



SAPIENZA
UNIVERSITÀ DI ROMA

PhD COURSE IN BIOCHEMISTRY

**MACROMOLECULAR INTERACTIONS
IN SOLUTION
INTERFEROMETRIC AND TURBIDIMETRIC STUDIES**

PhD candidate:

Roberta Piacentini

Cycle XXXVI (2020-2023)

Tutor:

Prof. Alberto Boffi

Coordinator:

Prof. Maria Luisa Mangoni



SAPIENZA
UNIVERSITÀ DI ROMA

PhD COURSE IN BIOCHEMISTRY

**MACROMOLECULAR INTERACTIONS
IN SOLUTION
INTERFEROMETRIC AND TURBIDIMETRIC STUDIES**

PhD candidate:

Roberta Piacentini

Cycle XXXVI (2020-2023)

Tutor:

Prof. Alberto Boffi

Coordinator:

Prof. Maria Luisa Mangoni

TUTTI I DIRITTI RISERVATI – ALL RIGHTS RESERVED

Summary

Acknowledgements.....	1
Chapter 1 Introduction	2
1.1 Protein-protein interaction.....	2
1.1.1 RBD-ACE2 interaction.....	6
1.2 Lactoferrin protein.....	13
1.3 Multivalency of human ferritin	18
1.4 Inhibition of PCSK9 pathway	22
Chapter 2 Materials and methods	29
2.1 Biolayer interferometry	29
2.1.1 Interference	29
2.1.2 BLI applications.....	31
2.1.3 Experimental setup for advanced kinetic assay	31
2.2 Nanoparticle-enhanced turbidimetry	36
2.2.1 Physical principle.....	37
2.2.2 Nanoparticle-enhanced turbidimetric assay	39
2.2.3 Experimental setup	44
2.3 Analytical model.....	45
2.3.1 Saturation curve	48
2.4 2D Zernike computational method.....	51
2.5 Proteins expression and purification	54
2.5.1 Commercial proteins.....	54
2.5.2 Proteins and peptides synthesis	54
2.5.3 Purification protocols.....	55
2.5.4 Protein characterization techniques	57

Chapter 3 Results and discussion.....	60
3.1 Investigation on inhibiting role of lactoferrin in the complex formation between ACE2 receptor and SARS CoV-2 recognition binding domain ...	60
3.1.1 Binding of lactoferrin to ACE2.....	61
3.1.2 Binding of lactoferrin to RBD	62
3.1.3 Effect of lactoferrin in the interaction between RBD and ACE2 ..	64
3.1.4 Computational recognition of the binding regions of ACE2 to lactoferrin.....	70
3.2 Design of protein-binding peptides with controlled binding affinity: the case of SARS-CoV-2 receptor binding domain and ACE2-derived peptides	72
3.2.1 Computational analysis of the ACE2-derived peptides	73
3.2.2 BLI measurements of RBD-peptide complexes	75
3.2.3 Ferritin chimeric nanoparticles	77
3.3 Engineered ferritin and inhibition of PCSK9 metabolic pathway	81
3.3.1 Purification of engineered nanoparticle	81
3.3.2 Structural studies on ferritin HFt-PBM conformation.....	82
3.3.3 Binding kinetics analysis	84
Chapter 4 Conclusion.....	90
4.1 Methods for investigating PPIs	90
4.2 Lactoferrin binding properties	92
4.3 The Zernike approach to the design of mutants	93
4.4 The multivalency of ferritin.....	94
Bibliography.....	97
List of abbreviations	113
Appendix I.....	115
Appendix II	127

Appendix III.....140

Acknowledgements

The experimental work presented in this thesis was carried on at the Department of Biochemical Sciences A. Rossi Fanelli of University of Rome “Sapienza” and supported by Istituto Italiano di Tecnologia (CLN2S@Sapienza, Roma and CHT@Erzelli, Genova).

I would like to begin by expressing my gratitude to Prof. Alberto Boffi, who has been a constant source of support from day one. His ideas and work ethic continue to inspire me, and I am truly fortunate to have had the opportunity to learn from him.

My deepest and sincere gratitude goes to Prof. Alessandra Bonamore and Prof. Alberto Macone, for the way they accepted me and made me feel part of the group with their trust in me.

Deepest thanks must be addressed also to Prof. Paola Baiocco and Dr. Giacomo Parisi for guiding the bashful 1st year student I was in the beginning with kindness and cordiality.

I want to extend a huge thank you to the other researchers, Masters and PhD students that have been or are currently part of the lab group and that have helped me along the way: Alessio Incocciati, Chiara Cappelletti, Lorenzo Barolo, Sofia Botta and many other PhD colleagues that I now consider close friends. The extension goes beyond the Biochemistry department and to the IIT institute, where the people I have cooperated with have instilled in me interest and curiosity in many research fields different from the ones I am in.

Finally, a huge hug must go to my family and friends outside of my work environment that supported me during the ups and downs I encountered in the last 3 years.

Chapter 1 Introduction

1.1 Protein-protein interaction

Proteins form the basic machinery of cells, and the precise interactions among them, known as Protein-Protein Interactions (PPIs), are fundamental for appropriate execution of all cellular mechanisms. These interactions play a pivotal role in governing a plethora of biological processes, ranging from signal transduction and enzymatic activity to immune responses and structural stability. Every PPI is based on a specific arrangement of amino acids within the protein sequences that form the binding sites. These binding sites fit together to create a lock-and-key mechanism that determines which proteins can interact with the appropriate partner, be it a receptor or another protein within the intracellular interactome. The term "interactome" refers to the complete set of molecular interactions that occur within a biological system. The interactome is a comprehensive map of how molecules in a biological system interact with each other to carry out various cellular processes and functions. These interactions can include protein-protein interactions, protein-DNA interactions, protein-RNA interactions, and more. Therefore, it is relevant to consider the intracellular network as well as the protein-receptor interaction in the context of the present dissertation, since experimental results will also be presented regarding protein-protein interactions that physiologically occur within the cytoplasm. The strength and specificity of these interactions may vary according to their purpose; a weak interaction might be transient and serve a signaling effect, whereas a strong interaction could lead to a stable complex with crucial functional implications. PPIs are categorized into different types based on their functions and outcomes. Enzyme-substrate interactions, for instance, drive biochemical reactions by bringing substrates

into proximity with the active site of an enzyme. An enzyme binds to its substrate to catalyze a reaction, after which the products are released, freeing the enzyme to interact with other partners. These types of interactions can be classified as “non-obligated” since they can be transient and reversible. Similarly, receptor-ligand interactions are central to cellular signaling pathways, where extracellular molecules bind to cell surface receptors, thus triggering a cascade of events inside the cell that in turn are mediated by PPIs within cytosolic protein partners, thus constituting the interactome network. Additionally, the formation of protein complexes contributes to processes such as DNA replication, transcription, and translation, ensuring that cellular machinery functions with precision and accuracy. Some of these interactions are labeled as “obligated” because they are critical for the function of multi-domain subunit complexes. Exemplary of this are the components of ribosomes, which are responsible for protein synthesis. Advanced techniques like yeast two-hybrid systems, co-immunoprecipitation, surface plasmon resonance, X-ray crystallography and cryo-EM have revolutionized our ability to study PPIs. These methods enable researchers to identify interacting partners, map binding sites, and even determine the three-dimensional structures of protein complexes at atomic level. Conversely, “biological techniques” such as two-hybrid systems and co-immunoprecipitation, though lacking the structural dimension, are precious tools in order to identify novel interacting partners. Moreover, computational approaches have recently evolved and currently represents invaluable tools, helping predict potential interactions based on protein sequences, structures, and evolutionary relationships. Such insights provide a deeper understanding of cellular processes and often unveil potential targets for drug development. In fact, many therapeutic strategies aim to

modulate PPIs, either by enhancing or disrupting them, in order to treat diseases like cancer, neurodegenerative disorders, and infectious diseases.

One way to disrupt PPIs is to select mutations at the interfaces that can lead to significant changes in binding affinity and result in physiological or pathological phenotypes (Forbes et al., 2015; Landrum et al., 2016). Indeed, the stability of PPIs is the result of a complex fine-tuning of chemical-physical properties at the interfaces and entropic effects (Desantis et al., 2022; Vangone & Bonvin, 2015), making the substitution of even just one residue potentially disruptive. Therefore, in the last years, several computational methods have been developed to predict the effects of mutations on binding, based on a wide variety of techniques (Brender & Zhang, 2015; Geng et al., 2019; Moretti et al., 2013; Rodrigues et al., 2019). The experimental results of this dissertation were obtained starting from computational theoretical predictions on the behavior of the molecular systems under consideration; these predictions were made by applying the 2D Zernike model. Such a protocol is based, beyond a coarse-grained evaluation of electrostatics compatibility, on the application of the 2D Zernike formalism, to obtain a compact representation of the local shape of molecular surfaces (Milanetti, Miotto, Di Rienzo, Monti, et al., 2021). In this framework, the geometry of a molecular region is described by an ordered set of orthogonal polynomials. In a 2D Zernike polynomial expansion, the 2D function or surface is decomposed into a sum of these Zernike polynomials, each multiplied by a coefficient. These coefficients determine the contribution of each Zernike polynomial to the overall shape of the function or surface. The expansion ensures an easy evaluation of the shape complementarity between two molecular regions calculating the distance between the corresponding Zernike descriptors. Therefore, when a residue is substituted, it is possible to evaluate whether the shape of the mutated bind-

ing site is more complementary to the molecular partner. In the past years, this formalism has proven its efficacy in similar optimization protocols (De Lauro et al., 2022; Di Rienzo et al., 2020; Di Rienzo, Miotto, Milanetti, et al., 2023), or more in general to evaluate the local similarity or complementarity (Daberdaku & Ferrari, 2018, 2019; Di Rienzo et al., 2022; Di Rienzo, Miotto, Desantis, et al., 2023; Miotto, Di Rienzo, Gosti, et al., 2021; Piacentini et al., 2022; Venkatraman et al., 2009).

Therefore, the present dissertation reports the results of 3 experimental works in which the protein-protein interactions were investigated theoretically by means of 2D Zernike models and experimentally by means of bilayer interferometry and turbidimetry techniques.

Bilayer interferometry is a cutting-edge analytical method utilized in biopharmaceutical and life science research. It is a label-free technique that measures molecular interactions in real-time by monitoring changes in the interference pattern of light as it passes through a thin bio-layer immobilized on a sensor surface. The technique provides quantitative data on binding kinetics, affinity, and concentration of biomolecules, such as proteins, nucleic acids, and small molecules. The interaction process is visualized as a shift in the interference pattern, enabling researchers to determine binding constants and characterize binding mechanisms. It is a novel technique that finds applications in drug discovery, antibody characterization, and biomolecular interaction studies.

Turbidimetry is an analytical technique used to measure the cloudiness or turbidity of a liquid sample caused by the presence of suspended particles. As light passes through the sample, these particles scatter and absorb light, reducing its transmission. Turbidimetry quantifies this reduction in transmitted light intensity and correlates it with particle concentration or size. The tech-

nique, particularly when enhanced by nanoparticles, finds applications in various fields, including environmental monitoring and analytical determinations in food and beverage industries, but it is still rarely employed in pharmaceutical and biochemical fields.

In **Chapter 2** one can find a more detailed exposition of the physical principles behind these experimental methodologies.

1.1.1 RBD-ACE2 interaction

A paramount example of a PPI is the complex formation between the SARS-CoV-2 Receptor-Binding Domain (RBD) of the S spike protein and the human angiotensin-converting enzyme 2 (ACE2) receptor protein.

The S protein consists of an extracellular N-terminus, a transmembrane (TM) domain anchored in the viral membrane, and a short intracellular C-terminal segment. The total size is between 180 and 200 kDa and it forms trimers that surround the viral particle. **Figure 1.1** depicts a schematic illustration of the SARS-CoV-2 S protein internalization process. The S protein normally exists in a metastable, prefusion conformation; once the virus interacts with the host cell, extensive structural rearrangement of the S protein occurs and the virus fuses with the host cell membrane. The spikes are coated with polysaccharide that help the molecules to evade the host immune system.

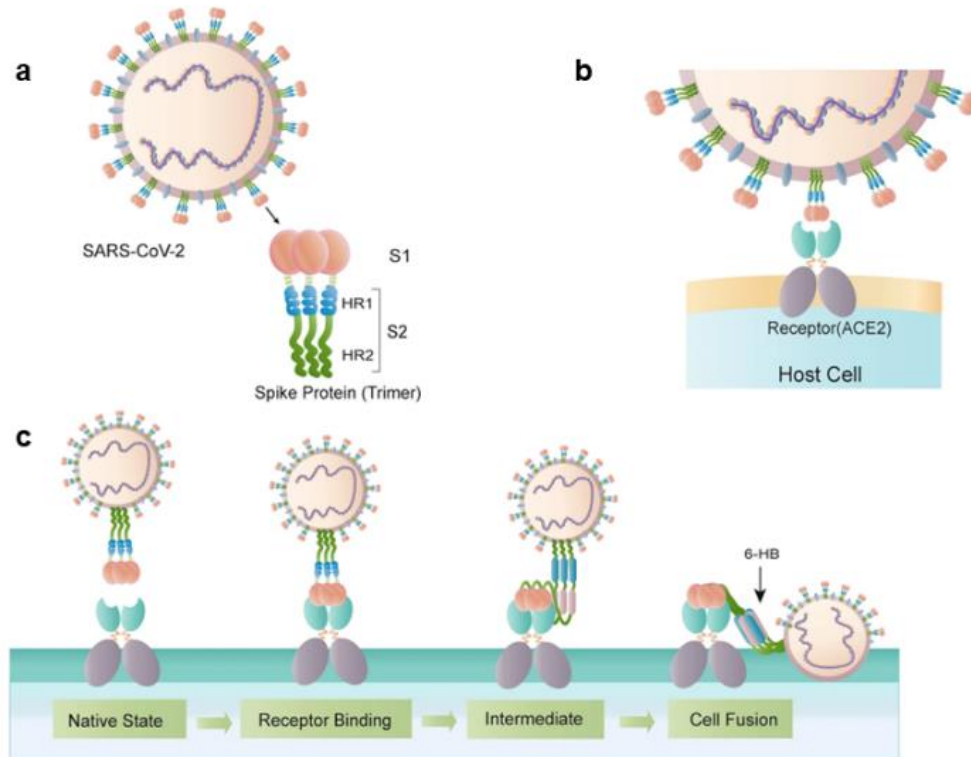


Figure 1.1. **a.** The schematic structure of the S protein. **b.** The S protein binds to the receptor ACE2. **c.** The binding and virus–cell fusion process mediated by the S protein. Figure adapted from (Huang et al., 2020).

The total length of SARS-CoV-2 S protein is 1273 a.a. and it is composed of two main subunits, an amino (N)-terminal S1 subunit and a carboxyl (C)-terminal S2 subunit, that are cleaved at the furin cleavage site (S1/S2 cleavage region). A detailed molecular structure of the spike gene with highlighted regions and locations is illustrated in **Figure 1.2**. The S1 subunit interacts with the ACE2 receptor subsequently promoting the viral infection: it switches between a “standing up” and a “closed down” RBD position, the former enabling ACE2 binding, and it undergoes spontaneous conformational transitions between the two forms. In contrast with the buried S2 subunit, the S1

subunit domains are located on the surface of the spike glycoprotein, effectively protecting the fusion apparatus.

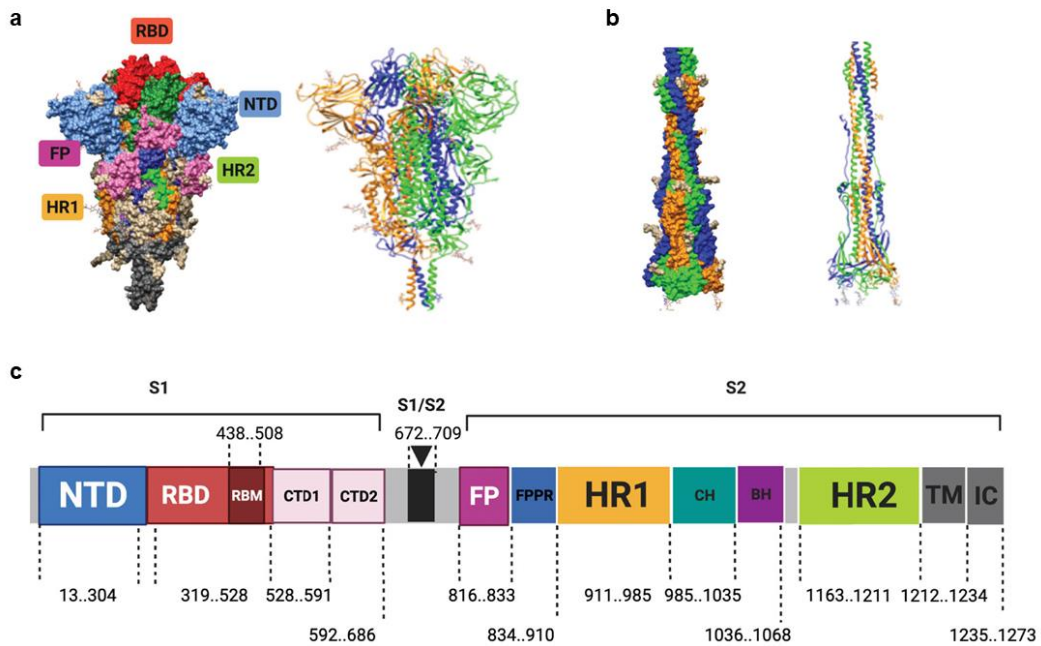


Figure 1.2. The structure of the spike glycoprotein and its coding gene. **a.** Schematic surface view of the spike ectodomain and its constituting domains and regions. **b.** Illustration of the structure of the native spike glycoprotein in its pre-fusion and post-fusion states. **c.** Detailed structure of the spike gene with the locations of the different regions is shown. The sequences of the two subunits S1 and S2 are highlighted. The S1 subunit includes the N-terminal domain (NTD), the receptor binding domain (RBD) which carries the receptor-binding motif (RBM), and two structurally conserved subdomains (the C-terminal domains 1 (CTD1) and the C-terminal domains 2 (CTD2)). The S2 subunit contains the N-terminal hydrophobic fusion peptide (FP), the fusion peptide proximal region (FPPR), the heptad repeat 1 (HR1) motif, the central helix region (CH), the β -hairpin region, the connector domain (CD) and the heptad repeat 2 (HR2) motif. These are followed by the transmembrane region (TM) and the intracellular regions (IC). Figure is adapted from (Hamdy et al., 2022).

The RBD of the S spike protein of SARS-CoV-2 is a critical component involved in the virus's ability to infect host cells. It is located on the outer surface of the spike protein (see **Figure 1.2a**) and structurally it is a compact, independently folded region that consists of about 193 amino acids. Its structure can be described as a globular domain. It has a twisted five-stranded antiparallel β sheet ($\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$ and $\beta 7$) with short connecting helices and loops that form the core. Between the $\beta 4$ and $\beta 7$ strands in the core, there is an extended insertion containing the short $\beta 5$ and $\beta 6$ strands, $\alpha 4$ and $\alpha 5$ helices and loops. The specific motif that interacts with human ACE2 receptor is the Receptor-Binding Motif (RBM): it comprises a cluster of amino acids that directly interact with the host cell receptor. **Figure 1.3** shows the structure of the complex RBD-ACE2. The RBD contains multiple disulfide bonds that help maintain its structural integrity. These covalent bonds between sulfur atoms in cysteine residues stabilize the overall structure of the RBD. When expressed in eukaryotic cells, the protein is glycosylated. Glycans attached to specific amino acids in the RBD can play a role in shielding the RBD from the host immune system and modulating its interaction with the ACE2 receptor.

