p-cymene impairs SARS-CoV-2 and Influenza A (H1N1) viral replication: *in silico* predicted interaction with SARS-CoV-2 nucleocapsid protein and H1N1 nucleoprotein

Athanasios Panagiotopoulos[1], Melpomeni Tseliou[2], Ioannis Karakasiliotis[3], Danai-Maria Kotzampasi[1], Vangelis Daskalakis[4], Nikolaos Kesesidis[3], George Notas[1], Christos Lionis[6,8], Marilena Kampa[1,8], Stergios Pirintsos[6,7,8], George Sourvinos[2,8], Elias Castanas[1,8]

1. University of Crete, School of Medicine, Laboratory of Experimental Endocrinology, Heraklion, Greece
2. University of Crete, School of Medicine, Laboratory of Clinical Virology, Heraklion, Greece
3. Laboratory of Biology, School of Medicine, Democritus University of Thrace, Alexandroupolis, Greece
4. Department of Chemical Engineering, Cyprus University of Technology, Limassol, Cyprus
5. Clinic of Social and Family Medicine, School of Medicine, University of Crete, Heraklion, Greece
6. University of Crete, Department of Biology, Heraklion, Greece
7. Botanical Garden, University of Crete, Rethymnon, Greece
8. Nature Crete Pharmaceuticals, Heraklion, Greece

Supplemental Material

Contents

Supplemental Methods

Molecular Dynamics of SARS-CoV-2 NC protein, with p-cymene and importin A

Molecular Dynamics computational protocol

Model setup

The crystal structure of the SARS-CoV-2 nucleocapsid protein (dimerization domain, residues 249-389) (pdb code: 6WJI, http://www.rcsb.org/structure/6WJI) and human importin A (residues 71-197) (pdb code: 4UAE) (Pumroy et al., 2015) were used as initial coordinates to build our models. A nucleocapsid – importin A complex was produced based on docking calculations with HEX (Ritchie, 2003), as described in the main text. The protonation states of titratable residues were simulated at neutral pH, thus all Glu, and Asp residues were left deprotonated, in accordance also with the PDB2PQR (propka 3.0 method, pH 7.3) predictions (Dolinsky et al., 2004). His-365 was protonated only at the Nε site, whereas His-170 and His-300 only at the Nδ site, to maintain the hydrogen bonding network within the crystal structures. p-cymene was docked to the nucleocapsid protein in the GalaxyWeb server, as detailed in the Material and Methods section. Thus, two models were built; one nucleocapsid-importin complex without p-cymene, and one with p-cymene. The all-atom models, as defined previously, were embedded in cubic boxes of volume around 10^3 nm^3. 30267 TIP3P water molecules (Mark and Nilsson, 2001) were used to hydrate each protein complex. Ion (K^+, Cl^-) concentration was set at the value of 150 mM to mimic the physiological salt content, in addition to a surplus of Cl^- (8 ions) to neutralize the protein charges in each sample, resulting in simulation unit boxes of up to 95186 atoms. The Amber03 (Duan et al., 2003) force field was used for the residues and ions. The Amber03 parameters for p-cymene were derived based on the ACPYPE algorithm (Sousa da Silva and Vranken, 2012).

Equilibration-Production Molecular Dynamics setup

The equilibration-relaxation for the all-atom systems is employed based on a published protocol for water-soluble proteins (Petratos et al., 2020). This contains a steepest descend energy minimization with a tolerance of 0.5 kJ mol^{-1} for 1000 steps, and a sequence of isothermal (nVT), isothermal-isobaric (nPT) runs with the gradual relaxation of the constraints on protein heavy atoms (from 10^4 in steps1-2 to 10^3 kJ mol^{-1} nm^{-2} in step-4) and Cα atoms (from 10^3 in step-5, to 10^2 in step-6, 10 in step-7, 1 in step-8 and 0 kJ mol^{-1} nm^{-2} in step-9) for around 30 ns, with a time step of 1.0 fs (steps 1-4) and 2.0 fs (steps 5-9).

