weak binding and small perturbations might already lead to the drug molecule's detachment. Additionally, the ligand is located in the poor confidence region of the receptor with the greatest structural variability. The ligand is, however, an enantiomer on the base of the previous ligand with the Pubchem ID 132074004; in other words, they are configurationally isomeric.

Considering that the ligands 132074004 and 21910880 are structurally similar, the racemat (21910880) is within the best five performing ligands for all three receptor models, which makes it the most robust ligand that potentially binds to all three versions of the receptor structure.

Conclusion

The choice of receptor structure is essential for the quality of a computational ligand screening experiment. We have demonstrated that differences in the secondary structure within receptor models of the RNA-depended RNA-polymerase of the tick-borne encephalitis virus and minor differences like the orientation of a single side chain may drastically affect the binding of drug ligands. The structural differences directly influence the size of binding pockets, the orientation and location of bound ligands, and, thus, the effect a ligand might have on the protein's functionality.

We have computationally constructed three different receptor structures with different modeling tools and compared their susceptibility to 2000 potential drug molecules. From the docking stage, the 50 best-performing ligands advanced to the molecular dynamics stage. While 17 of these best-performing ligands could be found across all three different receptor models, no ligand was among the best-performing ligands for all the three investigated structures after the MD stage. The results indicate that without endowing any level of realism on neither the Swiss Model-generated structure, nor the AlphaFold-generated structure, nor their hybrid variant, special care should be taken as a single model could lead to inconsistent results and should not be taken as the all-truth without further study. Additionally, all ligands could be compared regarding their physiochemical properties or their closest residues in the different receptor structures after each stage in ALISE's virtual screening workflow, which would go beyond a study focusing on important structural differences to drug binding in receptor structures.

As a silver lining, however, the ligand with the PubChem ID 21910880 turned out to be structurally identical to the ligand with the PubChem ID 132074004, except for a mirrored oxygen. If one considers these two highly similar compounds, a single potential drug molecule is identified which is robust over all three different receptor models and binds in all of them. Thus, it might be feasible to test the ligand with the PubChem ID 132074004 experimentally, as its performance is close to independent of the structure model. The ligand with the PubChem ID 10302205 performs robustly in two of the three models. In summary, without trusting the physiological realism of the receptor structures, the ligands 10302205 and 21910880 might be feasible candidates for further testing as potential treatments of the TBE.

Overall, trust in a receptor structure is hard earned. We demonstrated the effects of differences in receptors on the binding of drug molecules. Therefore, the results of virtual screening can only be trusted as far as the receptor structure model.

Code availability All used programs are cited. Some programs contain proprietary elements, which are not free for non-academic users.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Ruzek D, Avšič Županc T, Borde J et al (2019) Tick-borne encephalitis in Europe and Russia: review of pathogenesis, clinical features, therapy, and vaccines. Antiviral Res 164:23–51. https://doi.org/10.1016/j.antiviral.2019.01.014

2. Surveillance report. tick-borne encephalitis—annual epidemiological report for 2020. https://www.ecdc.europa.eu/sites/default/files/documents/Tick-borne-encephalitis-annual-epidemiological-report-2022.pdf. Accessed 27 July 2023

3. Gritsun T, Lashkevich V, Gould E (2003) Tick-borne encephalitis. Antiviral Res 57(1):129–146. https://doi.org/10.1016/S0166-3542(02)00206-1

4. Centers for Disease Control and Prevention (2023) Tick-borne encephalitis (TBE)—geographic distribution. https://www.cdc.gov/tick-borne-encephalitis/geographic-distribution/index.html. Accessed 27 July 2023

5. Im JH, Baek JH, Durey A, Kwon HY, Chung MH, Lee JS (2020) Geographic distribution of tick-borne encephalitis virus complex. J Vector Borne Dis 57:14–22. https://doi.org/10.4103/0972-9062.308794

6. Seedat J, Winkler M (2023) Aktuelle daten und informationen zu infektionskrankheiten und public health—epidemiologisches bulletin. https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2023/Ausgaben/09_23.pdf?__blob=publicationFile. Accessed 27 July 2023

7. Boldescu V, Behnam MAM, Vasilakis N, Klein CD (2017) Broad-spectrum agents for flaviviral infections: dengue, Zika and beyond. Nat Rev Drug Discov 16:565–586. https://doi.org/10.1038/nrd.2017.33

8. World Health Organization (2023) Tick-borne encephalitis. https://www.who.int/health-topics/tick-borne-encephalitis#tab=tab_1. Accessed 27 July 2023

9. Lipowski D, Popiel M, Perlejewski K et al (2017) A cluster of fatal tick-borne encephalitis virus infection in organ transplant setting. J Infect Dis 215(6):896–901. https://doi.org/10.1093/infdis/jix040

10. Lu G, Gong P (2013) Crystal structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. PLoS Pathog 9(8):e1003549. https://doi.org/10.1371/journal.ppat.1003549