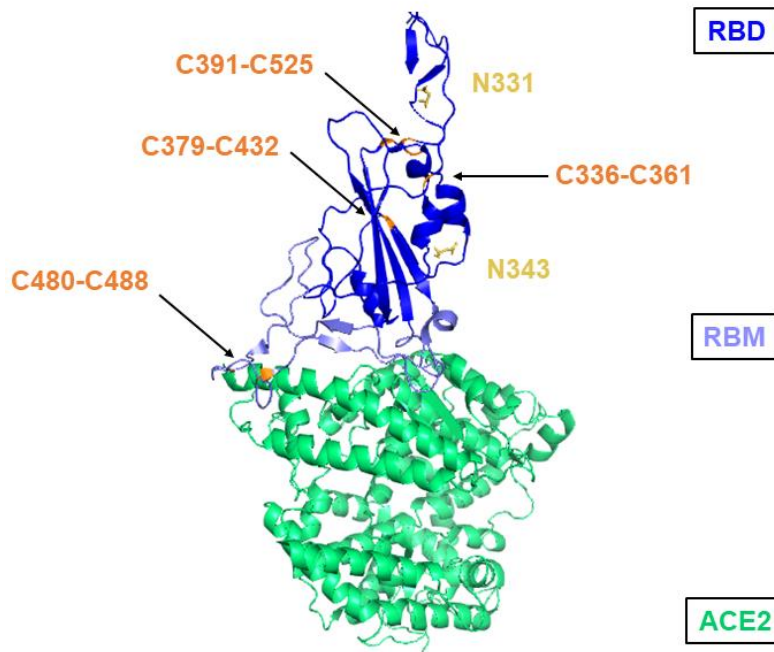


Figure 1.3. Complex structure of RBD interacting with ACE2 receptor. RBD is represented in blue ribbons; the receptor-binding motif belonging to the domain is colored in light blue; ACE2 receptor is in green. The glycosylation sites present in RBD structure are highlighted in yellow whereas the disulfide bonds are pointed by arrows and are labeled in orange.

The human angiotensin-converting enzyme 2 receptor is a crucial protein in the renin-angiotensin-aldosterone system (RAAS) and plays a significant role in regulating blood pressure and cardiovascular function. It is a membrane-bound protein, meaning it is anchored to the cell membrane of various human cells. It is typically found on the surface of cells in tissues like the lungs, heart, kidneys, and intestines. A major feature of the ACE2 structure is a deep channel on the top of the molecule that contains the catalytic site (see **Figure 1.4**). The channel is surrounded by ridges containing loops, helices, and a portion of a β -sheet. The long loop between N210 and Q221 is on the ACE2 surface. Potential N-glycosylation sites were identified at six positions: 53, 90, 103, 322, 432, and 546.

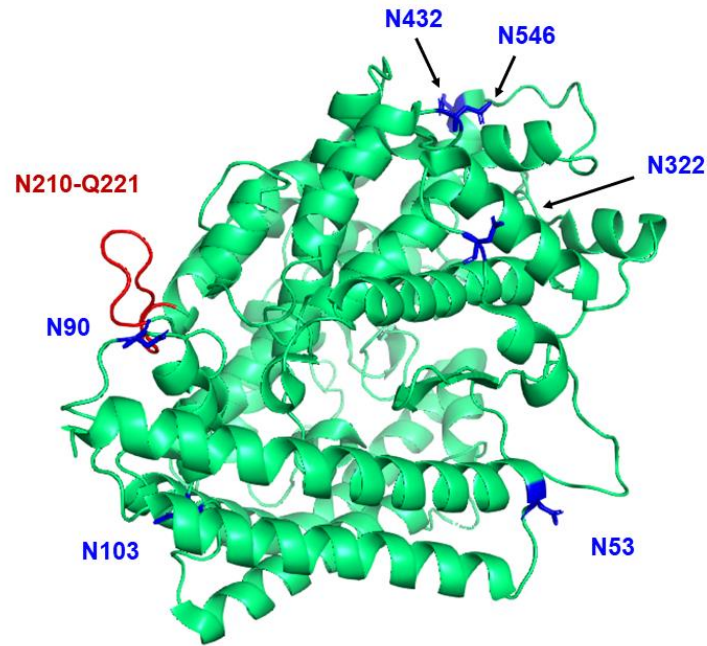


Figure 1.4. Top view of ACE2 protein with potential N-glycosylation sites highlighted in blue.

This interaction is crucial for viral infection, as it triggers a cascade of events ultimately leading to viral entry into the host cell (Walls et al., 2020). Also in this case, mutations in the RBD or ACE2 interface have been shown to affect virus infectivity and disease conditions (Barton et al., 2021). Considering the significance of this interaction, considerable efforts have been devoted to the development of computational methods for predicting mutation effects in the RBD-ACE2 interface (Bai et al., 2021; Miotto et al., 2022; Zou et al., 2020). To date, several experimental structures of the complex between ACE2 and the receptor-binding domain of the SARS-CoV-2 Spike protein have been determined, providing the structural basis for the specific interaction mechanism and highlighting the critical residues involved in the complex formation (Shang et al., 2020). Interestingly, it results that peptide "IEEQAK-TFLDKFNHEAEDLFYQSSLASWNYNTN" (residues 21-53) mimics the

major interacting portion of the ACE2 receptor to the SARS-CoV-2 Spike Receptor Binding Domain (RBD). Therefore, it is not surprising that peptides of different lengths, including residues 24 to 53 of the ACE2 receptor, have commonly been reported to exhibit high affinity binding to various regions of the SARS-CoV-2 Spike protein (de Campos et al., 2021; Larue et al., 2021). Many of these peptides, derived from the N-terminal α -helix of ACE2, have been tested in response to the SARS-CoV-2 pandemic (Tzotzos, 2022). Since SARS-CoV-2 enters cells through the interaction between the Spike glycoprotein and ACE2 ectodomain, disrupting the Spike/ACE2 interaction represents a major target for preventing cell infection (Gheblawi et al., 2020; Pappageorgiou & Mohsin, 2020).

In this dissertation, this PPI is experimentally investigated with two different approaches. In **Chapter 3.1** the investigation was dedicated to the direct experimental measurement of the computationally predicted inhibitory power of lactoferrin on the formation of the ACE2/RBD complex, which is manifested in its binding to the ACE2 receptor rather than to the RBD domain of the spike protein. The experimental methodologies applied are the previously introduced BLI and turbidimetry; the results demonstrate not only the validity of the *in silico* prediction, but also the potentiality of the two methodologies. The focus of **Chapter 3.2** is devoted to the exploration of the α -helix peptide derived from the N-terminus of the ACE2 sequence, specifically encompassing residues 21-43, in its interactions with the RBD protein. Upon application of the protocol design based on the Zernike model (extensively discussed in (Di Rienzo et al., 2021)), a set of five ACE2-derived peptide mutants, four of them endowed with predicted higher affinity for the RBD protein with respect to the WT peptide and one with predicted lower affinity, was generated. The peptides were subsequently synthesized and subjected to *in vitro* meas-

urements using BLI technique. These analyses experimentally confirmed the binding affinity between the peptides and the RBD protein.

1.2 Lactoferrin protein

Lactoferrins belong to the “transferrin superfamily”, a group of well conserved single-chain, glycosylated proteins that transport iron from plasma to cells or contribute to the regulation of iron transport in biological fluids (L. A. Lambert, 2012). Most members of the superfamily display a similar fold and consist of two homologous lobes (N- and C-lobe) connected by a short hinge region. Each of the two lobes is able to reversibly bind a single ferric ion. It is of interest to comment on the evolutionary path that underline this superfamily as follows. Early metazoans have been proposed to develop single-lobe iron binding proteins able to uptake iron from sea water. Subsequent gene duplication and fusion led to the evolution of the primordial single lobe iron binding protein into the bi-lobed mammalian transferrins (Tf) and lactoferrins (Lf) (L. A. Lambert et al., 2005; Park et al., 1985). In this framework, Lambert et al., proposed that Lf arose through a second gene duplication event in the mammalian lineage that occurred around 125 million years ago (L. A. Lambert, 2012; L. A. Lambert et al., 2005). In mammals, Lfs are expressed by epithelial cells in diverse tissues in mammals and are found in virtually all bodily exocrine secretions comprising colostrum and milk, tears, nasal and bronchial secretions and saliva (Masson et al., 1966). Furthermore, lactoferrins are also produced by the hematopoietic tissue of bone marrow and are found in granules of polymorphonuclear neutrophils (Berlov et al., 2007). Lower concentrations of lactoferrins are found in other exocrine secretions (Masson et al., 1966).

From the structural point of view, Lactoferrins are monomeric glycoproteins of about 690 amino acids (average molecular mass of ~80 kDa), and their three-dimensional structures have been reported for five species comprising human (Anderson et al., 1987, 1989; Haridas et al., 1995), cow (E. N. Baker et al., 1994; Moore et al., 1997), camel (Khan et al., 2001), buffalo (Karthikeyan et al., 1999, 2000), and horse (Sharma et al., 1999), all of them with very high structural similarity among species. The single polypeptide chain of lactoferrins is folded into two globular lobes sharing ~40% sequence homology and referred to as the N-lobe (amino acids 1–333) and C-lobe (amino acids 345–691, in human). Each lobe can be further divided into two domains, namely the N1 domain (amino acids 1–90, 251–333) and the N2 domain (amino acids 91–250), the C1 domain (amino acids 345–431, 593–689), and the C2 domain (amino acids 432–592). The two lobes are connected by a 10 amino acids long (334 to 344 in human), three-turn α -helix segment (see **Figure 1.5**). Each pair of N1 and N2, and C1 and C2 domains harbor an iron binding site. Four amino acids (two Tyr, Asp, His) are the main ligands of Fe^{3+} in each lobe. In addition, two oxygens from the synergistic carbonate anion complete the architecture of the binding site, accompanied by a portion of the N-terminus and an arginine side chain that contribute to the electrostatic coordination of the carbonate itself. Each lactoferrin molecule can, in principle, bind reversibly two iron ions, although the C-lobe site has an intrinsically much higher affinity for free ferric ions under physiological conditions (zeptomolar range, estimated $K_D \approx 10^{-22}$ M) due to a very slow kinetics of metal release (E. N. Baker & Baker, 2005; H. M. Baker & Baker, 2004; Mason et al., 2005). Conformational changes have been described in lactoferrins upon iron binding, indicating that iron-saturated species appear to

adopt a more closed structure with respect to the iron free species (Anderson et al., 1989; E. N. Baker & Baker, 2005).

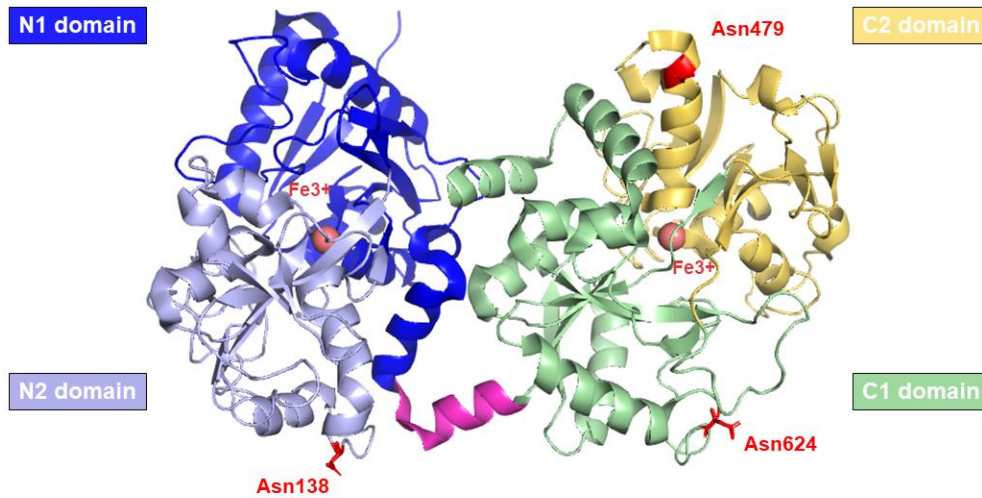


Figure 1.5. Structure of human holo-lactoferrin. In purple and light blue are highlighted the N1 domain (amino acids 1–90, 251–333) and the N2 domain (amino acids 91–250), respectively; C1 (amino acids 345–431, 593–689) and C2 (amino acids 432–592) domains are in light green and yellow, respectively; in pink, α -helix that joins the two lobes (amino acids 334–344); in red, the three potential N-glycosylation sites are indicated.

Lactoferrins are heterogeneously glycosylated with the number of glycosylation sites varying among species and, within a single species, can vary depending on the tissue in which the protein is expressed (Karav et al., 2017; Zlatina & Galuska, 2021). So far, human lactoferrin has three potential N-glycosylation sites comprising Asn138, Asn479, and Asn624, although the site at 479 is found to be glycosylated only in 5% of the mature proteins in body fluids (Berkel et al., 1996; Van Veen et al., 2004). In contrast, bovine lactoferrin displays five potential glycosylation sites, namely Asn233, Asn281, Asn368, Asn476, and Asn545, whose populations among the molecules within milk also shows heterogeneity (Wei et al., 2000; Ye et al., 1997).

The N-linked oligosaccharides belong to the class of high-mannose glycans, characterized by a N-acetylglucosamine (GlcNAc) core modified with mannose (Man) residues. By contrast, in the case of complex N-glycans, the common core pentasaccharide, consisting of Man₃GlcNAc₂, is elongated with GlcNAc at the α 1,3- and α 1,6-linked mannose residues. Additional monosaccharides, such as galactose, fucose and sialic acids, can also be added to complex N-glycans. The oligosaccharide moieties of Lf have been demonstrated modulatory effects on the biological functions of lactoferrin, besides the pivotal, intracellular, role in protein folding, oligomerization, quality control, sorting, and transport. N-glycans, in particular, have been shown to enhance iron-binding (Legrand et al., 1990; Li & Furmanski, 1995) and affect the anti-adhesive capacities of Lf towards bacterial or viral species (Barboza et al., 2012). The presence of sialic acid moieties within the oligosaccharide skeleton also appears to play a significant role in molecular recognition and protein stability.

As a whole, the body of high-resolution structural data currently available are providing the first clues on the molecular targets of Lfs that could explain some of the key biological activities reported so far. The first “iron dependent” molecular target of Lf has been identified in the bacterial protein Lactoferrin binding protein B (LbpB) within *Neisseria* pathogenic species (Yadav et al., 2021). The interaction between Lf and LbpB has been characterized in detail by high resolution crystallographic methods as shown in **Figure 1.6**; a set of “hot spot” have been identified on the molecular surface of Human Lf that summarize the current knowledge of Lfs interacting epitopes.

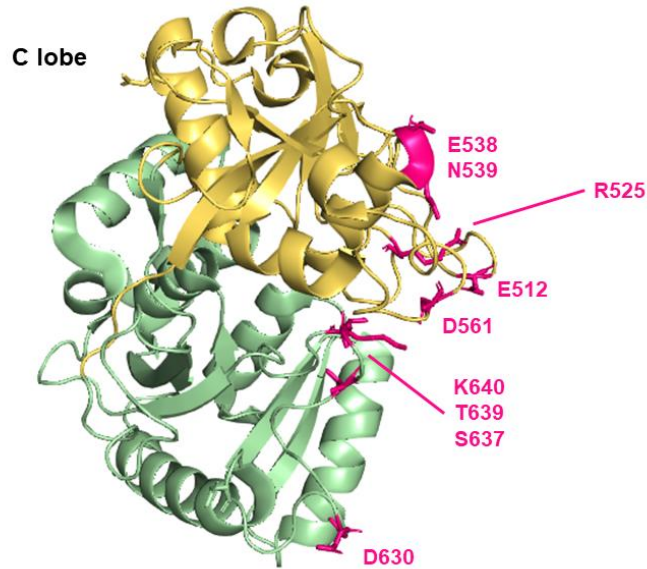


Figure 1.6. Side view of the C lobe of lactoferrin. The residues involved in the binding with *N. meningitidis* LbpB (NmLbpB) are highlighted in pink (PDB: 7JRD).

A few iron-independent pathways that lead to antimicrobial and antiviral activities by direct interaction with bacterial membranes or viral capsid epitopes or by modulation of host immunity will be dealt in **Chapters 3.1** and **3.2** In this framework, lactoferrins also impact immune homeostasis and modulate inflammatory response, that are likely to contribute to the host-parasite interaction. Again, these activities could be further discussed in the framework of iron-binding ability through the modulation of oxidative stress caused by reactive oxygen species within the inflammatory response or by modulation effect on the innate and adaptive immune responses by maturation and differentiation of immune cells.

Novel potential Lfs activities have been demonstrated recently in connection with reported anticancer properties. In this case, several targets have been proposed that are linked to cancer cell growth; induction of apoptosis; and

inhibition of cancer cell migration and invasion thus leading to confinement of metastasis. These biological activities have been suggested to be mediated by the recognition of specific receptors in human cells from different tissues belonging to the intelectin family (hIntL-1). The apparent affinity that has been measured for the hIntL-1 trimer is rather weak for a specific protein-protein interaction (K_D of ~ 500 nM). These initial data are however not yet consistent with a central role for lactoferrin–intelectin complexes in mediating the many functional roles for this interaction and should be further explored (Kowalczyk et al., 2022; Suzuki et al., 2005; Wesener et al., 2015). In this framework, it is clear that all biological activities in which Lfs are involved must be accounted for by specific PPIs with appropriate target proteins. To date, detailed structural information on such interactions have been demonstrated experimentally only in a few cases.

1.3 Multivalency of human ferritin

In the following **Chapters 3.2** and **3.3** the role of the human ferritin protein (Ft) in the molecular interactions at the center of the experimental studies exposed in this dissertation will be highlighted.

Ferritin is an iron storage protein that plays a key role in iron homeostasis and antioxidation of cells. It is a globular protein characterized by a typical tetraeicosameric assembly. Each subunit has a molecular weight of ≈ 20 kDa, while the assembled 24-mer has a total weight of ≈ 500 kDa; its hollow cage-like spherical structure has an outer diameter of roughly 12 nm and an inner diameter of 8 nm, with a thickness of 2-2.5 nm. **Figure 1.7** displays the outer and inner diameters.