In detail: (**step-1**) Constant density and temperature (nVT) Brownian dynamics (BD) at 100 K for 50 ps that employs the Berendsen thermostat ,[6] with a temperature coupling constant at 1.0 fs. (**steps 2-3**) Two short constant density (nVT) and constant pressure (nPT) runs for 100 ps each, with a weak coupling Berendsen thermostat and barostat (Berendsen et al., 1984) at 100 K employing time coupling constants of 0.1 ps for the temperature and isotropic 50.0 ps

coupling for the pressure with a compressibility of $4.6x10^{-5}$. (**step-4**) Heating from 100 to 250 K in a constant density ensemble (nVT) for 3 ns employing the v-rescale thermostat (Bussi et al., 2007), with a time coupling constant of 0.1 ps. (**step-5**) Heating from 250 to 310K in a constant pressure ensemble (nPT) for 2 ns, employing the v-rescale thermostat (Bussi et al., 2007) and Berendsen barostat (Berendsen et al., 1984), with time coupling constants of 0.1 ps for the temperature and 2.0 ps for the pressure, removing also all but the Cα-atom protein position restraints. (**step-6**) Equilibration at 310K (0.1 ps temperature coupling constant) for 5 ns (nPT, 1 atm, 2.0 ps coupling constant for pressure. (**steps 7-8**) Equilibration at 310K (0.5 ps temperature coupling constant) for 5 ns (nPT, 1 atm, 2.0 ps coupling constant for pressure). (**step-9**) Equilibration at 310K (0.5 ps temperature coupling constant) for 10 ns (nPT, 1 atm, 2.0 ps coupling constant for pressure). The barostats – thermostats employed for steps 6-9 were the same as in the production trajectories that follow.

For the production all-atom classical Molecular Dynamics (MD), the Newton's equations of motion are integrated with a time step of 2.0 fs at 310K. All classical MD production simulations were run with the leap-frog integrator in GROMACS 2020 (Berendsen et al., 1995) for 2.0 μs each trajectory, for seven replicas per sample (with, without p-cymene) starting from seven different structures extracted at the equilibration process. They were performed at the constant pressure nPT ensemble, with isotropic coupling (compressibility at $4.5x10^{-5}$) employing the v-rescale thermostat (Bussi et al., 2007) (310K, temperature coupling constant 0.5) and the *Parrinello-Rahman* barostat (Parrinello and Rahman, 1981) (1 atm, pressure coupling constant 2.0). Details for parameters can be found in earlier work (Petratos et al., 2020). Van der Waals interactions were smoothly switched to zero between 1.0-1.2 nm with the VERLET cut-off scheme. Electrostatic interactions were truncated at 1.2 nm (short-range) and long-range contributions were computed within the PME approximation (Darden et al., 1993; I-C and Berkowitz, 1999). Hydrogen bond lengths were constrained employing the LINCS algorithm (Hess et al., 1997).

Markov State Modeling

The Classical MD trajectories were analysed based on Markov State Modelling (MSM) (Pande et al., 2010; Prinz et al., 2011; Chodera and Noe, 2014) in order to enable the extraction of long-time-scale dynamics from rather short-time-scale MD trajectories of different states. The application and accuracy of the powerful MSM theory has been presented in many cases also by experiments that include protein−protein, or protein-drug binding kinetics, as well as protein folding rates and protein dynamics (Voelz et al., 2010; Plattner and Noe, 2015; Plattner et al., 2017; Durrant et al., 2020). Our objective was to approximate the slow dynamics in a statistically efficient manner. Thus, a lower dimensional representation of our simulation data was necessary. In order to reduce the dimensionality of our feature space, we employed the time-structure independent components analysis (tICA) which yields a representation of our molecular simulation data with a reduced dimensionality and can

