

Report P5.1.1: Realizzazione dei modelli strutturali delle proteine di superficie del virus.

Obiettivo 5.1 di PULVIRUS

L'obiettivo dello studio di modellistica molecolare *in silico* proposto è l'identificazione delle potenziali interazioni molecolari tra aerosol atmosferico (PM) e le proteine strutturali di superficie di SARS-CoV-2. Lo studio è articolato in tre step: 1) la modellazione della componente biologica, 2) la modellazione del particolato atmosferico (PM) e 3) la modellazione dell'interfaccia virus-PM mediante tecniche di tipo docking molecolare e dinamica molecolare simulata.

Questo documento riporta i risultati ottenuti nello step 1.

Step 1: Modellazione della Componente Biologica

Il virus SARS-CoV-2 ha una morfologia rotondeggiante con dimensioni di circa 150 nm di diametro. La superficie del virus è caratterizzata dal possedere una membrana lipidica nella quale sono inserite tre proteine di superficie

- la glicoproteina S (Spike),
- la proteina di membrana M (Membrane) e
- la proteina di rivestimento E (Envelope).

Queste componenti formano il capsido del virus.

All'interno del capsido è presente il materiale genetico e la proteina N (Nucleocapsid), una proteina strutturale che serve a impacchettare e proteggere il materiale genetico.

Nonostante le attuali tecniche computazionali consentano, in principio, di modellare l'intero virus, questo processo rimane ancora un lavoro estremamente complesso che richiede molto tempo e costi computazionali elevati. Per ridurre il costo computazionale ed accorciare i tempi di modellazione è stato realizzato un modello semplificato costituito da un frammento della membrana lipidica dell'involucro virale, nella quale sono stati inseriti i modelli strutturali delle tre proteine S, M e E.

Il modello della proteina E

La proteina E è una piccola proteina strutturale presente sulla membrana virale che svolge un ruolo fondamentale nelle fase di assemblaggio del virione.

La proteina E di SARS-CoV-2 è formata da 75 residui e comprende tre domini:

- un corto dominio idrofilico al N-terminale (NTD)
- un singolo dominio idrofobico transmembranale (TM)

- un dominio idrofilico molto lungo al C-terminale (CTD)

Un aspetto controverso è la topologia che la proteina E può assumere nelle membrane. Tra le topologie ipotizzate quella caratterizzata dall'aver il dominio NTD rivolto verso l'esterno e il dominio CTD rivolto nel citoplasma ($N_{\text{exo}}C_{\text{cyto}}$) appare essere la più plausibile (Ruch et al 2012). Diversi studi mostrano che il dominio idrofobico transmembranale (TM) può andare incontro a oligomerizzazione per formare canali ionici non-selettivi pentamerici (viroporine). Nel nostro studio, la struttura della intera proteina E è stata modellata utilizzando un approccio che si basa sull'omologia di sequenza (homology modeling) con strutture (templati) risolte con tecniche sperimentali (NMR o X-Ray). I templati utilizzati sono (1) la struttura della proteina E di SARS-CoV-2, ottenuta mediante NMR, contenente il solo dominio idrofobico TM in forma pentamerica (Mandala et al 2020) e (2) la struttura della proteina E di SARS-CoV-1, ottenuta mediante NMR, contenente il dominio TM in forma pentamerica e il dominio idrofilico CTD (Surya et al 2018). L'identità di sequenza con il template di SARS-CoV-1 superiore a 90% garantisce la creazione di un buon modello per la parte CTD. Il dominio idrofilico NTD mancante in entrambi i templati, è stato ricostruito mediante MODELLER (Fiser & Sali 2003). Sono stati generati 10 modelli distinti e ogni modello è stato ottimizzato e rifinito mediante appropriati algoritmi di minimizzazione dell'energia seguiti da dinamiche di simulazioni di annealing. I modelli ottenuti sono stati analizzati e validati con diversi tool, tra i quali PROCHECK (Laskowski et al 1993). Il modello migliore è stato scelto tenendo conto di diversi parametri tra i quali l'energia del sistema e la qualità stechiometrica. In **Fig. 1** è mostrato il miglior modello ottenuto con questa procedura di modeling e utilizzato per essere inserito nella membrana virale.