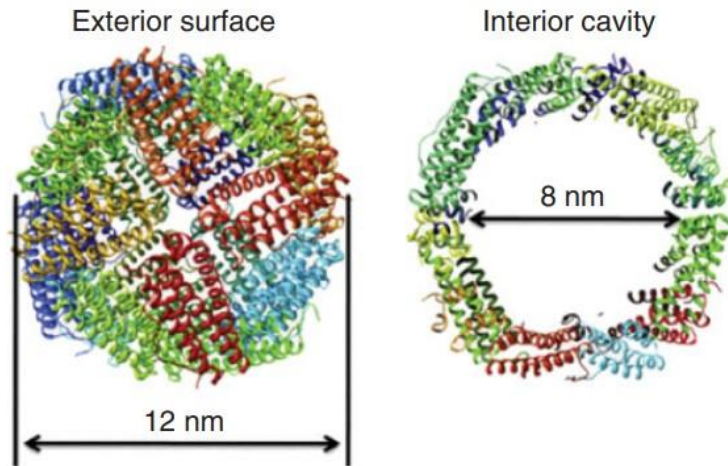


Figure 1.7. Ribbon diagrams of exterior surface view and interior cavity of human heavy chain ferritin. Picture from (Uchida et al., 2010).

The conformation of each subunit consists of a 4-helix bundle of alpha helices (see **Figure 1.8**): A (in humans, residues 10-39), B (45-72), C (92-120) and D (124-155); a fifth helix, E (res 160-169), lies at an acute angle to the bundle, pointing towards the internal cavity of the three-dimensional protein. A long loop L (res 73-91) connects B and C helices.

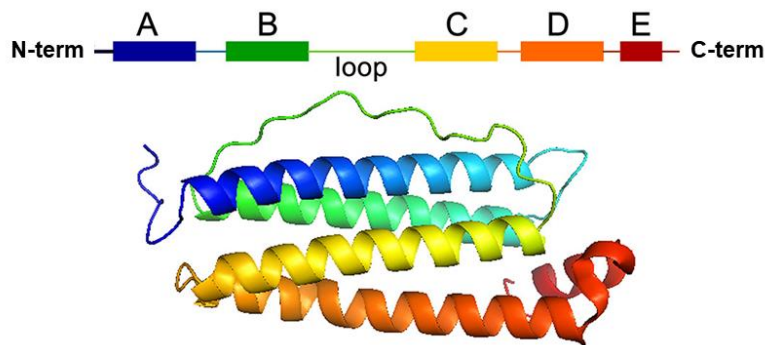


Figure 1.8. Ribbon diagram showing the tertiary structure of H-subunit of ferritin. The four α helices forming a bundle and a short C-terminal α helix are shown in different colors. Figure adapted from (Honarmand Ebrahimi et al., 2015).

In the self-assembly of the 24-meric shell, the first stable intermediate appears to be a dimer (Stefanini et al., 1987) through the interaction of the respective BC loops (deletion of two residues in the loop has been proved to abolish the formation of the whole 24-mer (Levi et al., 1989)).

The kinetics pathway from dimers to a complete 24-mer shell remains relatively unexplored. The disassembly in dimers is only partially reversible and occurs spontaneously at very acidic or basic pHs. Under neutral conditions, the inner surface of ferritin has a high negative charge density because acidic residues such as Glu and Asp are mainly distributed here, while the net charge on the exterior surface is close to zero or slightly positive.

Eukaryotes typically have two ferritin genes encoding the immunoglobulin heavy (H) (21 kDa) and light (L) (19 kDa) chains, which co-assemble to form heteropolymers under physiological conditions. The main function of H-chain is the catalyzation of Fe^{2+} to its oxidized state Fe^{3+} (Harrison & Arosio, 1996), which is enabled by its ferroxidase center. However, the L-chain lacks this center and catalytic ability, which therefore assists iron nucleation (Levi et al., 1994). However, to date, a putative LFt receptor on human cancer cells has not yet been identified, and whether or not LFt has the ability to site specifically delivery therapeutic agents to tumor is questionable (Fan et al., 2017).

Although ferritin has been well studied accurately in relation to its physiological properties and mineral core formation, its potential interest as a drug delivery vehicle has recently gained most attention. Its self-assembly ability, symmetrical spherical architecture, and high thermal stability are important aspects driving interest in ferritin nanocarriers. Furthermore, as an endogenous protein, ferritin also possesses excellent biocompatibility, biodegradability, and low toxicity, which are highly desirable features for nanocarriers in

clinical use. Due to its 8-nm diameter inner cavity, ferritin has the potential space to encapsulate many drug molecules, thus offering protection from degradation as well as limiting potential side effects to healthy cells. In addition, the sensitivity of the stable cage-like structure of ferritin to pH facilitates the application of various drug-loading methods. For example, under extreme environments, such as strong acidic pH, the quaternary structure of ferritin disassembles but, interestingly, reassembles once pH returns to physiological conditions. Thus, by manipulating the disassembly and reassembly of ferritin, it is possible to encapsulate therapeutic drugs inside its structure.

As already introduced, in **Chapter 3.2** the binding affinity of a set of five mutated peptide sequences was investigated, all derived from the wild-type sequences of the α -helix of the binding site of ACE2 receptor with RBD spike protein. As a proof of concept, the wild-type ACE2 peptide and the optimized peptide with the highest affinity for RBD were genetically fused at the N-terminus of human H-chain ferritin. The multivalent potency of ferritin, already vastly exploited in the field of drug delivery (Palombarini et al., 2020; Song et al., 2021) has been recently investigated in the context of SARS-CoV-2 treatment or vaccine development (Kalathiya et al., 2021; Khoshnejad et al., 2018; J. Kim et al., 2017; S. A. Kim et al., 2022). The fusion strategy applied with the mutated ACE2 peptides aimed to address the limitations associated with the therapeutic use of free peptides, including their susceptibility to rapid renal clearance and reduced bioavailability, while also benefiting from the multivalent effect conferred by the 24-meric ferritin structure. In fact, nanoparticles (NPs) have demonstrated their potential as conjugate scaffolds, enhancing peptide functionality and leveraging their intrinsic properties, often leading to synergistic effects (Jeong et al., 2018). Protein-based nanoparticles are particularly suitable, given their biocompatibil-

ity, ease of producing monodisperse forms through recombinant techniques, and the ability to be modulated via genetic engineering approaches. Among these protein nanoparticles, ferritin serves as an ideal platform for diagnostic and therapeutic applications (Affatigato et al., 2023; Calisti et al., 2018; Palombarini et al., 2020). Modifying a single subunit enables the functionalization of the entire nanoparticle. In the case of ACE2-derived peptides, fusing the peptide sequence to the N-terminus of human H ferritin offers the advantage of generating a chimeric protein that presents 24 peptides on its surface. This arrangement leads to a multivalent effect while concurrently reducing peptide clearance.

1.4 Inhibition of PCSK9 pathway

In **Chapter 3.3** of the present dissertation experimental results that again involve the multivalence property of human H ferritin protein are exposed and discussed.

In this experimental work, a human H ferritin variant displaying multiple copies of a Proprotein Convertase Subtilisin/Kexin type 9 (PCSK9) binding domain on its surface has been developed. PCSK9 is a kexin (that is, an endopeptidase) found in kidneys, liver, and intestine and it is a protein that plays a critical role in the regulation of cholesterol metabolism. Cholesterol is an important molecule for normal cell function and serves as a precursor for steroid hormones and bile acids. Indeed, it plays a crucial role in building cell membranes, producing hormones (like estrogen and testosterone), and aiding in the digestion of fats. Cholesterol is transported in the bloodstream by lipoproteins, primarily low-density lipoprotein (LDL) and high-density lipoprotein (HDL). High cholesterol levels, specifically high LDL cholesterol, can

lead to the buildup of plaque in arteries, increasing the risk of atherosclerosis, heart disease, and stroke. Conversely, excessively low levels of cholesterol, especially low HDL cholesterol, may raise the risk of mental health issues, hormonal imbalances, and impaired fat-soluble vitamin absorption. Therefore, maintaining a balanced cholesterol level is essential for overall health. Hypercholesterolemia is an example of a condition in which the levels of cholesterol are high in the blood, may be caused of diet, obesity, inherited (genetic) diseases (such as LDL receptor mutations in familial hypercholesterolemia), or the presence of other diseases such as type 2 diabetes and an underactive thyroid.

PCSK9 is a member of the proprotein convertase family and is predominately synthesized in the liver as a ~ 73 kDa zymogen consisting of a signal peptide (M1-A30), prodomain (Q31-Q152), catalytic domain (S153-Q454) and C-terminal domain (L455-Q692) (see **Figure 1.9**). After expression and cleavage of the signal peptide, PCSK9 undergoes intracellular autocatalytic processing between Q152-S153 of the prodomain. Although cleaved, the prodomain remains non-covalently bound to the catalytic domain, preventing further catalytic activity by blocking the catalytic site. The mature PCSK9 heterodimeric complex is necessary for PCSK9's cholesterol-regulatory function as it stabilizes the active structure for efficient LDLR interaction and likely assists with secreting PCSK9 into the extracellular space.

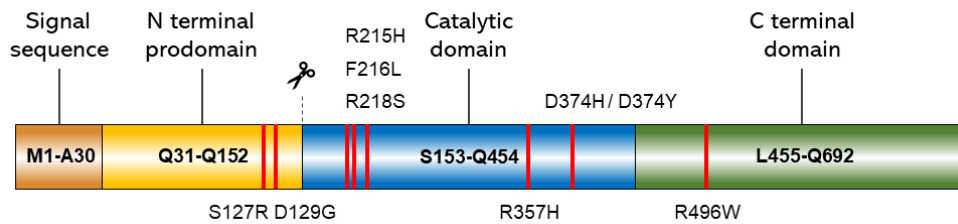


Figure 1.9. Sequence of the different domains of PCSK9. PCSK9 consists of a signal peptide (orange), N-terminal prodomain (yellow), catalytic domain (blue) and C-terminal domain (green). Autocatalytic cleavage between Q152–S153 results in the prodomain non-covalently bound to the catalytic domain. The naturally occurring mutants of PCSK9 that are associated with familial hypercholesterolemia are highlighted in their respective domains. Picture adapted from (Tombling et al., 2021).

Several separate crystal structures of full length, wild-type PCSK9 have recently been published (Cunningham et al., 2007; Hampton et al., 2007). The overall domain structure of PCSK9 is similar to other PC family members. It includes a signal peptide, followed by a prodomain, a subtilisin-like catalytic domain, and a variable C-terminal domain (termed V domain). The prodomain serves a dual role as a chaperone for folding and as an inhibitor of catalytic activity. Autocatalytic cleavage between Gln-152 and Ser-153 separates the prodomain from the catalytic domain, but the prodomain remains bound, occluding the catalytic site. For other PC family members, a second catalytic cleavage is required to release the prodomain, which unmarks the catalytic site, resulting in an active protease. No site of secondary cleavage has been identified that activates PCSK9.

The crystal structure of apo-PCSK9 revealed a tightly bound prodomain that is predicted to render the active site inaccessible to exogenous substrates. The structure of the PCSK9 prodomain and catalytic domain is similar to that of other subtilisin-like serine proteases. The V domain of PCSK9 consists of three subdomains of β -strands folded in a jelly roll motif, each stabilized by three internal disulfide bonds. The subdomains are arranged in quasi 3-fold

symmetry resulting in a barrel-like structure. The V domain of PCSK9 is rich in histidine amino acids, with the majority of them forming a patch on the surface of subdomain 2. The large number of histidines is particularly striking when one considers the 25–170-fold increase in binding affinity between PCSK9 and the LDLR as the experimental pH is dropped from neutral toward lysosomal pH, below the pKa of the histidine amino acid side chain (i.e., pH 6.0). This domain also shares structural homology to the adipokine resistin and has been speculated to mediate protein–protein interactions.

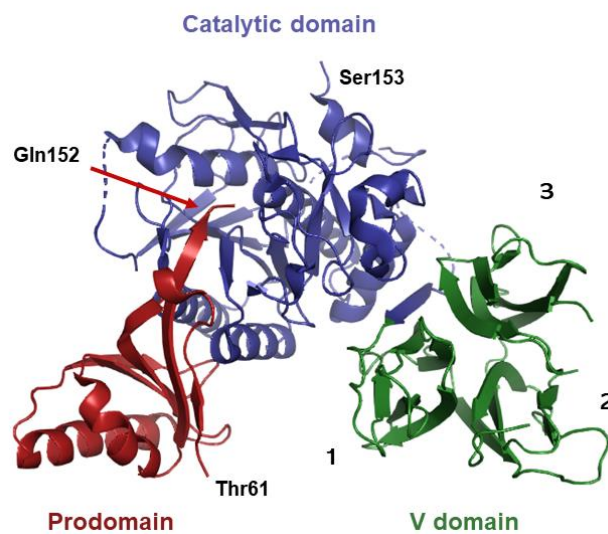


Figure 1.10. Ribbon structure of PCSK9. The prodomain is in red, the catalytic domain is in blue and the V domain is in green. Thr61 marks the first observed residue, and Gln152 marks the C terminus of the prodomain. Ser153 marks the N terminus of the catalytic domain. The subdomains of the V domain are numbered. Picture adapted from (G. Lambert et al., 2009).

Since PCSK9 is inactive when first synthesized, because a section of peptide chains blocks their activity, proprotein convertases are needed to remove that section to activate the enzyme. When activated, it acts to increase circulating levels of low-density lipoprotein cholesterol (LDL-C). Elevated PCSK9 lev-

els (or “gain of function” variants) are associated with an increased risk of cardiovascular disease and familial hypercholesterolemia. The mechanism of action of PCSK9 in hypercholesterolemia involves the binding of PCSK9 to the LDL receptor (LDLR) on the surface of liver cells. **Figure 1.11** displays the regular metabolic pathways that LDLR proteins undergo. Normally, the LDLR acts to remove LDL cholesterol from the bloodstream by specific uptake and transfer into the liver cells for breakdown and excretion. When LDL cholesterol is uptaken by the LDLR, both the LDL cholesterol and the LDLR are endocytosed. Once inside the cell, the LDLR and LDL cholesterol are separated; upon exposure to the acidic pH of the endosome, the LDL cholesterol is transported to the lysosome for degradation and the LDLR is recycled back to the cell surface to continue removing LDL cholesterol from the bloodstream. However, when PCSK9 binds to the LDLR, it causes a conformational change in the LDLR, marking it for lysosomal degradation, and preventing it from being recycled back to the cell surface. In fact, PCSK9 promotes the receptor degradation into the lysosomes and reduces its recycling on the cell surface (see **Figure 1.11** right panel). This, in turn, leads to an accumulation of LDL cholesterol in the bloodstream, which is a key risk factor for the development of cardiovascular disease.

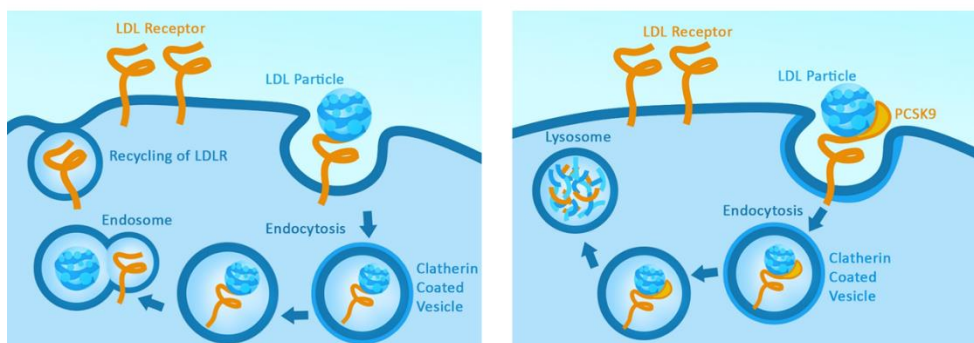


Figure 1.11. Schematic representation of how PCSK9 impacts LDLC metabolism by regulating LDLR levels. **Left panel.** ordinary metabolic pathway of the recycling of a LDLD receptor. **Right panel.** in hypercholesterolemic subjects where levels of PCSK9 are wrongly regulated, PCSK9 binds LDLR on hepatocyte surfaces, triggering their internalization and degradation by the lysosome.

Reducing LDL-C levels, most commonly by statins, has been shown to substantially reduce cardiovascular disease events. However, statin therapy alone is not adequate to lower LDL-C for patients with familial hypercholesterolemia.

In **Chapter 3.3** of this dissertation, the aim is to report experimental results given by the exploration of the potential of ferritin surface modification for recognizing PCSK9. The approach involved modifying the N-terminal sequence of the HfT subunits to incorporate a PCSK9-binding molecule (“PBM”) identified to be one of PCSK9 inhibitor (Ahamad & Bhat, 2022; Liu et al., 2022). The resulting HfT-PBM nanoparticles were characterized by size exclusion chromatography and electron microscopy to determine their size and morphology. The binding properties of the nanoparticles to PCSK9 were assessed by bilayer interferometry, measuring the kinetics and affinity of the interaction. The results obtained are very promising and relevant for the fact that they demonstrate the potential of human H ferritin modification for recognizing disease targets and its application as a therapeutic platform.

Chapter 2 Materials and methods

2.1 Biolayer interferometry

Biolayer Interferometry (BLI) is a label-free, optical detection technique that is becoming increasingly used in biochemistry and biophysics for studying macromolecular interactions. It provides real-time and quantitative information about binding kinetics, affinity, and concentration of biomolecules. It is based on the interference pattern of the wavefront of white light passing through a biosensor and hitting one or more layers of biomolecules that may interact on the tip of the biosensor.

In the following section, the physical principle and its application of this technique will be explained in depth.

2.1.1 Interference

The physical principle behind BLI is based on the interference of light waves. BLI instruments typically consist of an optical waveguide, such as a fiber optic, coated with a thin layer of biomolecules, referred to as a "biolayer". The biomolecules of interest can be proteins, antibodies, nucleic acids, small molecules, or other biological compounds. When a light beam, typically from a laser, is shone onto the waveguide, it travels through the waveguide and interacts with the biolayer on its surface. Some of the light is reflected back from the interface between the waveguide and the biolayer, while the rest is transmitted through the biolayer and is reflected from the interface between the biolayer and the surrounding medium (usually the buffer solution). These two reflected light waves recombine, leading to interference patterns that can be constructive or disruptive: when two light waves have same phase the in-

interference is constructive and the resulting wave is a positive combination (see **Figure 2.1**).

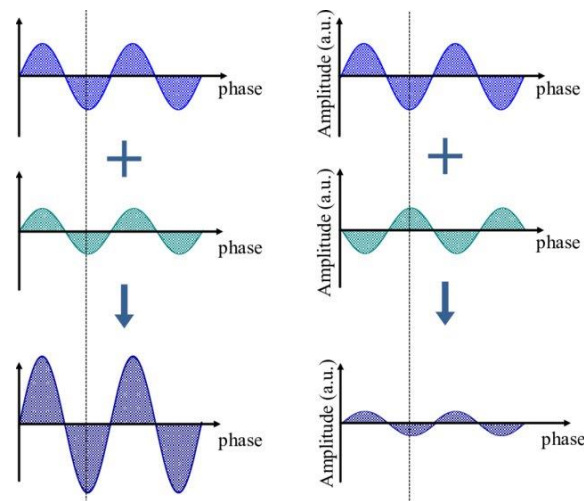


Figure 2.1. Constructive and destructive interference phenomenon.

The interference occurs due to the difference in the optical path length traveled by the two reflected light waves. This difference is influenced by the changes in the thickness and refractive index of the bilayer caused by the binding or dissociation of molecules on its surface. As the biomolecules bind to or dissociate from the bilayer, the thickness and refractive index change, leading to a shift in the interference pattern. By monitoring the changes in the interference pattern over time, BLI instruments can precisely measure the kinetics of molecular interactions, including association (binding) and dissociation (unbinding) rates, as well as equilibrium constants. This information provides insights into the strength and specificity of the binding interactions.