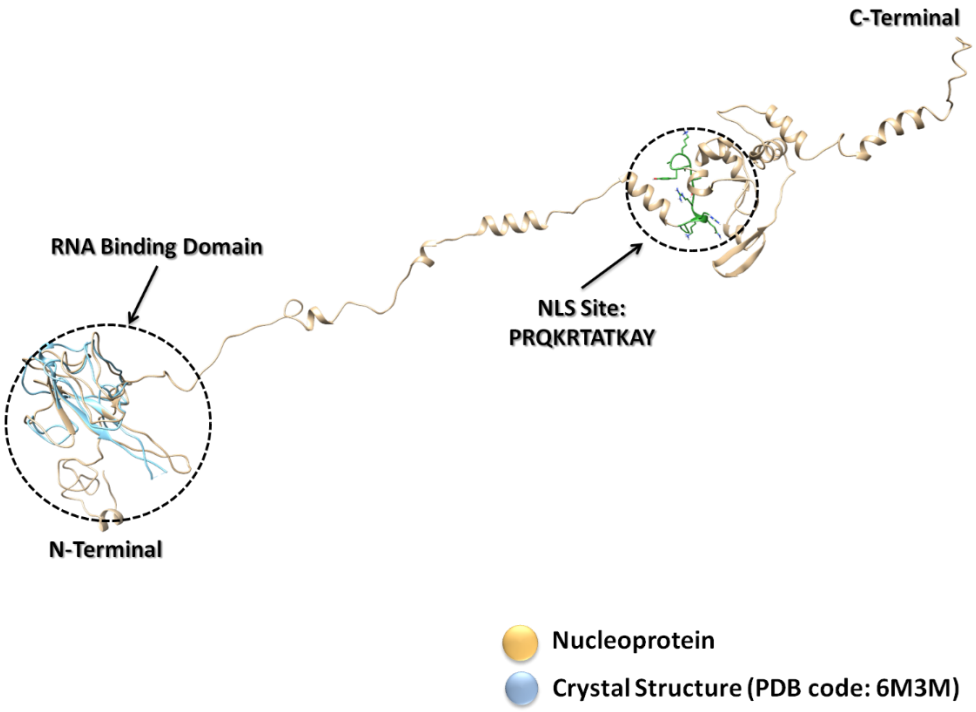
greatly facilitate the decomposition of our system into the discrete Markovian states necessary for MSM estimation. The conformations of the most rigid part of the nucleocapsid protein (71-197 and 249-389) were projected on these slowest modes as defined by the tICA method, then the trajectory frames were clustered into 50 cluster-centers (microstates) by the k-means clustering algorithm, as implemented in PyEMMA (Scherer et al., 2015). Conformational changes of a system can be simulated as a Markov chain, if the transitions between the different conformations are sampled at long enough time intervals so that each transition is Markovian. This means that a transition from one conformation to another is independent of the previous transitions. Therefore, an MSM is a memoryless model. The uncertainty bounds were computed using a Bayesian scheme (Noe, 2008; Trendelkamp-Schroer et al., 2015). We found that the slowest implied timescales converge quickly and are constant within a 95% confidence interval for lag times above 25ns. The validation procedure is a standard approach in the MSM field., a lag time of 25 ns was selected for Bayesian model construction, and the resulting models were validated by the Chapman-Kolmogorov (CK) test. Subsequently, the resulting MSMs were further coarse grained into a smaller number of three metastable states or microstates, using PCCA++ as implemented in PyEMMA. The optimum number of microstates (three) was proposed based on the VAMP2-score (Wu and Noé, 2020). Both the convergence of the implied timescales, as well as the CK test confirm the validity and convergence of the MSM. The CK test indicates that predictions from the built MSM agree well with MSMs estimated with longer lag times. Thus, the model can describe well the long-time-scale behavior of our system within error. The tICA method identified the torsional angles of the following Nucleocapsid residues: 82-83, 88, 100, 102-103, 107-108, 113, 118, 121, 162-164, 186-189, 191, 193-195, 197, 252, 256, 261, 264-265, 268-269, 270, 272-278, 280-281, 283-287, 293-294, 299, 307, 310, 313-315, 318-319, 322, 328-336, 338, 353, 356-358, 360-363, 365-366, 369-371, 373-376, 379-381, 384-385, 387 as the most important features in the Nucleocapsid-Importin interaction, by setting a series of thresholds for the coefficients in the tICA vectors. At first, we set a threshold of 0.09. We continued by setting a threshold of 0.07 for the coefficients in the tICA vectors of the filtered data and afterwards a threshold of 0.05. We end up with a VAMP score of 2.24. For the selection of these thresholds we checked for different thresholds the VAMP2-score and the states projected onto the first two independent components.