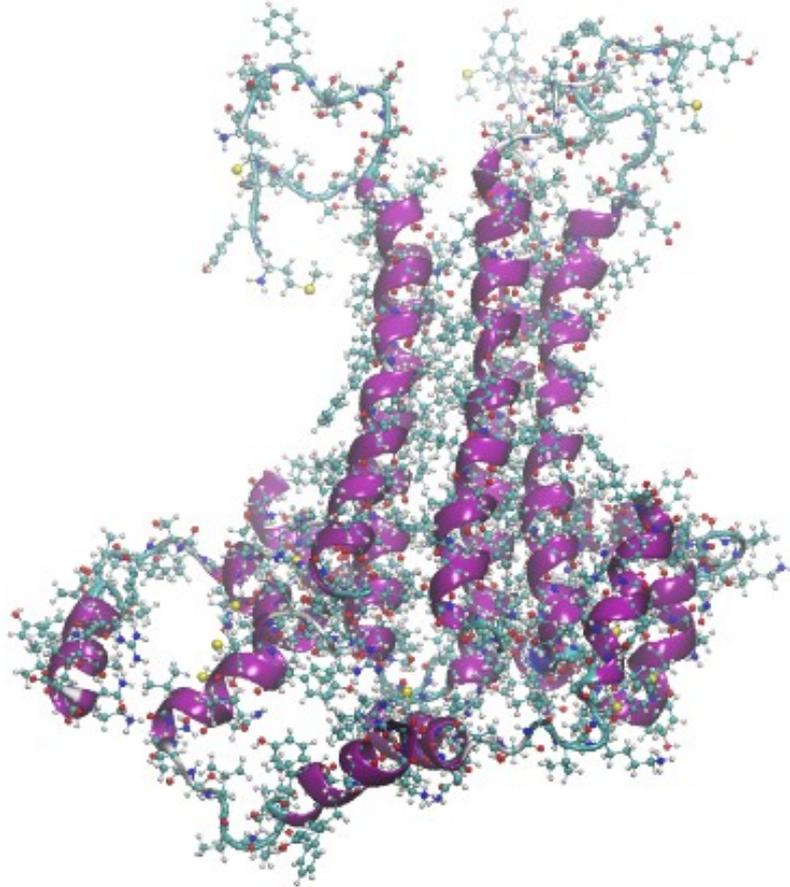


Fig. 1 Il miglior modello strutturale della proteina E, ottenuto mediante homology modeling.

Il modello della proteina M

La proteina M è la proteina più abbondante presente nell'involucro virale dei coronavirus e dirige l'assemblaggio del virus mediante interazioni con le altre proteine strutturali.

La proteina M di SARS-CoV-2 è formata da 222 residui e, in analogia alla omologa SARS-CoV-1 (Hu et al 2003), comprende sette domini:

- un dominio idrofilico al N-terminale (NTD)
- un dominio idrofobico transmembranale (TM-I)
- un dominio idrofilico citoplasmatico
- un dominio idrofobico transmembranale (TM-II)
- un dominio idrofilico extracellulare
- un dominio idrofobico transmembranale (TM-III)
- un dominio idrofilico molto lungo al C-terminale (CTD)

In analogia con la proteina M di SARS-CoV-1, anche la sequenza della proteina M di SARS-Cov-2, possiede tre domini transmembranalni ovvero una regione che attraversa per tre volte la

membrana virale. Pur avendo una identità di sequenza superiore al 90% con SARS-CoV-1, la struttura della proteina M di quest'ultimo virus rimane scarsamente caratterizzata. La mancanza di informazioni dettagliate sulla sua struttura e funzione è per gran parte dovuta al fatto che essendo una proteina del rivestimento virale tende a formare aggregati insolubili. Pertanto non esiste ad oggi una struttura di riferimento nemmeno per SARS-CoV-1. La mancanza di una struttura ottenuta mediante tecniche sperimentali che possa fungere da template e quindi da struttura di riferimento rende la costruzione del modello della proteina M uno step complicato e delicato. Esistono alcuni tentativi di costruzione di modelli strutturali della proteina M di SARS-CoV-2 in forma monomerica mediante metodologie template-free o *ab-initio* ([Mahtarin et al 2020](#) e [Hasan et al 2020](#)). Nel ricostruire un frammento di membrana virale di SARS-CoV-2 abbiamo scelto di modellare la proteina M sotto forma di omodimero. Per ricostruire la struttura del omodimero della proteina M abbiamo sfruttato il modello ottenuto con TrRosetta da [Mahtarin et al 2020](#) e un modello costruito con SWISS_MODEL utilizzando come template la struttura dimerica di un piccolo frammento della proteina M di SARS-CoV-1, ottenuta mediante X-Ray ([Liu 2011](#)). Queste due strutture sono state utilizzate come template per MODELLER e, come per la proteina E, sono stati prodotti 10 diversi modelli. Tutti i modelli prodotti sono stati sottoposti ad una analisi strutturale ed energetica. In **Fig. 2** è mostrato il migliore modello strutturale ottenuto con questa procedura di modeling.