2.1.2 BLI applications

BLI has numerous applications in biochemistry. Some of the key applications include:

Protein-protein interactions: BLI can be used to study the binding interactions between proteins, such as enzyme-substrate interactions, antibody-antigen interactions, and protein-protein complex formations.

Protein-small molecule interactions: BLI enables the characterization of the binding affinities and kinetics between proteins and small molecules, such as drug candidates or inhibitors.

Antibody characterization: BLI is valuable in antibody development and characterization, allowing the measurement of antibody-antigen binding kinetics and affinity. It helps assess the specificity and potency of antibodies for diagnostic and therapeutic purposes.

Protein-nucleic acid interactions: BLI can investigate the interactions between proteins and nucleic acids, including DNA-protein or RNA-protein interactions, providing insights into gene regulation and protein function.

Membrane protein studies: BLI allows the investigation of membrane protein interactions with ligands or other proteins, aiding in drug discovery and understanding signal transduction processes.

Protein-lipid interactions: BLI can analyze the binding interactions between proteins and lipids, providing insights into membrane biology, lipid-protein interactions, and lipid-mediated signaling pathways.

2.1.3 Experimental setup for advanced kinetic assay

The BLI system utilized for most of the experiments is the Octet® N1 model "BLItz" manufactured by Sartorius (**Figure 2.2**).



Figure 2.2. Octet® N1 model.

The instrument is equipped with:

- a black 96-well plate for hydration of biosensors,
- magnetic microplates with 4µL capacity ("drop holder") for loading and binding steps,
- black plastic vials with 500µL capacity for baseline and dissociation steps,
- trays of 96 biosensors provided by Sartorius and chosen according to the biochemical requirements of individual experiments. Shown below is a list of the biosensors that have been used for the experiments. All the characteristics reported are provided by Sartorius' biosensor manual.

Biosensor	Description	Application	Dynamic Range
HIS2	Anti-HIS	Quantitation of HIS-tagged proteins in crude matrices or buffer or column eluent (pre-coated with anti-His Ab from MBS)	Protein dependent, typically 0.1–200 µg/mL

Ni-NTA	Nickel NTA	Quantitation of HIS-tagged proteins in buffer or diluted matrix, capturing of HIS-tagged proteins for kinetic analyses with various analytes	Protein dependent, typically 0.5–1000 $\mu\text{g/mL}$
ProA	Protein A	Quantitation of IgG's of various species including human	0.5–4000 $\mu\text{g/mL}$
SAX2	High Precision Streptavidin 2.0	Immobilizing biotinylated molecules for high precision and reproducible kinetic characterization and custom quantitation	Protein dependent

In **Figure 2.3** it is reported a graphic explanation of how the binding signal is detected and elaborated from the interferometer.

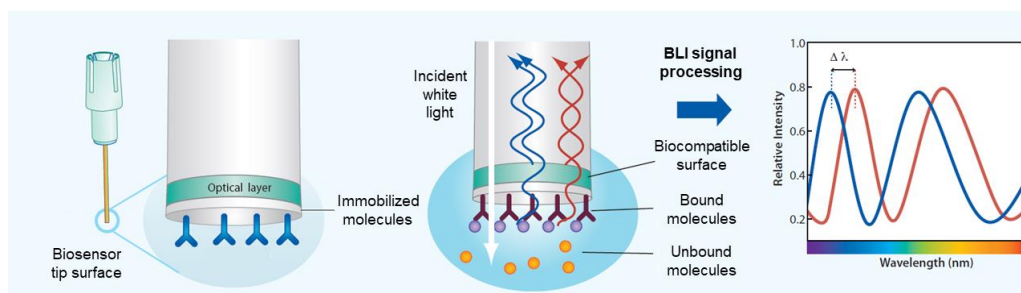


Figure 2.3. Schematics of the detection process of the binding signal. Picture is adapted from Sartorius manual.

The internal surface of the biosensor is semi-reflective, meaning that part of the incident light is reflected in back scattering (180°) to the detector. This

provides a reference signal which is always identical for all types of biosensors. The outer surface of the biosensor is coated with a layer of molecules or proteins upon which the biosensor tip is initiated. The choice of the specific biosensor type depends on the molecular system in question and the particular tag characteristic of each protein. Any alteration in the quantity of molecules adhered to the biosensor tip induces a modification in the interference pattern, which is detectable in real-time. The interaction between a ligand present on the biosensor tip's surface and an analyte within the solution leads to an increase in optical thickness at the tip. Consequently, this causes a shift in wavelength, directly indicative of the alteration in the biological layer's thickness. These interactions are continuously monitored, facilitating the precise and accurate observation of binding specificity, association and dissociation rates, and concentration. The Octet® System's interference pattern is only influenced by molecules that either bind to or dissociate from the biosensor, thus generating a distinctive response profile. Factors like unbound molecules, fluctuations in the refractive index of the surrounding medium, or changes in flow rate do not exert any impact on the interference pattern.

In the following **Figure 2.4** a schematic of a typical binding signal acquired with the instrument is shown. The assay usually is articulated into defined time steps and the experimental protocol is described.

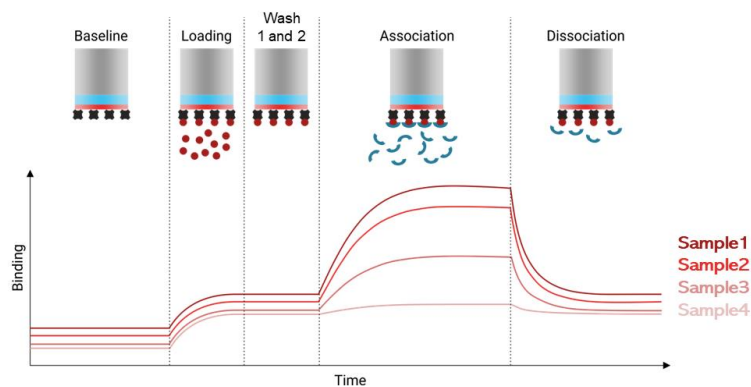


Figure 2.4. Example of time-course data acquisition in a BLI assay.

Before each acquisition, a new biosensor is left hydrating in kinetic buffer for a period of 10 min; this operation is needed in order to eliminate the saccharose in which the tips are coated when packed before usage. In a BLI assay, the initial step is a “baseline step” lasting 30 s. Then the first molecule is immobilized on the biosensor during the “loading step”; the loading process takes place in a drop holder with capacity 4 μ l. During all BLI experiments the concentration of the loaded protein has been maintained constant to 50 μ g/ml. Before the binding step, two washing steps (“wash 1 and 2”) occur. The washing is done in a black tube of volume 250 μ l containing the kinetic buffer provided by Sartorius (PBS with 0.02% Tween20, 0.1% BSA and 0.05% NaN₃). These steps provide baseline signals before the interaction with the analyte protein; also, between the two washing, the tube is emptied and filled with kinetic buffer again, so the biosensor tip already initialized with the receptor protein is put in a new environment cleaned of residual protein left unbound in the solution. After the washing there is the “association step” in which the initialized and cleaned biosensor is immersed in the solution containing the analyte protein (in another drop holed with capacity 4 μ l). The duration of this step varies depending on the velocity of the binding be-

tween the receptor and the analyte proteins. Finally, during the “dissociation step” it is possible to observe the decreasing of the signal when dissociation happens. This step occurs again in the black tube containing 250 μ l of kinetic buffer. After every assay, the equipment is cleaned with 0.5 M HCl solution. The shaker for the sample plate is set at 2200 rpm. All assays have been performed at room temperature (25°C).

The protocol for the BLI assay just described was common for all the experimental work exposed in the present thesis. The choice of biosensors, the concentration range of the analyte protein, the dilution buffer and other experimental characteristic that may differ will be specified in each chapter.

2.2 Nanoparticle-enhanced turbidimetry

Analysis based on turbidimetric phenomenon are being widely used with the intent to develop diagnostic test that are able to identify and quantify proteins present in biological fluids or solutions.

Turbidimetry has become a common tool in clinical diagnostics, since the technique is simple, low-cost and requires small amounts of materials. Moreover, in its most recent development (e.g., nanoparticle enhanced techniques) it presents several advantages, such as high sensibility to the analyte in solution, reduction of background noise caused by non-specific binding, and stability of nanospheres in solution.

In the following sections, the physical principle behind turbidimetric measurements and nanoparticles enhanced techniques will be discussed, and the protocol abided.

2.2.1 Physical principle

When the photons of the incident ray hit the sample whose particles are smaller than its wavelength, the diffusion occurs at the same frequency as the incident light, it is thus elastic and homogeneous (Rayleigh scattering), and the sample is spectrometrically "transparent". When aggregates start forming and the particles reach dimensions greater than the incident wavelength, anelastic diffusion becomes relevant and light transmitted along the incident direction decreases, therefore absorbance values raise. This scenario is described by the Mie model.

Light-scattering theory is dictated by the diameter, D , of the scattering elements in relation to the wavelength λ of the incident light. Indeed, theory is often specified in terms of the Mie size parameter $\alpha = 2\pi R/\lambda$, where R is particle radius. Depending on the particle size, two main situations can be distinguished.

Small particle size: for particles where $D < 0.05\lambda$, Rayleigh scattering theory is applicable to liquids with low concentrations of suspended particles which do not interact with each other. For such small particles, relatively symmetrical light-scattering distributions are obtained. If a visible light source (meaning λ between 400-700nm) is employed, the theory is applicable for particles where $D \leq 0.03\lambda$. The Rayleigh equation describing the angular distribution of resultant scattering is:

$$I = I_0 \left[\left(\frac{n'}{n} - 1 \right)^2 \right] \left(\frac{NV^2}{\lambda^4 R^2} \right) (1 + \cos^2 \theta)$$

where θ is the diffusion angle, n' and n are the refractive indexes of particles and suspension media respectively, R is the distance from the particles to the point of measurement in terms of N number of particles, each of volume V . The term $(1 + \cos^2\theta)$ provides the angular distribution of Rayleigh scattering. **Figure 2.5** represents a schematic illustration of how the angular distribution of diffused light is influenced by particle size.

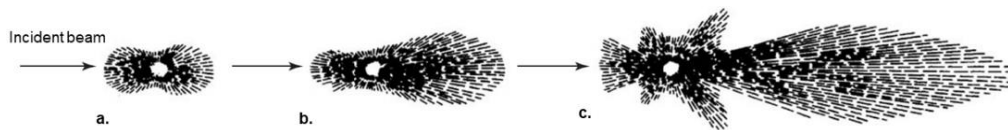


Figure 2.5. Influence of particle size on the angular distribution of scattered light. **a.** Small particles ($D < 0.1\lambda$); **b.** large particles ($D \sim 0.25\lambda$); and **c.** larger particles ($D > \lambda$).

Rayleigh's work thus shows that the intensity of scattered light varies with the square of the particle volume (and thus with the 6th power of the particle radius, assuming spherical shapes) and inversely with the 4th power of the light wavelength used. Rayleigh theory has since been developed to allow relative molecular masses and sizes to be determined.

Large particle size: for larger particles where $0.1\lambda < D < 0.8\lambda$, the angular distribution of scattered light becomes asymmetrical. Destructive interference of light scattered in the backward direction leads to a bias in forward-scattered light (**Figure 2.5b**). In these contexts, Mie scattering theory becomes more appropriate. For such larger particles, scattering intensity is less dependent on wavelength.

2.2.2 Nanoparticle-enhanced turbidimetric assay

Turbidimetric signals can be amplified by means of latex nanoparticles that allow the measurement of increase in turbidity (and therefore the absorbance intensity) of molecules that are too small to generate a scattering signal when freely dissolved in solution. In the current experimental set up, chloromethylated polystyrene nanospheres (provided by Ikerlat) with diameter of 103 nm and parking area of $95 \text{ \AA}^2/\text{group}$ were utilized. The following **Figure 2.6** illustrates the nanospheres as observed with transmission electron microscopy (TEM) and scanning electron microscopy (SEM).

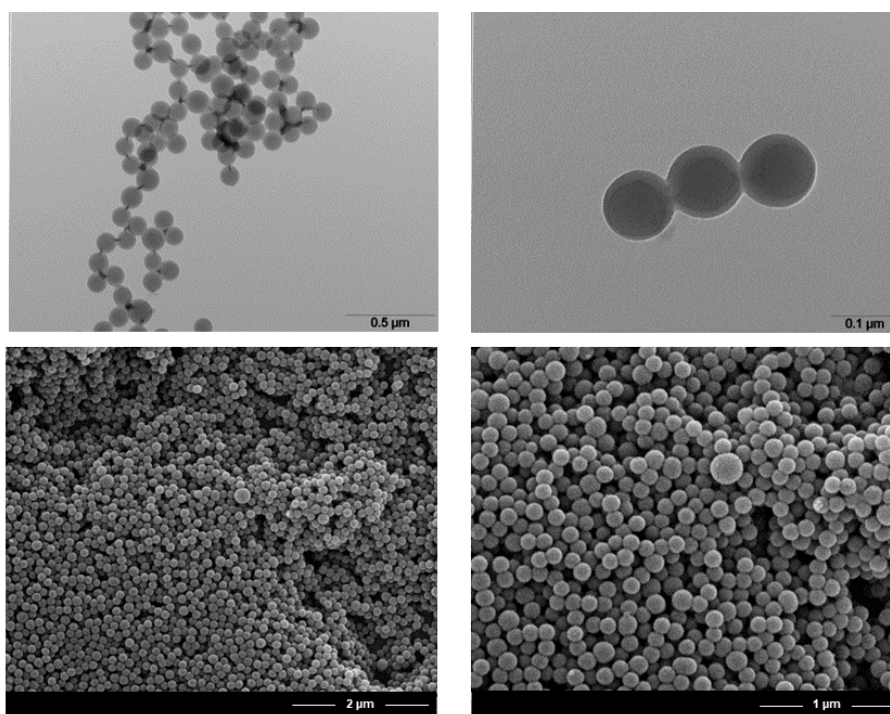


Figure 2.6. Top panels. TEM images of nanospheres with 100nm diameter. It is clearly noticeable the smoothness and homogeneity of the spheres' surface. **Bottom panels.** SEM images of a more concentrated solution.

TEM is a high-resolution microscopy technique that uses a focused beam of electrons to pass through thin samples, providing detailed information about the internal structure, composition, and morphology of materials. In SEM microscopy technique, electrons scan the sample's surface instead of going through it. It produces high-resolution, three-dimensional images of the sample's surface morphology, therefore it is commonly used to visualize the external features of samples, such as textures, topographies, and structural details at various scales. Polystyrene nanospheres must be functionalized with a specific binder, capable of coupling with the analyte with high affinity. Thus, it is necessary to link covalently the binder to the surface of the nanosphere, a process that is referred to as “coating”.

The process of the coating consists of covalently bond the molecule of interest to the nanosphere and letting the chloromethyl groups on the latex particle react with the ammine residues on the protein binder, with a reaction SN_2 as shown in **Figure 2.7**.

Then, nanoparticles functionalized with the protein are mixed in solution with the second analyte protein: the protein-protein interaction detected will be then observed between the free analyte protein in solution and the one bound to the nanospheres.

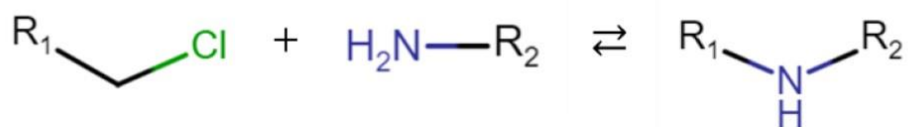


Figure 2.7. SN_2 reaction between chloromethyl group of the lattice nanosphere and the ammine group of the lateral chain on lysine.

When the interaction occurs, aggregates start forming and therefore the level of turbidity rises. This effect causes an increase in absorbance of the solution that is detectable with the spectrophotometer. The value of the absorbance is directly correlated not only to the concentration of a substance, but also to the scattering of the monochromatic light that hits the sample. By monitoring the change in absorbance, one is able to quantify the analyte protein and derive the thermodynamic constants of the reaction. **Figure 2.8** shows an example of data acquisition in the turbidimetric assays performed.

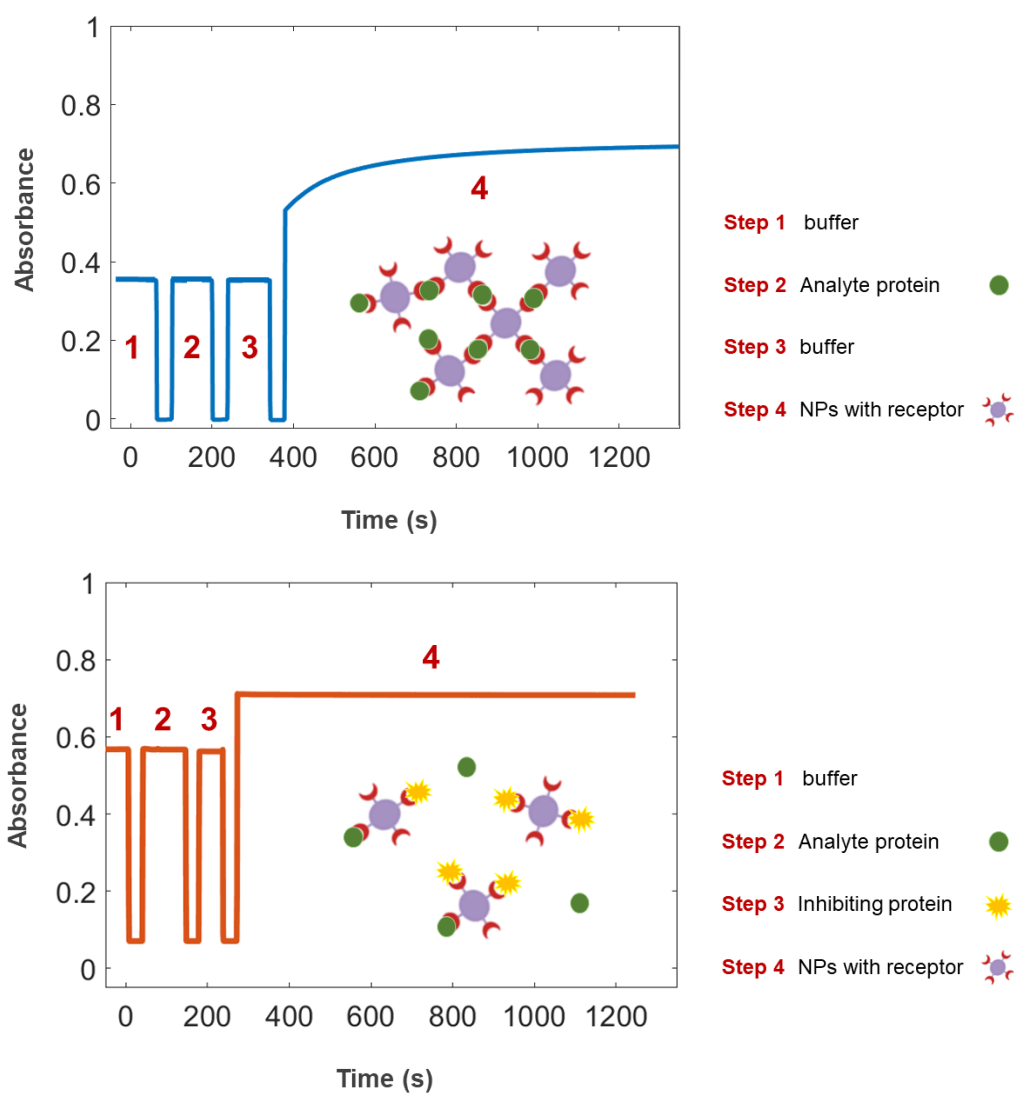


Figure 2.8. Examples of time-course acquisition in a turbidimetric assay in absence (**top panel**) and presence (**bottom panel**) of an inhibiting protein. Step 1 (lasting approximately 60 s) consists of acquisition of just the buffer signal. During Step 2 (lasting approximately 60 s) a suitable volume of the first protein is added, in order to monitor relevant absorbance signal caused by possible aggregation coming from the analyte protein alone (trimeric). In Step 3 the inhibiting protein is injected to monitor potential aggregation with analyte protein (monomeric). Finally in Step 4 the nanospheres coated with the receptor protein are added in the solution. During this final step absorbance increase due to possible interaction with the nanospheres is observable, since such interaction is translated in an increasing signal of

absorbance detected by the spectrophotometer. If enhancement is not detected, it is possible that the presence of the protein added in step 3 is interfering with the formation of analyte-receptor complex and therefore the aggregation of the latex nanoparticles.