The residues that seem to affect the Nucleocapsid-Importin dynamics are derived based on the same MSM/ tICA protocol described before also in the presence of p-cymene. These include: 78, 80, 82, 86-87, 100-102, 113, 115, 121, 146, 150, 156-157, 161, 185-187, 191-197, 259-263, 271-272, 274-278, 280, 283-287, 289, 292-293, 310-318, 332, 334-336, 339-340, 353-354, 361-364, 367, 375, 377, 382-385. At first, we set a threshold of 0.07. We continued by setting a threshold of 0.085 for the coefficients in the tICA vectors of the filtered data, a threshold of 0.125, and thus we concluded in the previously referred residues.

Multiple Walkers: Enhanced Molecular Dynamics Sampling

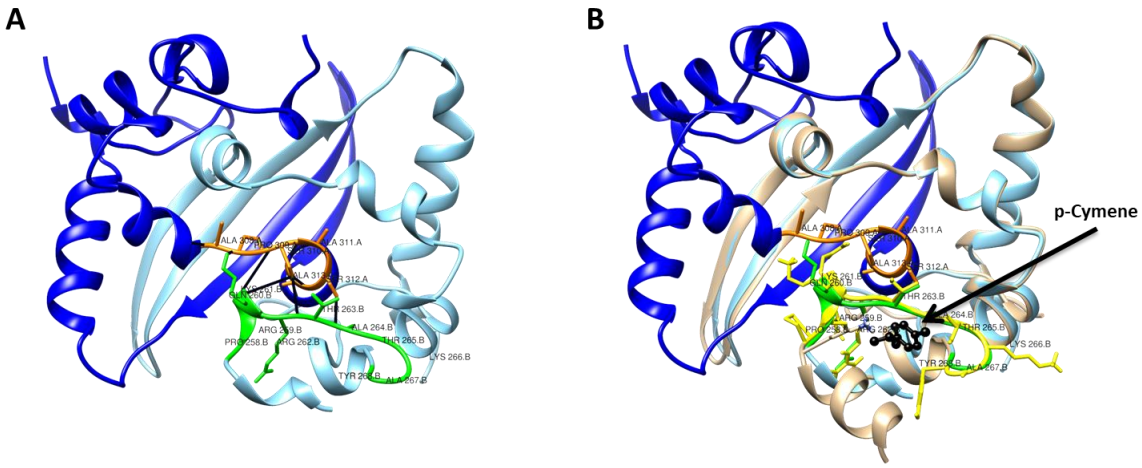
To enhance the conformational sampling on the SARS-CoV-2 Nucleocapsid dimer we employed the multiple walkers (MW) variant of metadynamics (Raiteri et al., 2006). Seven replicas per sample were run from different nucleocapsid-importin complex configurations (starting structures, *see above*). Replicas were allowed to exchange every 1000 steps for 0.5μs each, which gave an exchange probability around 20%, consistent with the large sample sizes. The Collective Variables (CVs) chosen for the MW runs were the radius of gyration of the complex (CV1, Rg) and the nucleocapsid – importing distance (CV2), calculated as an average of the following distances: Lys-361…Asp-83, Glu-280…Arg-125, Arg-259…Glu-124, Lys-256…Asp-193, Lys-387…Asp-83, Lys-369…Glu-74, Arg-276…Asp-83, Glu-323…Arg-125, Lys-370…Asp-109, Lys-374…Glu-74, Lys-249…Asp-193, -Lys-370…Glu-74. These are the common salt bridges that are formed between nucleocapsid and importin, identified by analysis of the classical MD trajectories via the VMD interface (Humphrey et al., 1996). Several residues within these pairs are also indicated as important sites out of the MSM analysis. A combination of the GROMACS 2020/ PLUMED 2.5 (Tribello et al., 2014) engines was employed. A bias factor of 25 at the well-tempered ensemble, along with Gaussians of 1.2 kJ/mol initial height, and sigma values (width) of 2.5 (CV1), 0.5 (CV2) in the CV space, deposited every 2 ps, was employed. The grid space for the CVs is defined between 1.5-10 (CV1) and 0-10 (CV2) at a resolution of 0.5 (CV1), 0.1 (CV2). The average Free Energy Surfaces (FES) in the absence and presence of p-cymene, were produced out of the seven individual replicas, employing the equation 9 of ref (Golla et al., 2020).