Fig. 2 Il miglior modello strutturale della proteina M, ottenuto mediante un approccio che combina tecniche ab-initio e homology modeling

INSERIMENTO delle singole proteine nella MEMBRANA VIRALE

Entrambe le proteine, E e M, sono state inserite all'interno di un doppio strato lipidico che funge da membrana virale. La composizione lipidica della membrana è stata ottenuta dalla letteratura ([Woo et al 2020](#)) ed è rappresentata da una miscela eterogenea di diversi fosfolipidi, sfingomieline e molecole di colesterolo. L'inserimento delle proteine E ed M è stato eseguito utilizzando gli algoritmi presenti nella piattaforma CHARMM-GUI ([Jo et al 2008](#)). Al sistema costituito da proteina e membrana è stato aggiunto il solvente (molecole di acqua) e del sale (NaCl) disciolto in concentrazione pari a 150 mM. L'intero sistema assemblato è stato sottoposto a una procedura di "equilibratura" durante la quale sono stati applicati diversi vincoli per posizionare i gruppi della testa lipidica paralleli all'asse Z e per impedire alle molecole di acqua di entrare nella regione idrofobica del sistema occupata dalla membrana. Il processo di equilibratura è stato eseguito in dinamica molecolare simulata con pressione e temperatura costante (NPT ensemble) a 303.15K. Alla fine della procedura sono stati generati i parametri per il forcefield e i file di input per

GROMACS. Nella **Fig. 3** è riportato il sistema ottenuto dopo la procedura di assemblaggio della proteina E inserita nella membrana virale.

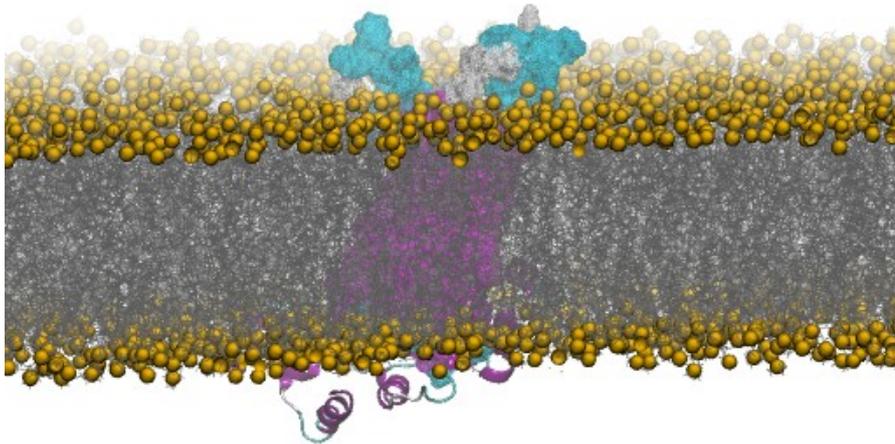


Fig. 3 Proteina E inserita nella membrana virale. Le sfere gialle rappresentano le teste polari dei fosfolipidi, mentre la catena idrofobica del fosfolipide è colorata in grigio

Il modello della proteina S in membrana virale

La proteina S possiede un singolo dominio transmembranale, si trova sull'involucro del virione e funziona da proteina di fusione e ancoraggio alla cellula ospite. La proteina S è la proteina che ha ricevuto maggiore attenzione da parte della comunità scientifica essendo la proteina che media il legame del virus al recettore e l'ingresso nella cellula ospite. Il recettore con il quale la proteina S interagisce è l'enzima di conversione dell'angiotensina umana- 2 (ACE2).