The time-course acquisition of light absorption signals was conducted with a spectrophotometer (Jasco V-750 UV-Visible/NIR) at fixed wavelength (340 nm) and bandwidth (10 nm) for 1300 s. The choice of the incident wavelength depends on the dimension of the latex particle; as discussed in the physical principle paragraph, different ratios between the diameter of the particles and the excitation wavelength is what determine the phenomena of elastic and anelastic scattering of light, according to the Rayleigh model and the Mie model.

The application of latex nanoparticle-enhanced turbidimetry technique is exposed in **Chapter 3.1**, where the inhibiting role of lactoferrin in the formation of the complex RBD-ACE2 is investigated. In the experiments, the nanospheres are coated with ACE2 protein and analyte proteins in solutions are RBD or Lf. The structure of lactoferrin is well defined, therefore it is reasonable to assume that when in solution with coated nanospheres, multiple interactions can occur and aggregation phenomena such as those depicted in **Figure 2.8** (top panel) may ensue. In the case of RBD, an oligomerization of the protein is currently under observation. Observed results described in the chapter, indicate that the presence of RBD protein in the turbidimetric assay as the analyte protein yields aggregation, thus suggesting that the protein is capable of bridging individual coated nanoparticles. As reported previously, dimerization of RBD has been observed (Grishin et al., 2022; Lan et al., 2020; Sinegubova et al., 2021).

2.2.3 Experimental setup

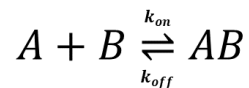
In this paragraph, the specifications for the experiment described in **Chapter 3.1** are collected.

The nanosphere stock concentration was 0.1 g/mL (10%). Before the assay, the nanospheres were coated with the receptor protein in 10 mM phosphate buffer (PB) at pH 7.4 in the case of ACE2 protein. A total of 2 mL of the solution of ACE2-coated nanospheres was obtained by adding 4 μ L of Ikerlat nanospheres to 1.92 mL of PB. Such an amount of nanospheres guarantees 0.02% of solids in the final solution. Then, 26 μ L of ACE2 (stock concentration 1.34 mg/mL) was added to the nanospheres solution (final ACE2 concentration 17.42 μ g/mL = 157.65 nM), and the solution was kept in gentle agitation on a tilting platform at room temperature for 3 h. Finally, 52 μ L of blocking buffer was added. Such buffer consisted of goat serum (Millipore) at a concentration of 20 mL/L and ProClin 300 biocide (Sigma-Aldrich) at 0.3 mL/L in PBS at pH 7.4, to fully block potentially unbound chloromethyl groups. This reaction was incubated again on a tilting surface, gently mixing overnight. In the case of the RBD protein, the coating buffer was 20 mM sodium carbonate at pH 9.2 containing 150 mM NaCl. For the coating step of nanospheres with RBD as the receptor protein, the procedure was the same but with different quantities: 1 μ L of Ikerlat nanospheres was added to 0.5 mL of bicarbonate and guaranteed 0.02% of solids in the final solution (the stock concentration was 0.1 g/mL). The concentration of RBD used for the coating procedure was 174.2 μ g/mL = 5.8 μ M. For spectra acquisition, polymethyl methacrylate (PMMA) cuvettes with 500 μ L capacity and 0.6 cm optical path were used. Nanoparticle aggregation results in an increasing signal of absorbance attributed to RBD–ACE2 complex formation and was analyzed in phenomenological terms as a saturation curve. Thus, a turbidimetric

assay was performed with nanospheres coated with fixed concentrations of either ACE2 or RBD, while Lf was present at different concentrations. The analysis was carried out in phosphate-buffered saline at pH 7.25 with 0.1% glycine, 0.1% NaN₃ and polyethylene glycol (PEG) 6000 at a concentration of 5%. The analysis of the turbidimetric data was performed by means of a MATLAB custom program.

2.3 Analytical model

BLI can be used to measure kinetic binding constants (k_{on} , k_{off}) and the affinity constant K_D . A simple 1:1 binding model of a 2nd order reversible reaction is described by the expression:



The association step is governed by a rate constant called the association rate constant, denoted by the parameter k_{on} . The binding is reversible and the rate of breakdown of the complex is governed by a rate constant termed the dissociation rate constant, denoted as k_{off} . The units of the association rate constant are $M^{-1} s^{-1}$ and of the dissociation rate constant s^{-1} . The larger the value of the association rate constant, the more rapid the binding event. The study of bimolecular reactions is usually carried out under pseudo-first order (PFO) conditions by keeping the initial concentration of one of the two reagents in large excess (50–100 fold) (Kumaraswamy & Tobias, 2015; Malatesta, 2005; Müller-Esparza et al., 2020). Under this approximation, a simple exponential time course is observed, with the observed rate constant being directly related to the product of the initial concentration of the reagent in excess to the forward bimolecular rate constant. In the case of a simple reversible or irreversi-

ble bimolecular reaction, a plot of the observed rate constant versus the concentration of the reagent being varied will result in a linear relationship, wherever pseudo-first order conditions are met. The slope of the linear plot yields the apparent bimolecular rate constant for the reaction, under the chosen experimental conditions.

According to the law of mass action, the rate of association of A with B is thus the product of the concentrations of A and B and a rate constant (the association rate constant, k_{on}). The rate of dissociation is the product of the concentration of AB and a rate constant (the dissociation rate constant, k_{off}). The differential equations for [AB] is the rate of change of the target-ligand complex concentration over time, given by the rate of association minus the rate of dissociation:

$$\frac{d[AB]}{dt} = k_{on}[A][B] - k_{off}[AB]$$

Eq. 2.1

For the dissociation assay, the analytic equation needs to incorporate the amount of AB at the initiation of the dissociation phase. The equation defining the dissociation time course is derived using the same strategy as described above for the association process. The rate of change of the target-ligand complex concentration over time is simply the rate of decay of the complex: Therefore, the second differential equation describing the dissociation phase is

$$\frac{d[AB]}{dt} = -k_{off}[AB]$$

Eq. 2.2

The resolving functions of the differential equations describe the association and dissociation phases and their analytical form is a single exponential function:

$$y = y_0 + s(1 - e^{-k_{\text{obs}}t})$$

$$y = y_0 + se^{-k_{\text{off}}t}$$

Eq. 2.3

where $k_{\text{obs}} = k_{\text{on}}[B] + k_{\text{off}}$.

All observed time courses will eventually reach equilibrium asymptotically. In this case, the forward binding transition should be balanced by the backward unbinding transition thus holding the relation:

$$k_{\text{on}}[A][B] = k_{\text{off}}[AB]$$

From this, the affinity constant is defined as the ratio:

$$K_D = \frac{[A][B]}{[AB]} = \frac{k_{\text{off}}}{k_{\text{on}}}$$

Eq. 2.4

There is a fixed amount of receptor protein on the sensor surface, so there is a maximum possible amount of analyte protein binding at equilibrium. The equilibrium levels reached will depend on the concentration relative to the affinity constant K_D ; at concentration around 100 times K_D , the equilibrium will be at a saturating concentration.

It is clear that the advantage of working under PFO conditions is that time-course can be expressed with simple exponential equations, with the observed rate constant being directly related to the exponential parameter in the binding process. In the case of a simple bimolecular reaction, a plot of the

observed rate constant versus the concentration of the reagent being varied will result in a linear relationship (see **Eq. 2.3**). The slope and the intercept of this plot yield the apparent association and dissociation rate constants for the reaction.

2.3.1 Saturation curve

According to the simple system described in the previous paragraph, the saturation fraction is defined by:

$$Y = \frac{[AB]}{[A] + [AB]} = \frac{[AB]}{[A_{\text{tot}}]}$$

as in the ratio between the concentration of occupied binding sites by the ligand and the total available binding sites. Since $K_D = \frac{[A][B]}{[AB]}$, substituting $[AB]$ from this expression in Y , the saturation curve is:

$$Y = \frac{[B]}{K_D + [B]}$$

Eq. 2.5

where B is the ligand protein, the independent variable of the experiment. It can happen however that **Eq. 2.5** is not always sufficient to describe the experimental time courses observed in a binding assay. For macromolecules having two or more binding sites, a phenomenon called “cooperativity” may occur. Cooperativity is characterized by a change of the intrinsic (site specific) equilibrium binding constant as a function of the reaction progress, meaning that the affinity of a given binding site for a ligand will be affected by the occupancy of other sites by the same or different ligands. To conciliate this

kind of data with the sigmoidal shape of the binding curve, it may be necessary to introduce a cooperativity coefficient, or Hill coefficient, in the fraction saturation equation.

$$Y = \frac{[B]^n}{k_{\text{obs}} + [B]^n}$$

Eq. 2.6

Figure 2.9 exemplifies the types of binding interactions between sites, showing the fractional saturation of binding sites in a macromolecule as a function of the free ligand concentration. In the curves in the left panel it is possible to observe the hyperbolic shape of the saturation curve, characteristic of independent sites without interactions ($n = 1$), and the deviations produced by positive and negative cooperativity. It is worth noting that the sigmoidal shape of the curve is distinctive for positive cooperativity in the direct representation. When the curves are represented as a function of the logarithm of ligand concentration (right panel), an apparent sigmoidal trend appears for all types of binding.

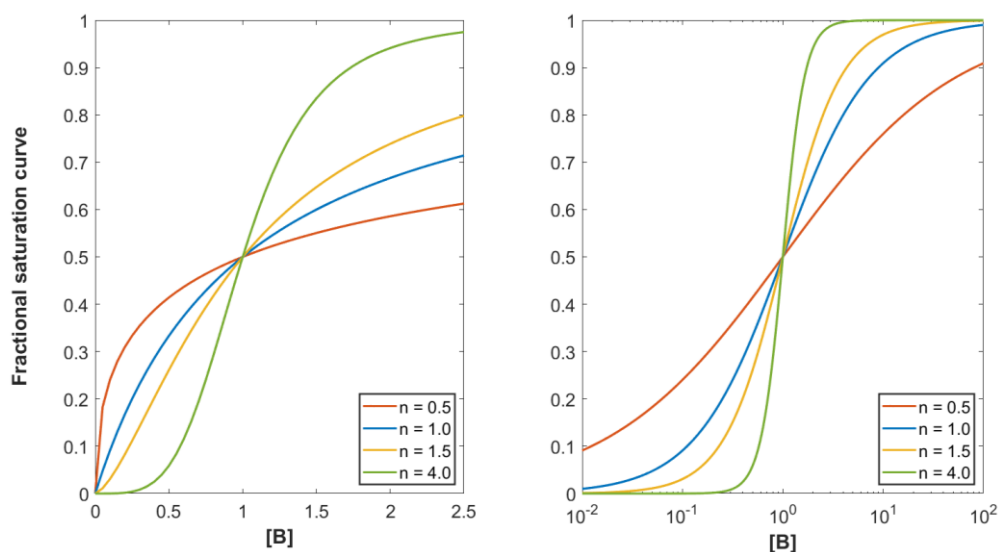


Figure 2.9. Saturation curves for the three characteristic types of interactions between sites. The fractional saturation of a macromolecule by a ligand was simulated using **Eq. 2.6** for identical and independent sites ($n = 1$, blue line), negative cooperativity ($n < 1$, red line) and positive cooperativity ($n > 1$, yellow and green lines). The right panel includes saturation values for a wide range of ligand concentration in a logarithmic scale.

For the turbidimetric data analyzed in **Chapter 3.1**, **Eq. 2.6** was used instead that explicitly takes into account a cooperativity coefficient. Thus far, the apparent thermodynamic constants could not be appropriated described as “ K_D ” and is referred to as “ k_{obs} ”.

When available, the concentration range for each binding experiment was chosen based on K_D values obtained from the literature or experimentally determined for different scenarios where K_D values were unknown. The recorded data were always analyzed using the Octet software to extrapolate the kinetic parameters. All association and dissociation curves were fitted using a single exponential function. Pseudo-first order (PFO) conditions were met when the initial concentration of one of the two reagents was significantly higher (between 50- and 100-fold) than the other.

2.4 2D Zernike computational method

In this paragraph, the new unsupervised computational method is described, that is behind the computational predictions at the basis of the experimental work carried out and presented in the results section. Such method efficiently characterizes the shape of any portion of molecular surfaces and allows one to evaluate the shape complementarity of protein-protein interfaces employing the 2D Zernike formalism. A 2D Zernike polynomial expansion refers to a mathematical technique that associates each portion of molecular surfaces with an ordered set of numerical descriptors, invariant under rotation, allowing easy metric comparison between the shape of different protein regions for similarity or complementarity evaluation. The algorithm reduces the computational cost that would be required if all possible relative angles between the surfaces would be considered. In a 2D Zernike polynomial expansion, the 2D function or surface is decomposed into a sum of these Zernike polynomials, each multiplied by a coefficient. These coefficients determine the contribution of each Zernike polynomial to the overall shape of the function or surface. The expansion allows for a concise and systematic representation of complex shapes or irregularities. **Figure 2.10** depicts how the algorithm work in a step-by-step illustration.

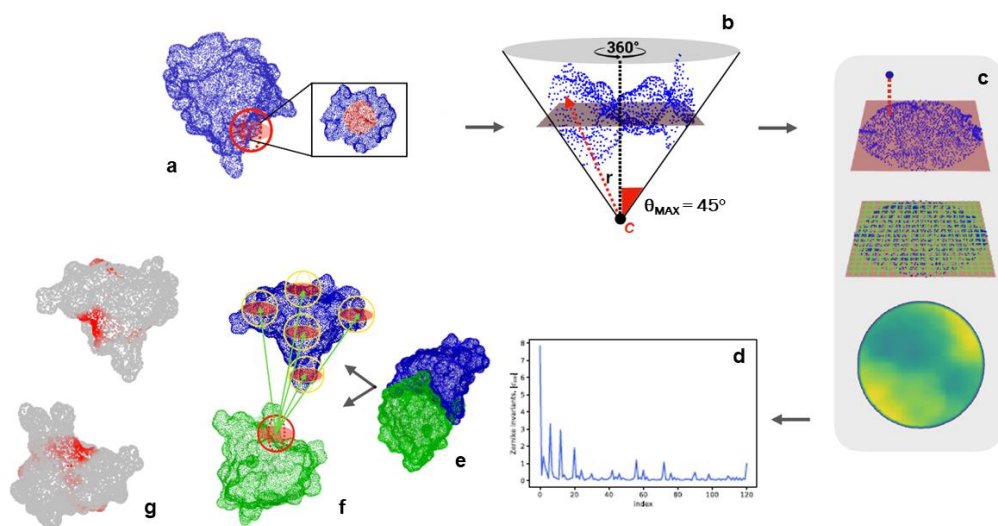


Figure 2.10. Computational protocol for the characterization of each surface region and the blind identification of the binding sites. **a.** Molecular solvent-accessible surface of a protein (in blue) and example of patch selection (red sphere). **b.** The selected patch points are fitted with a plane and reoriented so that the z-axis passes through the centroid of the points and is orthogonal to the plane. The point C along the z-axis is defined, such as that the largest angle between the perpendicular axis and the secant connecting C to a surface point is equal to 45° . Finally, to each point, its distance, r with point C is evaluated. **c.** Each point of the surface is projected on a squared grid fit plane. To each pixel, the average of the r values of the points inside the pixel is associated. **d.** The resulting 2D projection of the patch can be represented by a set of 2D Zernike invariant descriptors. **e-f.** For a given protein-protein complex under study, a patch centered on the surface vertex is selected and Zernike descriptors are computed. To blindly identify the two binding sites, each sampled patch is compared with all the patches of the molecular partner and each vertex is associated with the minimum distance between its patch and all the patches of the molecular partner is associated with. **g.** A Smoothing process of the surface point values is applied to highlight the signal in the regions characterized mostly by low distance values. Picture adapted from (Milanetti, Miotto, Di Rienzo, Nagaraj, et al., 2021).

For a more detailed explanation of the protocol and applications of the algorithm, (Milanetti, Miotto, Di Rienzo, Monti, et al., 2021; Milanetti, Miotto, Di Rienzo, Nagaraj, et al., 2021) suggested readings are here enclosed. By applying the Zernike method, a more exhaustive overview of the binding between ACE2 and spike protein was obtained. In **Figure 2.11** there is a zoom

on the binding region between the receptor binding domain of SARS-CoV-2 spike protein and the ACE2 receptor protein, as obtained from X-ray crystallography.

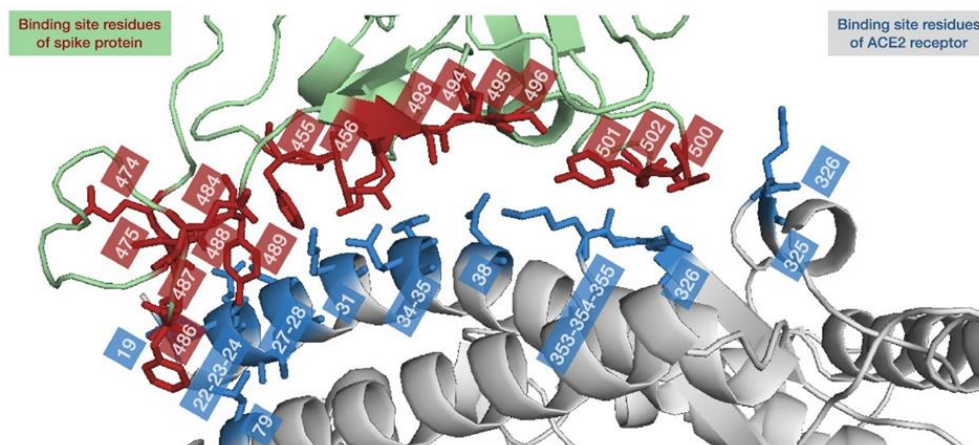


Figure 2.11. Ribbon representation of the binding site of the SARS-CoV-2 spike protein bound to human ACE2 (PDB id: 6M0J). Interacting residues whose α -carbons have a distance lower than 9 Å are represented with red sticks for the spike protein and with blue ones for the ACE2 receptor. Figure from (Miotto et al., 2022).