Supplementary Figures



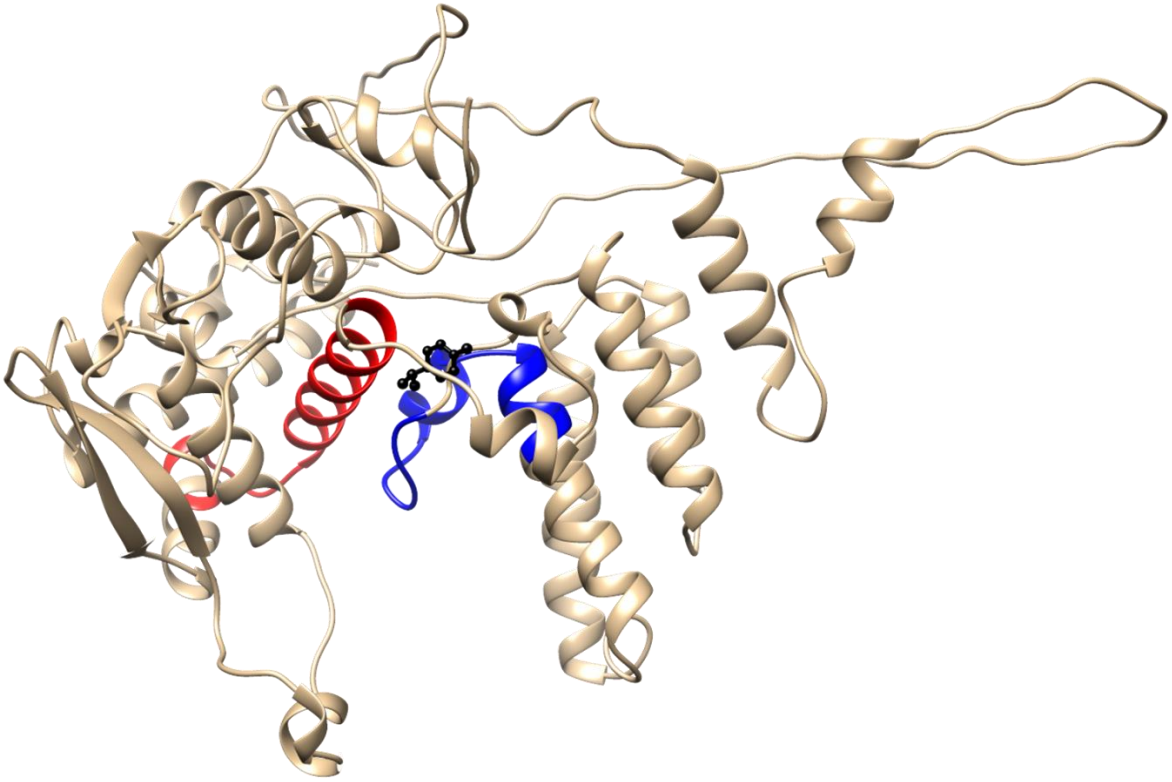
Supplemental Figure 1

3-D modeling of SARS-CoV-2 nucleocapsid protein. The structured N-terminal RNA-binding and the C-terminal dimerization domains are shown, together with the PDB codes of available crystals.