11. Upadhyay AK, Cyr M, Longenecker K, Tripathi R, Sun C, Kempf DJ (2017) Crystal structure of full-length Zika virus NS5 protein reveals a conformation similar to Japanese encephalitis virus NS5. Acta Crystallogr Sect F Struct Biol Commun 73(Pt 3):116–122. https://doi.org/10.1107/S2053230X17001601

12. Badshah SL, Ahmad N, Rehman AU, Khan K, Ullah A, Alsayari A (2019) Molecular docking and simulation of Zika virus NS3 helicase. BMC Chem 13:67. https://doi.org/10.1186/s13065-019-0582-y

13. Chen KC, Lin YUF, Huang ANC (2019) Molecular interaction of the antiviral compound CW-33 and its analogues with the NS2B-NS3 protease of the Japanese encephalitis virus. Int J Mol Med 43:2024–2032. https://doi.org/10.3892/ijmm.2019.4113

14. Varadi M, Anyango S, Deshpande M et al (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50(D1):D439–D444. https://doi.org/10.1093/nar/gkab1061

15. Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. Neuron 99(6):1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

16. Yoshii K (2019) Epidemiology and pathological mechanisms of tick-borne encephalitis. J Vet Med Sci 81:343–347. https://doi.org/10.1292/jvms.18-0373

17. Waterhouse A, Bertoni M, Bienert S et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46(W1):W296–W303. https://doi.org/10.1093/nar/gky427

18. Bienert S, Waterhouse A, de Beer Tjaart A. P, Tauriello G, Studer G, Bordoli L, Schwede T (2016) The SWISS-MODEL repository—new features and functionality. Nucleic Acids Res 45(D1):D313–D319. https://doi.org/10.1093/nar/gkw1132

19. Guex N, Peitsch MC, Schwede T (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis 30(Suppl 1):S162-73. https://doi.org/10.1002/elps.200900140

20. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T (2020) QMEANDisCo-distance constraints applied on model quality estimation. Bioinformatics 36(6):1765–1771. https://doi.org/10.1093/bioinformatics/btz828

21. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T (2017) Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci Rep 7(1):10480. https://doi.org/10.1038/s41598-017-09654-8

22. Jumper J, Evans R, Pritzel A et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2

23. ...Ahdritz G, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, O'Donnell TJ, Berenberg D, Fisk I, Zanichelli N, Zhang B, Nowaczynski A, Wang B, Stepniewska-Dziubinska MM, Zhang S, Ojewole A, Guney ME, Biderman S, Watkins AM, Ra S, Lorenzo PR, Nivon L, Weitzner B, Ban YEA, Sorger PK, Mostaque E, Zhang Z, Bonneau R, AlQuraishi M (2022) OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. bioRxiv. https://doi.org/10.1101/2022.11.20.517210

24. Lim SP, Koh JHK, Seh CC, Liew CW, Davidson AD, Chua LS, Chandrasekaran R, Cornvik TC, Shi P-Y, Lescar J (2013) A crystal structure of the dengue virus non-structural protein 5

(NS5) polymerase delineates interdomain amino acid residues that enhance its thermostability and de novo initiation activities. J Biol Chem 288(43):31105–31114. https://doi.org/10.1074/jbc.M113.508606

25. Zhao B, Yi G, Du F, Chuang Y-C, Vaughan RC, Sankaran B, Kao C-C, Li P (2017) Structure and function of the Zika virus full-length NS5 protein. Nat Commun. https://doi.org/10.1038/ncomms14762

26. Jacobsen L, Hungerland J, Bačić V, Gerhards L, Schuhmann F, Solov'yov IA (2023) Introducing the automated ligand searcher (ALISE). J Chem Inf Model 63:7518–7528. https://doi.org/10.1021/acs.jcim.3c01317

27. Korol V, Husen P, Sjulstok E, Nielsen C, Friis I, Frederiksen A, Salo AB, Solov'yov IA (2020) Introducing VIKING: a novel misc platform for multiscale modeling. ACS Omega 5(2):1254–1260. https://doi.org/10.1021/acsomega.9b03802

28. Yang J, Jing X, Yi W, Li XD, Yao C, Zhang B, Zheng Z, Wang H, Gong P (2021) Crystal structure of a tick-borne flavivirus RNA-dependent RNA polymerase suggests a host adaptation hotspot in RNA viruses. Nucleic Acids Res 49(3):1567–1580. https://doi.org/10.1093/nar/gkaa1250

29. Choi KH, Rossmann MG (2009) RNA-dependent RNA polymerases from flaviviridae. Curr Opin Struct Biol 19(6):746–751. https://doi.org/10.1016/j.sbi.2009.10.015

30. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28(1):235–242. https://doi.org/10.1093/nar/28.1.235

31. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009) Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc 4:1–13. https://doi.org/10.1038/nprot.2008.197