This analysis and the consequent simulations were the basis of the experimental work whose results are reported in **Chapter 3**.

2.5 Proteins expression and purification

Some of the proteins utilized during the experiments were provided by manufacturers. The ferritin nanoparticles were purified following the protocols found in the literature and possibly modifying some steps.

2.5.1 Commercial proteins

Lactoferrin. Human milk lactoferrin (Lf) protein was from Sigma-Aldrich, with molecular weight between 82.4 and 87 kD; the estimated extinction coefficient was $\epsilon^{\text{mM}} = 110.96 \text{ mM}^{-1} \text{ cm}^{-1}$ (280 nm), and the pI was 8.7. The storage buffer was 10 mM phosphate buffered saline at pH 7.4; therefore, all dilutions were executed in the same buffer.

RBD. The wild-type receptor-binding domain (RBD) of SARS-CoV-2 spike protein was provided by GenScript. The protein is a C-term HIS-tagged recombinant protein with a predicted molecular weight of 30 kD. The storage buffer was phosphate-buffered saline at pH 7.2, and the calculated pI was 8.91; the stock concentration was 0.89 mg/mL according to the manufacturer.

ACE2. The angiotensin-converting enzyme 2 (ACE2) was also provided by GenScript. The ACE2 ectodomain is a C-term Fc-tagged recombinant protein with predicted molecular weight of 110.5 kD at a concentration of 1.34 mg/mL, in a storage buffer consisting of 20 mM Tris-HCl, 300 mM NaCl, 1 mM ZnCl₂ and 10% glycerol, pH 7.4. The protein pI is 5.49.

2.5.2 Proteins and peptides synthesis

ACE2-derived peptides. WT, HA1, HA2, HA3, HA4, LA1 were synthesized by GenScript and provided in the lyophilized form. Before use, they were re-suspended in PBS buffer containing either 3% NH₄OH or 5% NH₃ (depend-

ing on the total net charge of the peptides). The final stock concentration for each peptide was 5mg/mL. The peptides were either subjected to a biotinylation reaction using the BTAG biotinylation kit (Sigma-Aldrich) or were provided pre-biotinylated by the manufacturer (GenScript).

Chimeric ferritins with ACE2-derived peptides. The synthetic genes encoding for HFt-peptHA1 and HFt-peptWT chimeric proteins were synthesized by GenScript and optimized for expression in *Escherichia coli* cells. Peptide sequences were inserted at the N-terminal region of the ferritin through a 4 glycine-spacer. The genes were then subcloned into pET22 b vector and transformed in BL21(DE3) competent cells. Ferritins were expressed as previously reported (Palombarini et al., 2019) and the purification protocols are exposed in the following paragraphs.

Engineered ferritin with PBM. The synthetic genes encoding for HFt-PBM protein was provided by the manufacturer with different characteristics from the chimeric ferritin.

2.5.3 Purification protocols

HFt-peptWT purification protocol. Bacterial paste from 1 L culture was resuspended in 50 mL of 20 mM sodium phosphate buffer, pH 7.4 containing 150 mM NaCl and protease inhibitors (Roche©) and disrupted by sonication. The soluble fraction was precipitated with 20% $(\text{NH}_4)_2\text{SO}_4$ for 1 h at 4°C. The pellet was resuspended in 20 mM sodium phosphate buffer pH 7.4 containing 150 mM NaCl and extensively dialyzed versus the same buffer overnight at 4°C. Finally, the protein was loaded onto a HiLoad 26/600 Superdex 200 pg column equilibrated with 20 mM sodium phosphate buffer pH 7.4 containing 150 mM NaCl, using an AKTA-Pure apparatus (Cytiva). Protein concentration was calculated by measuring the UV absorption at 280 nm (ϵ_{280})

= 20400 M⁻¹ cm⁻¹). Protein purity was checked by SDS-PAGE and the correct quaternary assembly was evaluated by high-performance size exclusion chromatography (HP-SEC).

HFt-peptHA1 purification protocol. Bacterial paste from 1 L culture was resuspended in 50 mL of 50 mM Tris/glycine buffer, pH 7.5 containing 150 mM NaCl, 0.5 mM TCEP and protease inhibitors (Roche©) and disrupted by sonication. The insoluble fraction was resuspended in 50 mL of 100 mM Tris/glycine buffer pH 8.5, containing 0.5 mM TCEP and 0.5 M urea. The solution was stirred for 30 min at room temperature. The resulting washed pellet was resuspended in 50 mL of 100 mM Tris/glycine buffer pH 8.5, containing 0.5 mM TCEP and 4 M urea, and stirred for 30 min at room temperature. After centrifugation, the soluble fraction was dialyzed overnight versus 50 mM Tris/glycine buffer pH 9 at 4°C. The protein sample was then dialyzed for 4 h versus sodium phosphate buffer pH 7.4, concentrated, and loaded onto a HiLoad 26/600 Superdex 200 pg as described above. The protein fractions eluting at the retention time of ferritin were pooled, concentrated using AmiconUltra-15 centrifugal filter devices (100 kDa cut-off), sterile filtered, and stored at 4°C. Protein concentration was calculated by measuring the UV absorption at 280 nm ($\epsilon_{280} = 20400 \text{ M}^{-1} \text{ cm}^{-1}$). Protein purity was checked by SDS-PAGE and the correct quaternary assembly was evaluated by high-performance size exclusion chromatography (HP-SEC).

HFt-WT purification protocol. Bacterial paste from 1 L culture of HFt was resuspended in 50 mL of 20 mM sodium phosphate buffer, pH 7.4, containing 150 mM NaCl and protease inhibitors (Roche©) disrupted by sonication. The soluble fraction was treated with 50% and 70% (NH₄)₂SO₄. The 70% (NH₄)₂SO₄ pellet was resuspended in 20 mM sodium phosphate buffer pH 7.4 containing 150 mM NaCl and extensively dialyzed versus the same buffer

overnight at 4°C. After dialysis, the protein sample was subjected to heat treatment at 75°C for 10 min. The resulting soluble fraction was digested with 50 µg/mL deoxyribonuclease I (Merck) for 1 h at 3°C with the addition of 2 mM MgCl₂. After digestion, the protein sample was loaded onto a Hi-Load 26/600 Superdex 200 pg column previously equilibrated with 20 mM sodium phosphate buffer pH 7.4 containing 150 mM NaCl, using an AKTA-Pure apparatus (Cytiva). The protein fractions eluting at the retention time of ferritin were pooled, concentrated using AmiconUltra-15 centrifugal filter devices (100 kDa cut-off), sterile filtered, and stored at 4°C. Protein concentration was calculated by measuring the UV absorption at 280 nm ($\epsilon_{280} = 19000 \text{ M}^{-1} \text{ cm}^{-1}$) and protein purity was checked by SDS-PAGE and High-performance size exclusion chromatography (HP-SEC).

HFt-PBM purification protocol. The purification protocol for the engineered ferritin with PMB was based on the HFt-WT with several alterations.

2.5.4 Protein characterization techniques

High-performance size exclusion chromatography (HP-SEC). The purity and aggregation state of protein samples were analyzed by high-performance size exclusion chromatography (HP-SEC). HP-SEC analyses were performed by means of an Agilent Infinity 1260 HPLC apparatus equipped with UV detectors using an Agilent AdvanceBio SEC 300A 2.7 µm 4.6x150 mm column. Isocratic analysis was carried out with 20 mM sodium phosphate buffer pH 7.4 as the mobile phase. The flow rate was 0.7 mL/min over an elution window of 10 min. Ferritin elution was followed using UV detection at 220 nm and 280 nm.

Native Polyacrylamide Gel Electrophoresis (PAGE). Native Polyacrylamide Gel Electrophoresis was performed with a 4-15% nondenaturing

acrylamide gel (Mini-PROTEAN TGX stain-free) in a Tris-glycine pH 8.3 running buffer at room temperature for 30-40 min at 150-200 V in a Bio-Rad Mini-Protean tetra-cell electrophoresis apparatus.

Chapter 3 Results and discussion

3.1 Investigation on inhibiting role of lactoferrin in the complex formation between ACE2 receptor and SARS CoV-2 recognition binding domain

In the following chapter, the investigation focused on the analysis of the interactions among bovine lactoferrin, SARS-CoV-2 receptor-binding domain (RBD) and human angiotensin-converting enzyme 2 (ACE2) receptor will be discussed in order to assess possible mutual interactions. The rationale for this experimental set up was based on previous proposals that suggested possible mechanisms of antiviral action of lactoferrin (Campione et al., 2020; Kell et al., 2020; Rascón-Cruz et al., 2021; Y. Wang et al., 2020) and on computational studies (Campione et al., 2021; Milanetti, Miotto, Di Rienzo, Nagaraj, et al., 2021; Miotto, Di Rienzo, Bò, et al., 2021; Tortorici et al., 2019) that revealed significant surface complementarity among lactoferrin and two binding regions on the RBD domain and several possible binding sites to ACE2. Kinetic and thermodynamic parameters for the pairwise interactions were measured by means of biolayer interferometry and latex nanoparticle-enhanced turbidimetry techniques. The results obtained indicate that Lf is able to bind the ACE2 receptor ectodomain with significantly high affinity, whereas no direct binding to the RBD was observed up to the maximum achievable lactoferrin concentration range. Lactoferrin, above 1 μM concentration, thus appears to directly interfere with RBD–ACE2 binding, bringing about a measurable, up to 300-fold increase of the K_D value relative to the isolated RBD–ACE2 complex formation. The observed results could thus

provide a molecular basis of the reported preventative effect of lactoferrin against CoV-2 infection.

3.1.1 Binding of lactoferrin to ACE2

The molecular interactions of Lf with ACE2 and RBD were studied separately via both BLI and turbidimetric assays. **Figure 3.1** report the time course of the binding and dissociation signals between Lf and ACE2. Protein ACE2 is loaded on protein A biosensors (“ProA”) at fixed concentration of 50 μ g/ml, Lf is present in solution at increasing concentrations, starting from 33 μ M down to 0.33 μ M, obtained by consecutive dilutions in 1X kinetic buffer. The association and dissociation steps duration were set at 120 s and 300 s respectively. The value of K_D obtained from the BLItz software was $K_D = 27.64$ μ M.

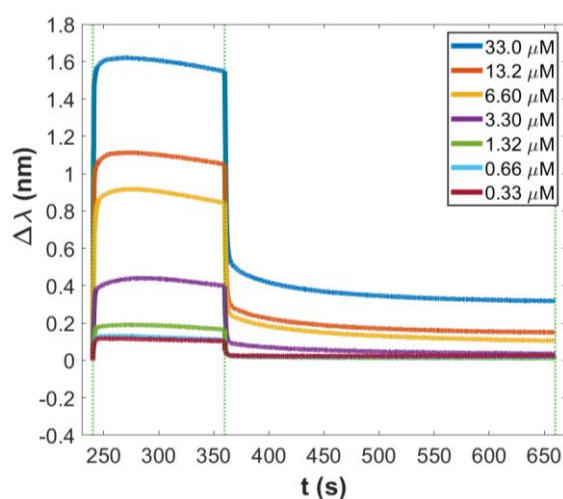


Figure 3.1. Time-courses of the BLI assay showing binding and dissociation trend between Lf and ACE2. The analyte protein is Lf at different concentrations, whereas ACE2 is loaded on the biosensor (ProA). The vertical dashed lines indicate the duration of the binding step (120 s) and of the dissociation step (300 s).

The time-courses relative to the turbidimetric assay are shown in **Figure 3.2**. Nanospheres with a suitable diameter of 103 nm were coated with ACE2 and Lf was added to the cuvette at concentrations ranging from 21.6 μM to 1.96 nM. The increase in light absorbance at 340 nm was measured.

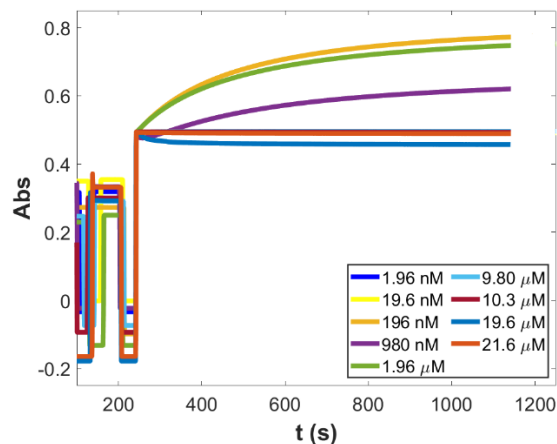


Figure 3.2. Turbidimetric assay performed with Lf as the analyte protein at variable concentrations in solution with latex nanospheres coated with ACE2. From the data analysis performed using **Eq. 2.5**, the value of K_D obtained was 46.12 μM .

These results show that, taken individually, the Lf-ACE2 interaction clearly occurs when measured by both techniques and it is quantifiable with an affinity constant K_D in a micromolar order of magnitude.

3.1.2 Binding of lactoferrin to RBD

The same experimental set up was applied to study the binding between LF and RBD. **Figure 3.3** and **Figure 3.4** show the BLI and turbidimetric analysis for the association of LF to RBD. Concerning the interferometry assay, LF as the protein in solution spanned a concentration ranged from 10 μM to 50 nM, while RBD protein was loaded on anti-HIS tag biosensors ("HIS2") at fixed

concentration of 50 $\mu\text{g/ml}$. The duration of the binding and dissociation steps were set for 180 s and 120 s, respectively. For the turbidimetric assay, nanospheres were coated with RBD protein and Lf was present in solution concentrations starting from 21.6 μM down to 540 nM.

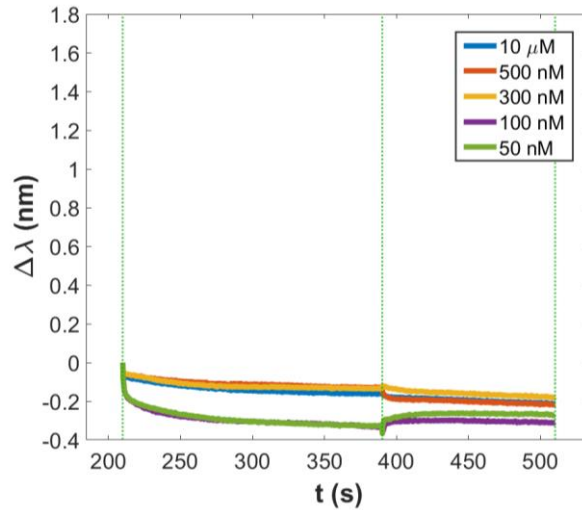


Figure 3.3. BLI signals of Lf as the analyte protein at different concentrations with RBD loaded on the biosensor (HIS2). The binding step was 180 s, the dissociation step was 120 s.

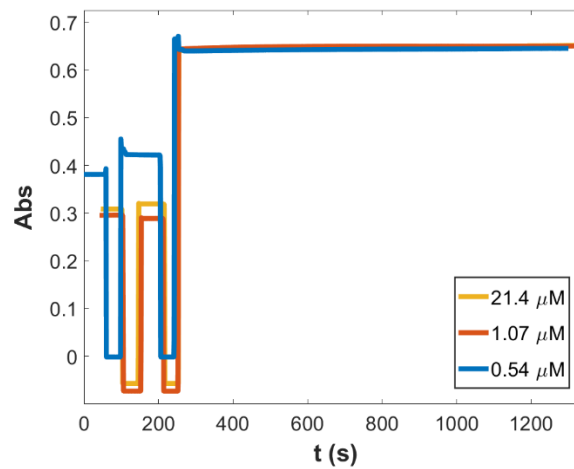


Figure 3.4. Turbidimetry assay performed with Lf as the analyte protein at variable concentrations in solution with latex nanospheres coated with RBD.

It is clear from the trends of the reported curves represented that no signal could be detected between Lf and RBD within the concentration range under study.

In **Table 1** are displayed the kinetic parameters of Lf–ACE2 interaction. The corresponding K_D values were 27.64 μM for the interferometric experiment and 46.12 μM in the turbidimetric measurement.

	k_{on} ($\text{M}^{-1} \text{s}^{-1}$)	k_{off} (s^{-1})	K_D (μM)
INTERFEROMETRY	(166.90 ± 4.79) 10^2	0.461 ± 0.007	27.64 ± 0.91
TURBIDIMETRY	33.41 ± 5.41	(1.54 ± 0.39) 10^{-3}	$46.12 \pm$ 12.12

Table 1. Kinetic parameters relative to Lf and ACE2 interaction. The k_{on} and k_{off} parameters were estimated by single exponential fit of the curves (see **Eq. 2.3**), and the corresponding K_D values were calculated according to **Eq. 2.5**.

These results clearly indicated that Lf showed a quantifiable interaction with the ACE2 protein but no binding at all with the RBD protein within the observed concentration range.

3.1.3 Effect of lactoferrin in the interaction between RBD and ACE2

The ability of Lf to effectively inhibit the interaction between ACE2 and RBD was further investigated by means of interferometric and turbidimetric measurements. Before evaluating that, however, the affinity between RBD and ACE2 was verified to fall in the nanomolar range, as published in previous reports (Saponaro et al., 2020). BLI analysis is presented in **Figure 3.5** and was performed by loading ACE2 on ProA biosensors and allowing the

association/dissociation of RBD at concentrations ranging from 1.67 μM to 10 nM.

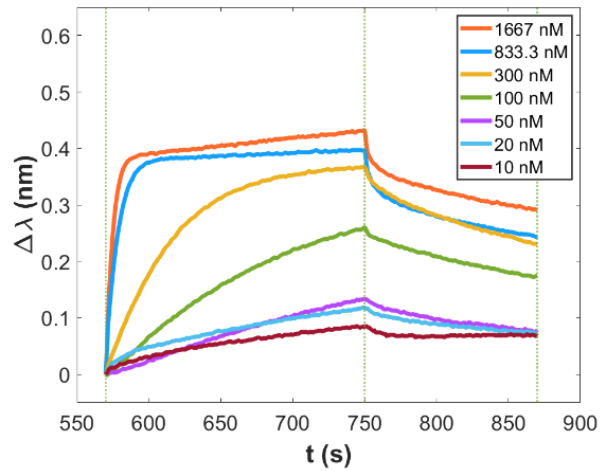


Figure 3.5. Time-courses of the reaction between RBD in solution with ACE2. Figure shows signals of the binding and dissociation experiment performed via BLI with ACE2 loaded on ProA biosensors, and RBD present in solution at decreasing concentrations. The vertical dashed lines indicate the time interval of the binding step (180 s) and of the dissociation step (120 s).

From data fitting analysis performed by BLItz software, the resulting affinity constant was 27.06 nM, in good agreement with previously observed values. In **Figure 3.6**, aggregation time-courses from turbidimetric assays are shown.