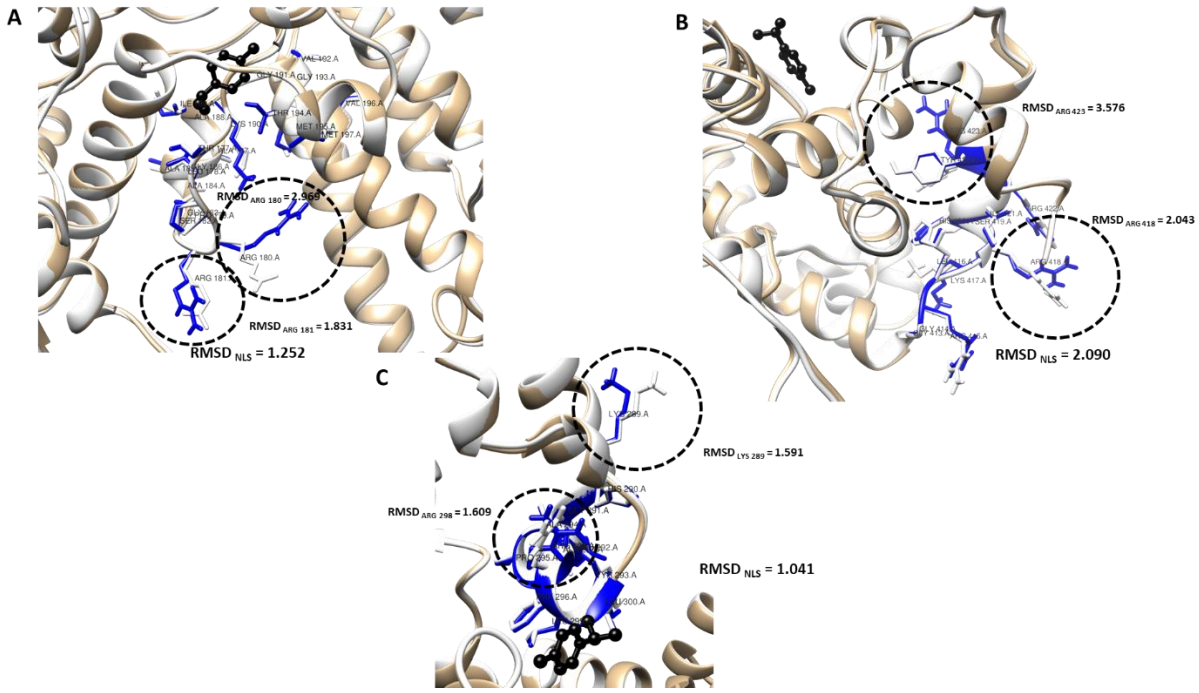
Supplemental Figure 2

Simulation of the dimerization of the N-CTD. In **A**, the antiparallel dimerization of two N-CTD proteins is presented in blue and cyan. The simulated solution presents only minor discrepancies from the reported crystal (Zhou et al., 2020) (total RMSD=0.0.762 Å). The animoacids of the interacting η1 domain are shown. In **B**, after p-cymene binding ((blaqck ball and stick), the impairment of binding of the second N-CTD protein is shown.



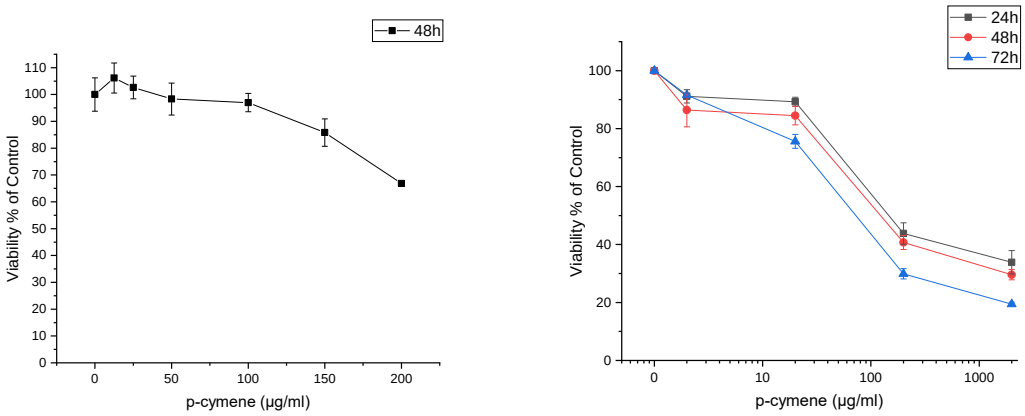
Supplemental Figure 3

Simulation of the binding domains on H1N1 NP for RNA (red) and Importin binding (blue). The p-Cymene also showed in this figure (black, ball and stick).