La proteina Spike è un omotrimerico. Ogni monomero è costituito da 1273 residui e possiede due subunità (S1 e S2) separate da un sito di scissione riconosciuto dalle proteasi dell'ospite. Un certo numero di studi strutturali, che sfruttano la microscopia elettronica criogenica (cryo-EM), hanno fornito numerosi dati e una buona comprensione della struttura della proteina S ad una risoluzione quasi atomica. La subunità S1 responsabile del legame del recettore (il dominio ectoplasmatico) è composta dal peptide segnale (SP), dal dominio N terminale (NTD) e dal dominio di legame del recettore (RBD) e la subunità S2 responsabile della fusione di membrana è composta dal peptide di fusione (FP), due ripetizioni heptad (HR1 e HR2), dal dominio transmembranale (TM) e dal dominio citoplasmatico (CP o CTD). Diverse strutture del dominio ectoplasmatico, dei complessi RBD/ACE2 e altri domini extracellulari (HR1 e HR2) della proteina Spike sono disponibili nel

database di strutture proteiche PDB. Dalla analisi di queste strutture si evince che i tre domini RBD posti sulla parte superiore della testa della proteina S sono conformazionalmente variabili. Nelle conformazioni chiuse, tutti e tre i domini RBD rimangono piatti con il motivo di legame del recettore occluso dai RBD sui monomeri vicini. Nelle conformazioni aperte, uno o più RBD sollevano ed espongono i motivi di legame del recettore. Della proteina S oltre alle diverse strutture cristallizzate del dominio ectoplasmatico sono disponibili anche alcuni modelli della proteina intera e glicosilata realizzati da un gruppo Koreano ([Woo et al 2020](#) e [Choi et al 2020](#)) di modellisti e messi a disposizione della comunità scientifica. Nella **Fig. 4** è riportato uno dei modelli strutturali utilizzato in questo studio e inserito nel frammento virale.

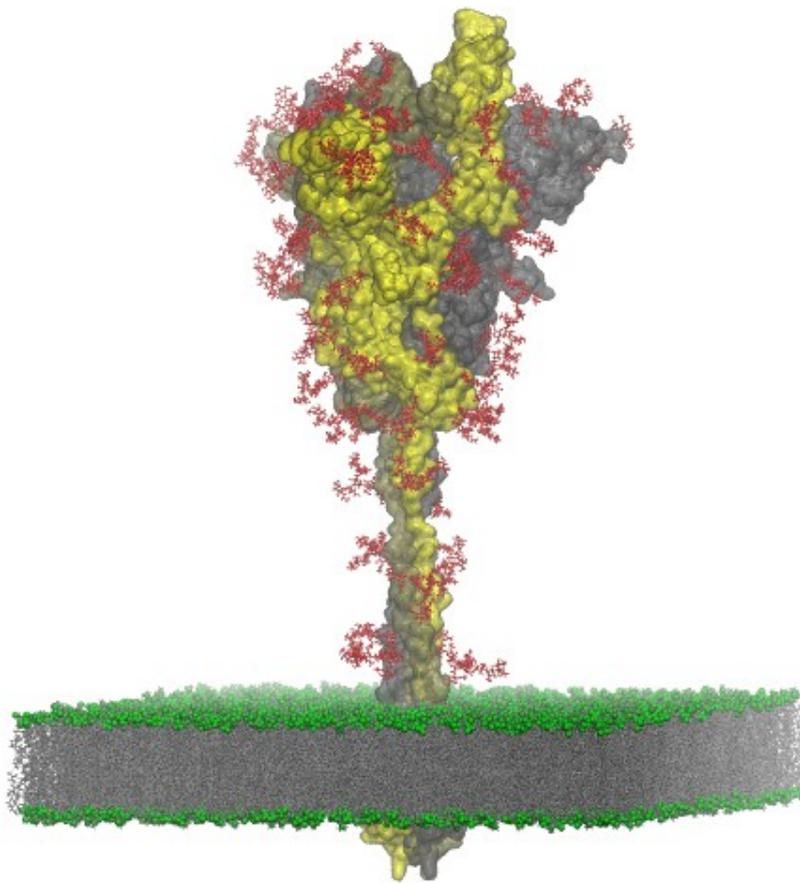


Fig. 4 Modello strutturale della intera proteina S glicosilata di SARS-CoV-2 in membrana virale. Il modello 6svb con il dominio RBD/NTD nella conformazione aperta. Le tre catene individuali della proteina S sono colorate in giallo, grigio e bianco, rispettivamente mentre i glicani sono colorati in rosso. I siti di palmitoilazione sono evidenziati in colore celeste. I gruppi polari e le catene idrofobiche della membrana virale sono colorati in verde e grigio rispettivamente.