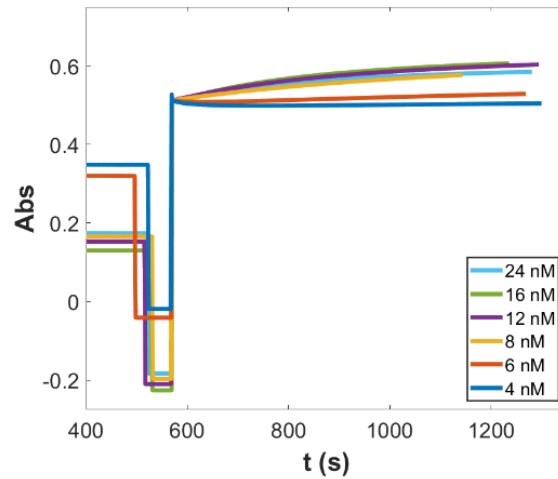


Figure 3.6. Aggregation signals from turbidimetric assays of the reaction between RBD in solution with ACE2. Latex nanospheres were coated with the ACE2 protein and mixed in solution with RBD at decreasing concentrations. The graphs refer to measurements of latex nanospheres coated with ACE2 protein and mixed in solution with RBD at decreasing concentrations. The resulting affinity constant was 18.15 nM.

Ultimately, Lf was introduced into a solution containing RBD, in order to explore its role in the interaction with ACE2. In the following graphs, one can observe the gathering of data through BLI and turbidimetry concerning the protein system. In this scenario, ACE2 served as the receptor (located either on the sensor tip or on the nanoparticles' surface), while RBD remained at fixed concentration. Variable concentrations of Lf were introduced. The impact of Lf on the formation of the ACE2–RBD complex was initially assessed using BLI analysis, as depicted in **Figure 3.7**.

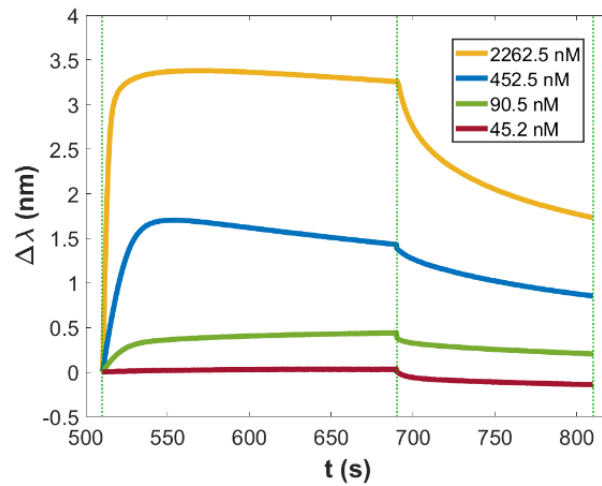


Figure 3.7. Time-courses of the reaction between RBD and ACE2 in the presence of Lf in solution. ACE2 is loaded on ProA biosensors, RBD is present in solution at constant concentration along with lactoferrin, whose decreasing concentrations are reported in the legend box. The vertical dashed lines indicate the start of the binding step (180 s) and of the dissociation step (120 s).

ACE2 was once again loaded onto ProA biosensors, and RBD was introduced to the biosensors during the association step, maintained at a steady concentration of 452.5 nM, and Lf was mixed with diminishing concentrations. The outcome involved both Lf and RBD binding to the ACE2 immobilized on the biosensors, leading to an observable k_{obs} wherein the binding affinities of both proteins to the receptor became intertwined. Interestingly, when Lf was present in solution at a concentration as low as 45.2 nM, the association signal recorded for the mix was significantly lower than that obtained in the presence of RBD alone, thus suggesting a strong inhibiting effect of Lf. The value of K_D obtained from the BLItz software was $K_D = 101.30$ nM. The same mixing approach was then applied to the turbidimetric methodology, by mixing in the cuvette nanospheres coated with ACE2 as receptor and a solution at a constant concentration of RBD (19.6 nM) in the presence of decreas-

ing concentrations of Lf. An inhibitory effect on the aggregation of ACE2 with RBD was observed, as shown in **Figure 3.8**.

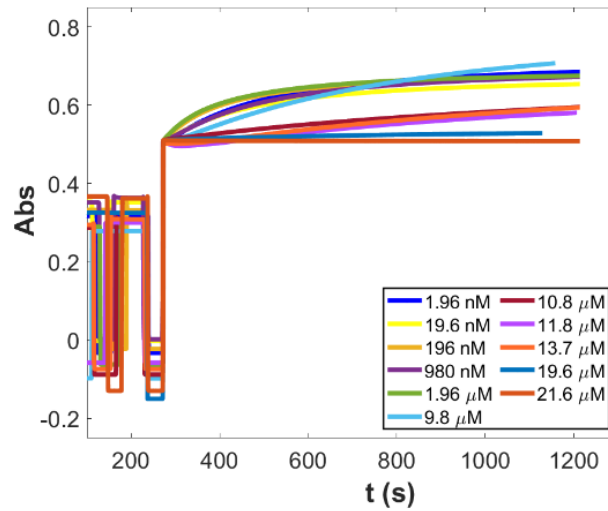


Figure 3.8. Turbidimetry assay performed with Lf at variable concentrations and RBD present at a fixed concentration, in solution with latex nanospheres coated with ACE2.

For a more in-depth examination of the turbidimetric data, curve fitting of the saturation curve and of the observed rates was executed. Results are shown in **Figure 3.9**. According to **Eq. 2.5**, the saturation of binding sites expressed in terms of ΔAbs exhibited a hyperbolic dependence on the concentration of the ligand protein, lactoferrin, whereas the curve appeared sigmoidal when protein concentration is expressed in logarithmic form.

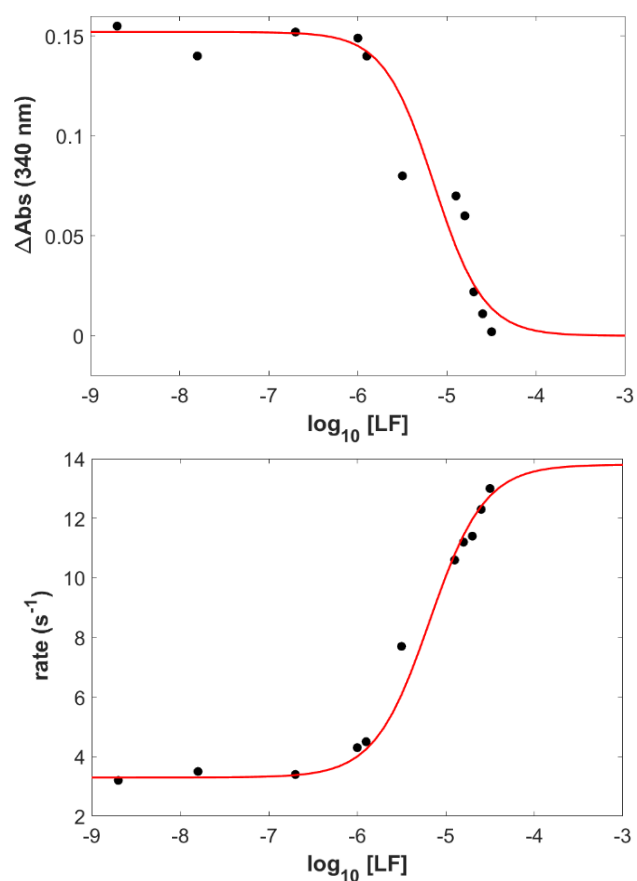


Figure 3.9. Inhibition of RBD-ACE2 complex formation in the presence of increasing concentrations of lactoferrin. The turbidimetric data of **Figure 3.8** were used to analyze the effect of increasing lactoferrin concentration (in log scale). The absorbance amplitudes of the reactions are shown in **top panel**, whereas the initial rates of the same curves are depicted in **bottom panel**. Interpolating red curves represent the best fit to the data obtained by **Eq. 2.6**. The data analysis, performed via custom MATLAB program, yielded apparent k_{obs} values of $9.5 \pm 1.5 \mu\text{M}$ (top) and $6.3 \pm 1.2 \mu\text{M}$ (bottom), respectively, with cooperativity coefficient n of 1.44 ± 0.11 and 1.54 ± 0.08 , respectively.

The flex point corresponded to an “apparent” dissociation constant k_{obs} of the reaction. In the top panel of **Figure 3.9**, the values of ΔAbs were plotted against $\log_{10}(\text{Lf})$. The difference in absorbance for each concentration of Lf was calculated between the start of step 4 of the turbidimetric assay and 600 s after it.

3.1.4 Computational recognition of the binding regions of ACE2 to lactoferrin

To conclude the chapter, it is reported in **Figure 3.10** a representation of the regions of ACE2 protein showing the highest binding propensity to Lf, according to the computational approach based on Zernike method. The molecular structures of human ACE2 (PDB id: 1R42) and human holo lactoferrin (PDB id: 1LFG) were inspected for portions of the molecular surfaces with high shape complementarity (Milanetti, Miotto, Di Rienzo, Monti, et al., 2021). In particular, a set of ACE2 receptor regions with high binding propensity were identified. Panel **a** shows the residues of ACE2 having the highest binding propensity scores with human Lf. Panel **b**, **c** and **d** show the identified regions on the molecular surface of human ACE2.

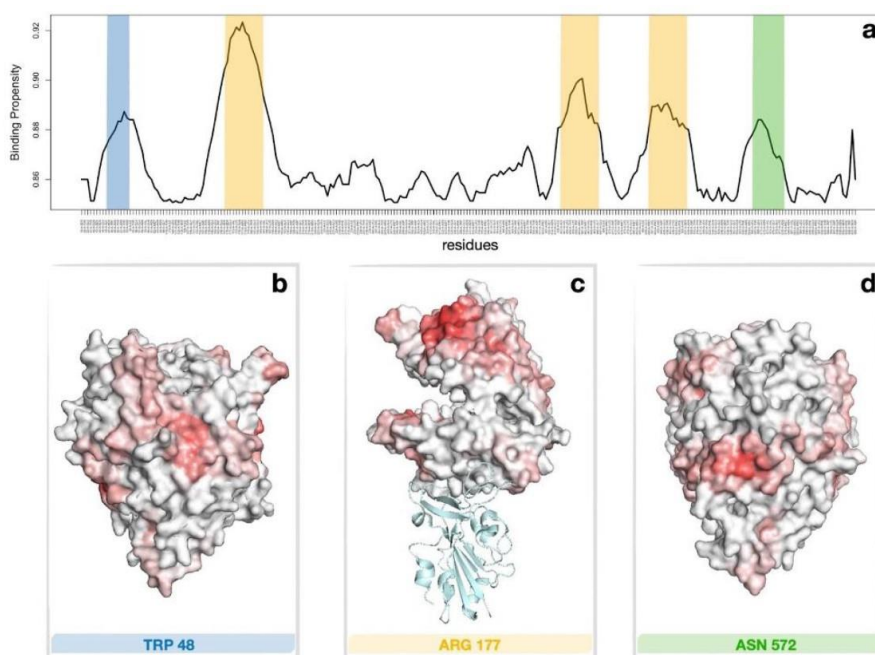


Figure 3.10. Regions of high binding propensity of ACE2 to Lf. **a.** Binding propensity of human ACE2 residues for human lactoferrin obtained on the basis of local shape complementarity of the molecular

surfaces (Miotto, Di Rienzo, Gosti, et al., 2021). Only residues whose binding propensity is higher than 0.85 are reported. Blue, orange, and green bands highlight three different portions of the molecular surfaces characterized by high binding propensity. **b.** Molecular surface of the extracellular region of human ACE2 colored according to the Zernike binding propensity score. The color turns from white to red as the local binding propensity increases. The surface was oriented to show the region around residue Tpr48, which is comprised in the blue band in panel **a.** **c.** Same as in **b** but displaying the region around residue Arg177 and marked with the orange bands; a cartoon representation of the RDB of the spike protein bound to ACE2 is also shown. **d.** Same as in **b**, but for the green band of the panel **a** marking the region around residue Asn572.

The model suggested that the region of binding of Lf does not strictly overlap with the well-known ACE2–RBD binding interface. As such, two different inhibition scenarios can be hypothesized: on one side, Lf binding on an ACE2 surface region far from the interaction surface with RBD (such as in **Figure 3.10c**) can lead to structural rearrangements on the ACE2 ectodomain that could effectively prevent spike attachment to host cells; on the other hand, Lf attachment on regions highlighted in **Figure 3.10b, d** could hamper directly ACE2–RBD interaction.

3.2 Design of protein-binding peptides with controlled binding affinity: the case of SARS-CoV-2 receptor binding domain and ACE2-derived peptides

The experimental work based upon the 2D Zernike formalism applied in order to design high affinity peptides to specific targets is described in the present chapter. The algorithm was based on shape complementarity optimization and electrostatic compatibility and focused on the interaction between the SARS-CoV-2 Spike Receptor-Binding Domain (RBD) and the human angiotensin-converting enzyme 2 (ACE2) receptor. On the bases of known 3D structures obtained by X-ray crystallography, a 23-residue peptide that structurally emulates the key interacting segment of the ACE2 receptor was identified. Starting from the wild-type sequence, a set of five different peptide mutants, each possessing controlled binding affinities, was generated. This computational forecast was subjected to validation through both molecular dynamics simulations and experimental *in vitro* methodologies. In this chapter it will be exposed the experimental determination of binding affinity (K_D) executed using biolayer interferometry corroborated the predictions generated *in silico*. This approach has been extensively employed in the investigation of antibody-antigen interactions (Concepcion et al., 2009; Kamat & Rafique, 2017; Petersen, 2017), with a specific emphasis on the interactions between SARS-CoV-2 variants and antibodies (Dzimianski et al., 2020; Ginex et al., 2022; R. Wang et al., 2022; Y. Wang et al., 2020). Finally, as a further experimental development, the peptide sequence exhibiting the highest affinity was expressed as a fusion protein with the human H ferritin (HFt) 24-mer. Analytical assessments performed on the resulting fusion protein confirmed the

consistent behavior of the peptides, thereby amplifying their effectiveness in binding to the RBD.

3.2.1 Computational analysis of the ACE2-derived peptides

The design protocol was applied onto residues into the α -helix peptide range 21-43. The spatial arrangement of the peptide within the ACE2-RBD complex can be observed in **Figure 3.11**. Within the aminoacidic range considered, the protocol performed the substitution in each position of the WT molecule with the 19 other possible lateral chains, selecting, among the substitutions with compatible electrostatics, the ones driving the highest increase in shape complementarity. Using this approach, a set of four mutated peptide sequences (HA1, HA2, HA3, and HA4) that exhibited higher nominal affinity towards the RBD compared to the wild-type sequence (WT), and one peptide sequence with lower affinity (LA1) were generated.

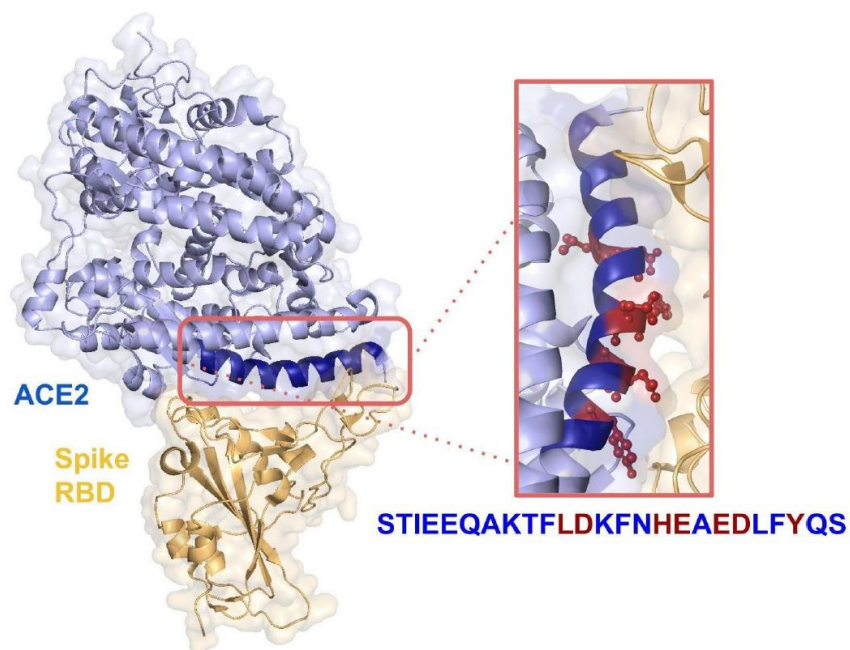


Figure 3.11. Crystal structure of SARS-CoV-2 spike receptor-binding domain bound with ACE2 (PDB id: 6MOJ). Relevant side chains of the residues in the α -helix are shown in the orange box and are colored in red.

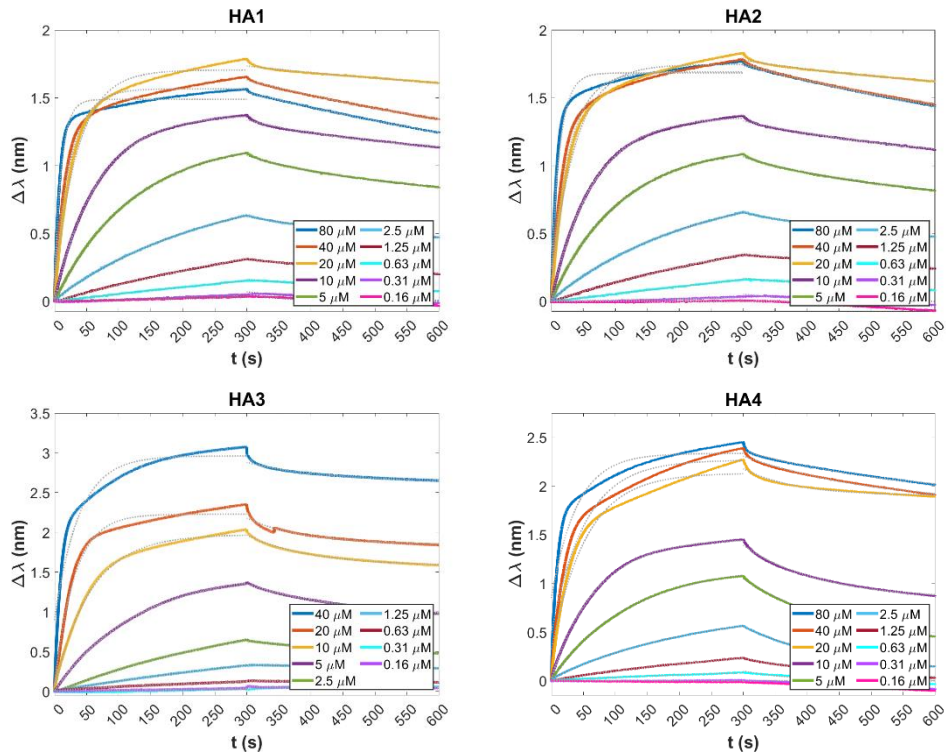
The sequences of each of the six peptides are detailed in **Table 2**, where the specific mutated amino acids are highlighted in red.

WT	STIEEQAKTFLDKFNHEAEDLFYQS
HA1	STIEEQAKTFLDKFN ILAL DLFYQS
HA2	STIEEQAKTFLDKFN VLAL DLFYQS
HA3	STIEEQAKTF DD KFN IL AEDLFYQS
HA4	STIEEQAKTF YD KFN VL AEDLFYQS
LA1	STIEEQAKTFL G KFNHEAE YLFR QS

Table 2. Sequences of the wild type α -helix peptide and the list of mutated sequences HA1-4 and LA1. The mutated amino acids are highlighted in bold red.