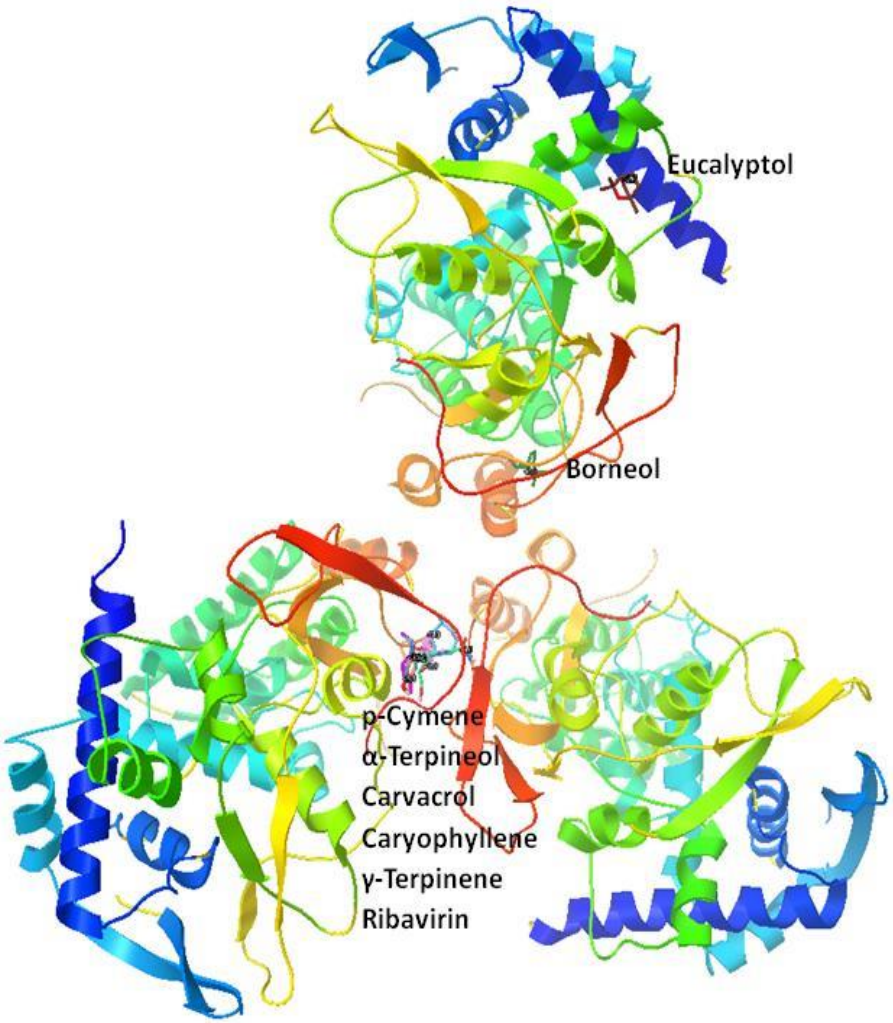
Supplemental Figure 4

Docking of p-cymene with Nucleoproteins of H1N1 (**A**), Ebola (**B**) and Rabies (**C**) viruses. NLS sequences are shown in blue and p-cymene is shown with black atoms and with ball and stick style. With white color, shown the unliganded protein and with gold color, shown the protein docked with p-cymene. The structure's changes (RMSD) refer to changes of the NLS sequence, while in black circles the changes in key aminoacids are presented.



Supplemental Figure 5

In vitro dose-response effect of p-cymene in VERO E6 (left panel) and MDCK cells proliferation (right panel). Cells were cultured in 24-well plates and treated with the indicated concentrations of p-cymene (0-2 mg/ml) for 48 h (VERO E6 cells, as this is the time point in which significant results were reported) and 24, 48 and 72 hours (MDCK cells). Cell viability was measured by MTT assay. Data represent three independent experiments (mean ± SE).



Supplemental Figure 6

Molecular simulation of binding of the main constituents of an essential oil combination of three aromatic plants on Influenza A nucleoprotein (Tseliou et al., 2019). A fully flexible docking of the different compounds was performed in the GalaxyWeb environment. Then, we have extracted the ligand final pose and superimposed in on the nucleiprotein- p-cymene structure.