In conclusione la componente biologica dell'interfaccia Virus-PM è stata modellata con successo.

References:

- 1) Choi, Y.K., Cao, Y., Frank, M., Woo, H., Park, S.-J., Yeom, M.S., Croll, T.I., Seok, C., Im, W., 2020. Structure, Dynamics, Receptor Binding, and Antibody Binding of Fully-glycosylated Full-length SARS-CoV-2 Spike Protein in a Viral Membrane (preprint). *Biophysics*. <https://doi.org/10.1101/2020.10.18.343715>
- 2) Fiser, A., Sali, A., 2003. Modeller: generation and refinement of homology-based protein structure models. *Meth. Enzymol.* 374, 461–491. [https://doi.org/10.1016/S0076-6879\(03\)74020-8](https://doi.org/10.1016/S0076-6879(03)74020-8)
- 3) Giri, R., Bhardwaj, T., Shegane, M., Gehi, B.R., Kumar, P., Gadhave, K., Oldfield, C.J., Uversky, V.N., 2020. Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. *Cell. Mol. Life Sci.* <https://doi.org/10.1007/s00018-020-03603-x>
- 4) Hasan, S., Hossain, M.M., 2020. Protein Structure Analysis Method of SARS-COV-2 M Protein for Possible Clues Regarding Virion Stability, Longevity and. *Bio-protocol* e5007–e5007.
- 5) Hu, Y., Wen, J., Tang, L., Zhang, H., Zhang, X., Li, Y., Wang, Jing, Han, Y., Li, G., Shi, J., Tian, X., Jiang, F., Zhao, X., Wang, Jun, Liu, S., Zeng, C., Wang, Jian, Yang, H., 2003. The M Protein of SARS-CoV: Basic Structural and Immunological Properties. *Genomics, Proteomics & Bioinformatics* 1, 118–130. [https://doi.org/10.1016/S1672-0229\(03\)01016-7](https://doi.org/10.1016/S1672-0229(03)01016-7)
- 6) Jo, S., Kim, T., Iyer, V.G., Im, W., 2008. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* 29, 1859–1865. <https://doi.org/10.1002/jcc.20945>
- 7) Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26, 283–291. <https://doi.org/10.1107/S0021889892009944>
- 8) Liu, J., Sun, Y., Qi, J., Chu, F., Wu, H., F, G., T, L., J, Y., Gf, G., 2010. The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *J Infect Dis* 202, 1171–1180. <https://doi.org/10.1086/656315>
- 9) Mahtarin, R., Islam, S., Islam, M.J., Ullah, M.O., Ali, M.A., Halim, M.A., 2020. Structure and dynamics of membrane protein in SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics* 0, 1–14. <https://doi.org/10.1080/07391102.2020.1861983>
- 10) Mandala, V.S., McKay, M.J., Shcherbakov, A.A., Dregni, A.J., Kolocouris, A., Hong, M., 2020. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nature Structural & Molecular Biology* 27, 1202–1208. <https://doi.org/10.1038/s41594-020-00536-8>
- 11) Ruch, T.R., Machamer, C.E., 2012. The Coronavirus E Protein: Assembly and Beyond. *Viruses* 4, 363–382. <https://doi.org/10.3390/v4030363>
- 12) Surya, W., Li, Y., Torres, J., 2018. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1860, 1309–1317. <https://doi.org/10.1016/j.bbamem.2018.02.017>
- 13) Woo, H., Park, S.-J., Choi, Y.K., Park, T., Tanveer, M., Cao, Y., Kern, N.R., Lee, J., Yeom, M.S., Croll, T.I., Seok, C., Im, W., 2020. Developing a Fully Glycosylated Full-Length

SARS-CoV-2 Spike Protein Model in a Viral Membrane. *J. Phys. Chem. B* 124, 7128–7137.
<https://doi.org/10.1021/acs.jpcc.0c04553>