32. Barragan AM, Crofts AR, Schulten K, Solov'yov IA (2015) Identification of ubiquinol binding motifs at the Qo-site of the cytochrome bc1 complex. J Phys Chem B 119(2):433–447. https://doi.org/10.1021/jp510022w

33. Hopkins E, Sanvictores T, Sharma S (2022) Physiology, acid base balance. https://www.ncbi.nlm.nih.gov/books/NBK507807/. Accessed 14 Apr 2023

34. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) Propka3: consistent treatment of internal and surface residues in empirical pka predictions. J Chem Theory Comput 7(2):525–537. https://doi.org/10.1021/ct100578z

35. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pka values. J Chem Theory Comput 7(7):2284–2295. https://doi.org/10.1021/ct200133y

36. Phillips JC, Hardy DJ, Maia JDC et al (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys 153(4):044130. https://doi.org/10.1063/5.0014475

37. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26(16):1781–1802. https://doi.org/10.1002/jcc.20289

38. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph Model 14(1):33–38, 27–28. https://doi.org/10.1016/0263-7855(96)00018-5

39. Brooks BR, Brooks CL III, Mackerell AD Jr et al (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30(10):1545–1614. https://doi.org/10.1002/jcc.21287

40. MacKerell AD, Bashford D, Bellott M et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616. https://doi.org/10.1021/jp973084f

41. MacKerell AD Jr, Feig M, Brooks CL III (2004) Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc 126:698–699. https://doi.org/10.1021/ja036959e

42. Bajusz D, Rácz A, Héberger K (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations ? J Cheminform 7:20. https://doi.org/10.1186/s13321-015-0069-3

43. Ellingson SR, Smith JC, Baudry J (2013) VinaMPI: facilitating multiple receptor high-throughput virtual docking on high-performance computers. J Comput Chem 34(25):2212–2221. https://doi.org/10.1002/jcc.23367

44. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Software news and updates AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791. https://doi.org/10.1002/jcc

45. Boyle NMO, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminform 3:33

46. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 10(5):449–461. https://doi.org/10.1517/17460441.2015.1032936

47. Vanommeslaeghe K, Hatcher E, Acharya C et al (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31(4):671–690. https://doi.org/10.1002/jcc.21367

48. Foloppe N, MacKerell AD (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J Comput Chem 21(2):86–104. https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2h86::AID-JCC2i3.0.CO;2-G

49. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, MacKerell AD (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. J Chem Theory Comput 8(9):3257–3273. https://doi.org/10.1021/ct300400x

50. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, MacKerell AD (2012) Optimization of the CHARMM additive force field for DNA: improved treatment of the BI/BII conformational equilibrium. J Chem Theory Comput 8(1):348–362. https://doi.org/10.1021/ct200723y

51. Pavelites JJ, Gao J, Bash PA (1996) A molecular mechanics force field for NAD+, NADH, and the pyrophosphate groups of nucleotides. J Comput Chem 18(2):221–239

52. MacKerell AD, Banavali NK (2000) All-atom empirical force field for nucleic acids: II. application to molecular dynamics simulations of DNA and RNA in solution. J Comput Chem 21(2):105–120. https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P

53. Denning EJ, Priyakumar UD, Nilsson L, MacKerell AD (2011) Impact of 20-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. J Comput Chem 32(9):1929–1943. https://doi.org/10.1002/jcc.21777

54. MacKerell AD, Feig M, Brooks CL (2004) Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc 126(3):698–699. https://doi.org/10.1021/ja036959e

55. Schuhmann F, Kattnig DR, Solov'yov IA (2021) Exploring post-activation conformational changes in pigeon cryptochrome 4. J Phys Chem B 125(34):9652–9659. https://doi.org/10.1021/acs.jpcb.1c02795

56. Schuhmann F, Tan X, Gerhards L, Bordallo HN, Solov'yov IA (2022) The same, but different, but still the same: structural and dynamical differences of neutrophil elastase and cathepsin G. Eur Phys J D 76(7):1–14. https://doi.org/10.1140/epjd/s10053-022-00452-0

57. Shahu MK, Schuhmann F, Scholten A, Solov'yov IA, Koch KW (2022) The transition of photoreceptor guanylate cyclase type 1 to the active state. Int J Mol Sci 23(7):1–17. https://doi.org/10.3390/ijms23074030

58. Hanić M, Schuhmann F, Frederiksen A, Langebrake C, Manthey G, Liedvogel M, Xu J, Mouritsen H, Solov'yov IA (2022) Computational reconstruction and analysis of structural models of avian cryptochrome 4. J Phys Chem B 126(25):4623–4635. https://doi.org/10.1021/acs.jpcb.2c00878

59. Hanić M, Antill LM, Gehrckens AS et al (2023) Dimerization of European Robin cryptochrome 4a. J Phys Chem B 127(28):6251–6264. https://doi.org/10.1021/acs.jpcb.3c01305