Supplementary References

Berendsen H, Postma J, van Gunsteren W, DiNola A and Haak J (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**:3684–3690.

Berendsen H, van der Spoel D and van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* **91**:43–56.

Bussi G, Donadio D and Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* **126**:014101.

Chodera JD and Noe F (2014) Markov state models of biomolecular conformational dynamics. *Curr Opin Struct Biol* **25**:135-144.

Darden T, York D and Pedersen L (1993) Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J Chem Phys* **98**:10089–10092.

Dolinsky TJ, Nielsen JE, McCammon JA and Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**:W665-667.

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J and Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* **24**:1999-2012.

Durrant JD, Kochanek SE, Casalino L, Ieong PU, Dommer AC and Amaro RE (2020) Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Cent Sci* **6**:189-196.

Golla VK, Prajapati JD, Joshi M and Kleinekathofer U (2020) Exploration of Free Energy Surfaces Across a Membrane Channel Using Metadynamics and Umbrella Sampling. *J Chem Theory Comput* **16**:2751-2765.

Hess B, Bekker H, Berendsen H and Fraaije J (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* **18**:1463–1472.

Humphrey W, Dalke A and Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**:33-38, 27-38.

I-C Y and Berkowitz M (1999) Ewald summation for systems with slab geometry. *J Chem Phys* **111**:3155–3162.

Mark P and Nilsson L (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem A* **105**:9954–9960.

Noe F (2008) Probability distributions of molecular observables computed from Markov models. *J Chem Phys* **128**:244103.

Pande VS, Beauchamp K and Bowman GR (2010) Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**:99-105.

Parrinello M and Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **52**:7182–7190.

Petratos K, Gessmann R, Daskalakis V, Papadovasilaki M, Papanikolau Y, Tsigos I and Bouriotis V (2020) Structure and Dynamics of a Thermostable Alcohol Dehydrogenase from the Antarctic Psychrophile Moraxella sp. TAE123. *ACS Omega* **5**:14523-14534.

Plattner N, Doerr S, De Fabritiis G and Noe F (2017) Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat Chem* **9**:1005-1011.

Plattner N and Noe F (2015) Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun* **6**:7653.

Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schutte C and Noe F (2011) Markov models of molecular kinetics: generation and validation. *J Chem Phys* **134**:174105.

Pumroy RA, Ke S, Hart DJ, Zachariae U and Cingolani G (2015) Molecular determinants for nuclear import of influenza A PB2 by importin alpha isoforms 3 and 7. *Structure* **23**:374-384.

Raiteri P, Laio A, Gervasio FL, Micheletti C and Parrinello M (2006) Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J Phys Chem B* **110**:3533-3539.

Ritchie DW (2003) Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2. *Proteins* **52**:98-106.

Scherer MK, Trendelkamp-Schroer B, Paul F, Perez-Hernandez G, Hoffmann M, Plattner N, Wehmeyer C, Prinz JH and Noe F (2015) PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J Chem Theory Comput* **11**:5525-5542.

Sousa da Silva AW and Vranken WF (2012) ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes* **5**:367.

Trendelkamp-Schroer B, Wu H, Paul F and Noe F (2015) Estimation and uncertainty of reversible Markov models. *J Chem Phys* **143**:174101.

Tribello G, Bonomi M, Branduardi D, Camilloni C and Bussi G (2014) PLUMED 2: New feathers for an old bird. *Comput Phys Commun* **185**:604–613.

Tseliou M, Pirintsos SA, Lionis C, Castanas E and Sourvinos G (2019) Antiviral effect of an essential oil combination derived from three aromatic plants (Coridothymus capitatus (L.) Rchb. f., Origanum dictamnus L. and Salvia fruticosa Mill.) against viruses causing infections of the upper respiratory tract. *J Herb Med* **17-18**:100288.

Voelz VA, Bowman GR, Beauchamp K and Pande VS (2010) Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J Am Chem Soc* **132**:1526-1528.

Wu H and Noé F (2020) Variational Approach for Learning Markov Processes from Time Series Data. *J Nonlinear Sci 30, 23–66* **30**:23-66.

Zhou R, Zeng R, von Brunn A and Lei J (2020) Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Mol Biomed* **1**:2.