3.2.2 BLI measurements of RBD-peptide complexes

Binding kinetics measurements were carried out by using the biolayer interferometry technique, in order to assess the affinity between the ACE2-derived peptides and RBD fragment. Biotinylation was performed on the peptides to enable their immobilization on the streptavidin biosensors. Subsequently, kinetic measurements were conducted at various concentrations of RBD spanning the range from 100 μM to 0.16 μM . In **Figure 3.12** are show the association and dissociation steps for each peptide, while in **Table 3** are presented the obtained K_D values, representing the affinity between the peptides and RBD.



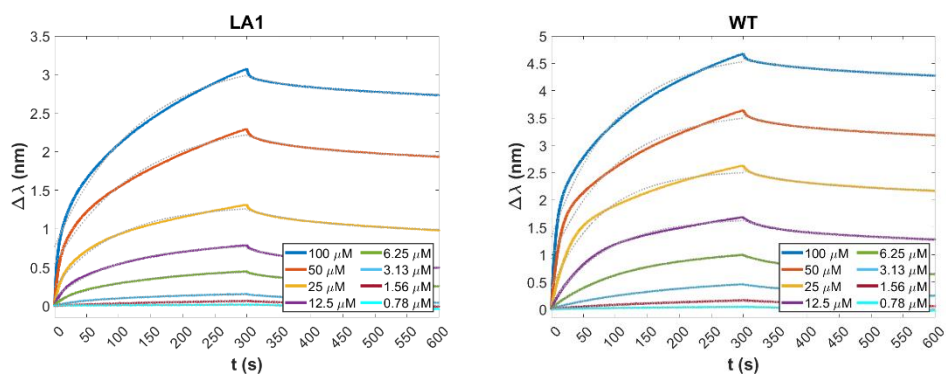


Figure 3.12. Time-course of the reaction between peptides and RBD as the analyte protein in solution at various concentrations. For all measurements, a 300 s time period was set for both association and dissociation steps. Peptides have been loaded first at fixed concentration (50 $\mu\text{g}/\text{ml}$) on biosensors initialized with streptavidin molecule (SAX2).

peptide	K_D (μM)
LA1	74.0 ± 7.2
WT	33.0 ± 1.9
HA4	7.2 ± 1.0
HA3	5.0 ± 1.0
HA2	2.70 ± 0.45
HA1	2.20 ± 0.44

Table 3. The values of K_D with standard errors from data analysis of BLI assay traces are reported in the first row. K_D were obtained directly within the Octet software.

Noticeably, the affinity constants exhibit the same trend as the one predicted via the computational approach, as in the HA4, HA3, HA2, and HA1 mutants displayed increased affinity towards RBD compared to the WT peptide. Conversely, LA1 exhibited significantly lower affinity.

3.2.3 Ferritin chimeric nanoparticles

On the basis of the relevant information obtained from each single peptide, the next objective was to provide evidence for the opportunity of merging these peptides with human ferritin H.

In the present case, the genetic fusion was applied for both the WT (wild-type) sequence and the HA1 sequence, which exhibits the highest affinity for RBD, to the N-terminus of human H ferritin. To ensure structural integrity, a flexible spacer consisting of four glycine residues was included in the fusion construct. HFt-peptWT and HFt-peptHA1 constructs were overexpressed in *Escherichia coli*. Interestingly, the HFt-peptWT was successfully expressed in a soluble form, while, HFt-peptHA1 was obtained in inclusion bodies, necessitating additional steps for purification. Nevertheless, both proteins were purified to a high degree of purity, enabling subsequent structural and functional characterization. In the following **Figure 3.13a** it is demonstrated the successful purification of the constructs, illustrating the expected increase in molecular weight, due to the presence of the peptide chain, of the individual subunits of each chimeric ferritin in the SDS-PAGE gel electrophoresis analysis. Moreover, HP-SEC analysis and native gel electrophoresis confirmed the correct assembly of these ferritins into their 24-mer form (Incocciati et al., 2022). The HPLC elution profile revealed a higher molecular weight for the chimeric ferritins compared to human H ferritin. Similarly, the native gel electrophoresis demonstrated that the chimeric variants migrated slower than

ferritin alone (**Figure 3.13b, c**), as expected. The disparities in electrophoretic mobilities (both native and SDS-PAGE) between the two mutants can be attributed to their differing masses and charges.

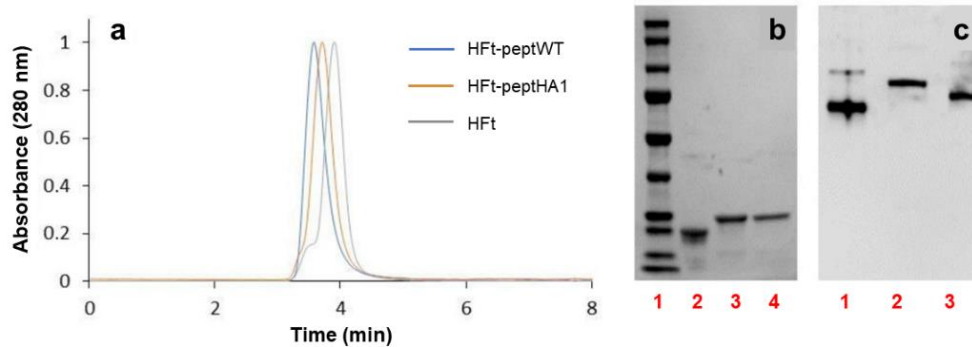


Figure 3.13. Biochemical characterization of HfT-peptWT and HfT-peptHA1 compared to HfT. **a.** HP-SEC analysis. The HfT-peptWT variant displays a weight increase of 77266 Da (retention time 3.53 min) while the HfT-peptHA1 shows an increase of 75924 Da (retention time 3.67 min). **b.** SDS-PAGE (lane 1: marker, lane 2: HfT, lane 3: HfT-peptWT, lane 4: HfT-peptHA1). **c.** native gel electrophoresis (lane1: HfT, lane 2: HfT-peptWT, lane 3: HfT-peptHA1). The disparity of the heights of the bands present on the gel demonstrates that the chimeric variants migrate slower than ferritin alone.

After characterization, the binding capacity of the two chimeric ferritins to the Spike receptor-binding domain (RBD) was assessed. The analysis was performed using the Biolayer Interferometry (BLI) technique, wherein His-tagged RBD was immobilized on the Ni-NTA biosensor, and the concentrations of the chimeric ferritins were systematically varied. Subsequently, the binding kinetics were examined, and the results are illustrated in **Figure 3.14**. The analysis of the kinetic plots revealed a dissociation constant (K_D) of 10.81 nM for HfT-peptWT and 8.32 nM for HfT-peptHA1, providing quantitative insights into the affinity of the respective constructs for RBD (see **Table 4**).

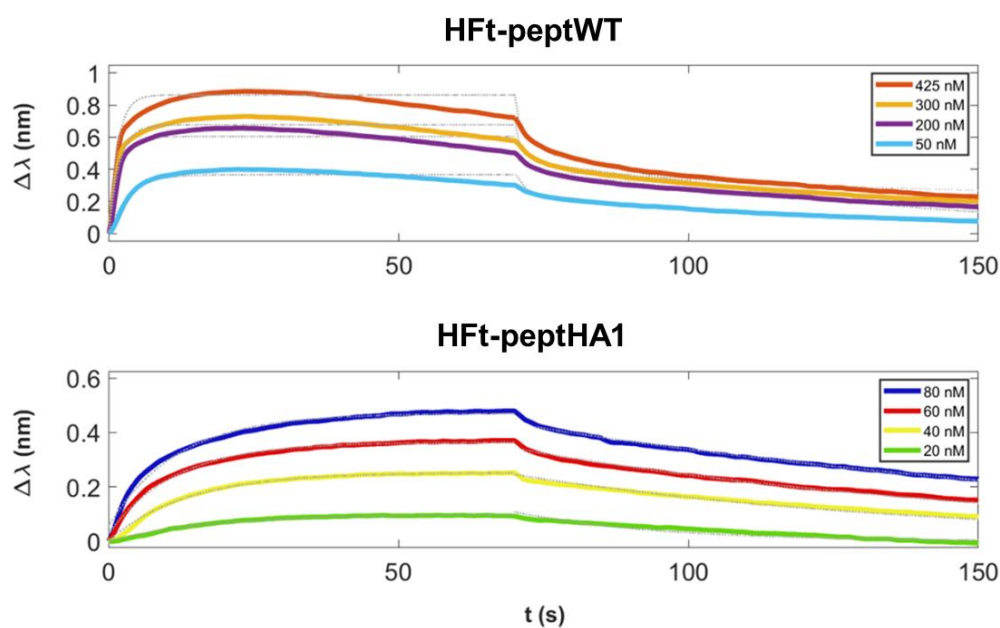


Figure 3.14. Time courses of the reaction between RBD and ferritin constructs. RBD protein has been loaded first at fixed concentration (50 $\mu\text{g}/\text{ml}$) on Ni-NTA biosensors. **Top panel:** ferritin with WT peptide construct is the analyte protein in solution at various concentrations. A 70 s time interval for the association step and 80 s for dissociation were set. **Bottom panel:** ferritin with HA1 peptide construct is the analyte protein in solution at various concentrations. Association and dissociation steps have duration respectively of 70 s and 80 s. Temperature was 25 $^{\circ}\text{C}$.

Complex	HFt-peptWT	HFt-peptHA1
K_D (nM)	10.81 ± 0.06	8.32 ± 0.03

Table 4. Values of K_D with standard errors from data analysis of BLI assay traces.

As can be inferred from the results obtained, the merging of the WT and the best mutant peptide at the N-terminus of human H ferritin 24-mer made it possible to limiting the peptide clearance and at the same time exploiting the multivalent effect conferred by the ferritin structure. Indeed, the fusion of the peptide with the ferritin moiety resulted in a strong increase in the affinity

with RBD, as measured by BLI. These results are strongly encouraging about the effectiveness of the design protocol and the use of ferritin can ensure a broad clinical applicability.

3.3 Engineered ferritin and inhibition of PCSK9 metabolic pathway

In the next section the results of the experiments on the purification of the ferritin containing the PCSK9-binding molecule and the related affinity experiments carried out with the PCSK9 molecule are presented. The binding efficiency was evaluated by measuring the affinity constant via BLI assay. Furthermore, as control experiments, the results obtained from tests using WT ferritin and comparing the value of K_D with the CD71 transferrin receptor are shown.

3.3.1 Purification of engineered nanoparticle

This study aimed to modify a human H ferritin nanoparticle to recognize PCSK9 through the addition of a molecule at the N-terminal of the protein. The sequence was selected based on previously reported PCSK9-targeting sequences obtained by combinatorial library screenings.

The synthetic gene encoding for HFt-PBM was optimized for the expression in *E. coli* cells and the recombinant protein was expressed at a very high level. The results of the purification process are shown in **Figure 3.15**.

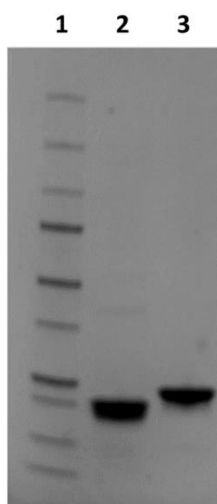


Figure 3.15. SDS-PAGE of the products of purification of HFt-WT (lane 2) and HFt-PBM (lane 3). Compared to human HFt, HFt-PBM has a higher molecular weight. Lane 1 shows the marker.

3.3.2 Structural studies on ferritin HFt-PBM conformation

The transmission electron microscopy (TEM) examination demonstrated that the altered ferritin nanoparticle retained its distinctive spherical shape, measuring approximately 12 nm in diameter (see **Figure 3.16**). The electron microscopy analysis further revealed a uniform dispersion of the modified ferritin nanoparticle throughout the sample, confirming the successful synthesis and purification of the modified ferritin nanoparticle.

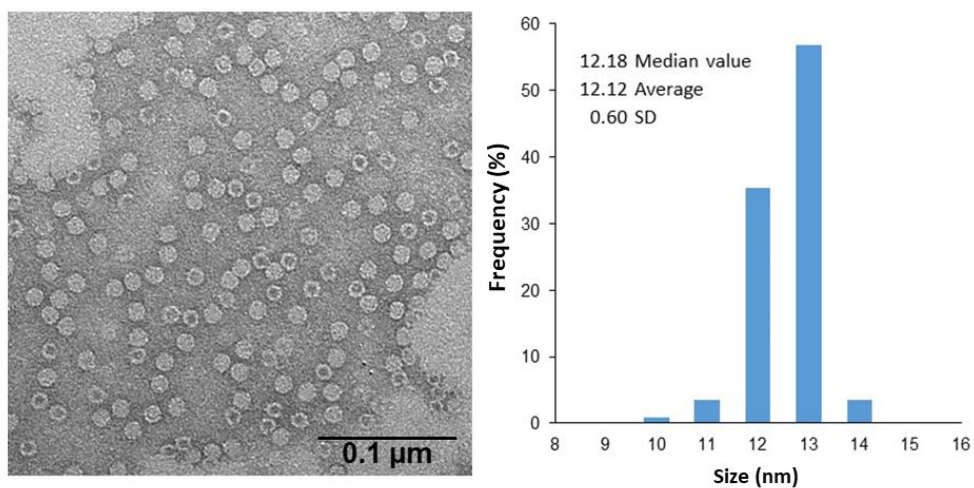


Figure 3.16. TEM imaging of purified HfT-PBM protein (**left panel**) and calculation of median and average diameter size (**right panel**).

Additionally, the HP-SEC analysis (see **Figure 3.17**) showed a higher molecular weight for the modified ferritin nanoparticle compared to the unmodified ferritin, indicating the successful attachment of the amino acid sequence to the ferritin nanoparticle.

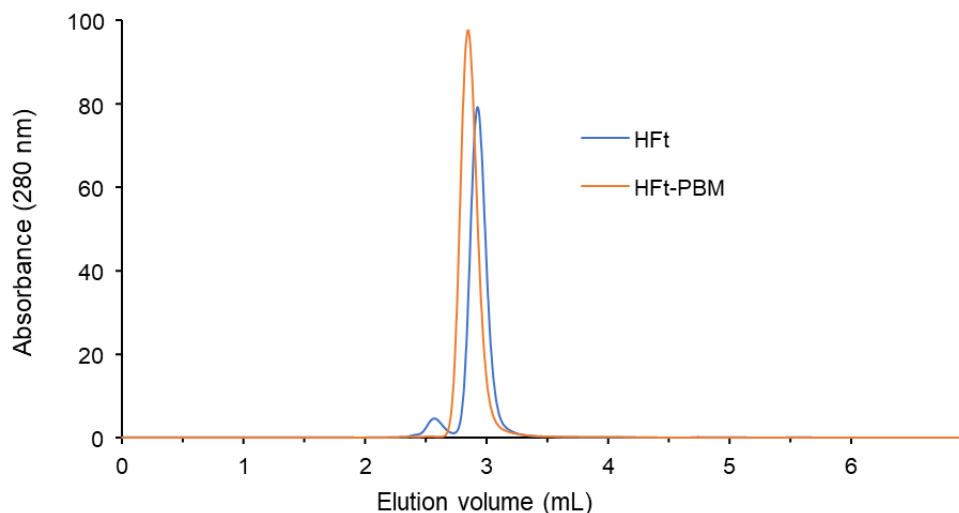


Figure 3.17. HP-SEC spectra of HfT-WT (blue line) and HfT-PBM (orange line). It is visible the difference in elution volume of the two proteins, in agreement with the fact that the modified ferritin has higher molecular weight.

Notably, the PCSK9-recognizing domain did not alter the overall structure of the ferritin nanoparticle, suggesting that the modification did not cause any significant changes to the protein's self-assembling behavior. This is a critical finding as it suggests that the modified ferritin retains its ability to self-assemble correctly, which is essential for maintaining its stability and preventing aggregation in solution.

3.3.3 Binding kinetics analysis

To determine the affinity of HfT-PBM towards PCSK9, a BLI assay was performed. In the experiment, his-tagged PCSK9 protein was immobilized on Nickel NTA biosensors' tip and then immersed in a solution containing the mutated ferritin. Time-courses are shown in **Figure 3.18**.

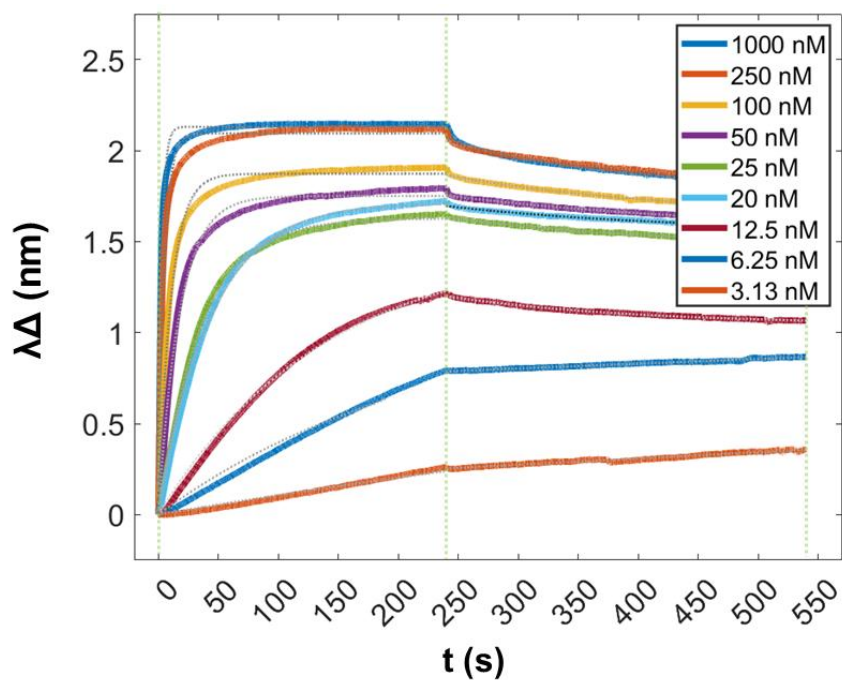


Figure 3.18. Time-courses of the reaction between and PCSK9 loaded on the biosensor and HFt-PBM present in solution at decreasing concentrations. The vertical dashed lines indicate the time interval of the binding step (240 s) and of the dissociation step (300 s).

From data analysis, the results show that HFt-PBM has a high affinity for the PCSK9 protein, with a dissociation constant K_D of 0.66 ± 0.30 nM. This suggests that the mutant has a strong binding ability to PCSK9, making it a promising candidate for use as a therapeutic agent for hypercholesterolemia.

In contrast to monoclonal antibodies, ferritin is an innate protein that can be taken up by cells through TfR1 receptor-mediated endocytosis. Given that HepG2 cells, human liver cancer cell line, possess this receptor on their surface, it is crucial to investigate whether the addition of the PBM molecule to the N-terminus of ferritin could impact its binding to the Tf receptor 1 (TfR1). This consideration holds particular significance because HFt-PBM needs to interact with extracellular PCSK9 to inhibit its interaction with the

LDL-receptor. The internalization of HFt-PBM via TfR1 may potentially reduce its effectiveness in lowering cholesterol levels, therefore an evaluation of the affinity strength with the His-tagged TfR1 receptor was performed. **Figure 3.19** displays the time-course acquisition of HFt-PBM protein binding to CD71 receptor loaded on Ni-NTA biosensors.

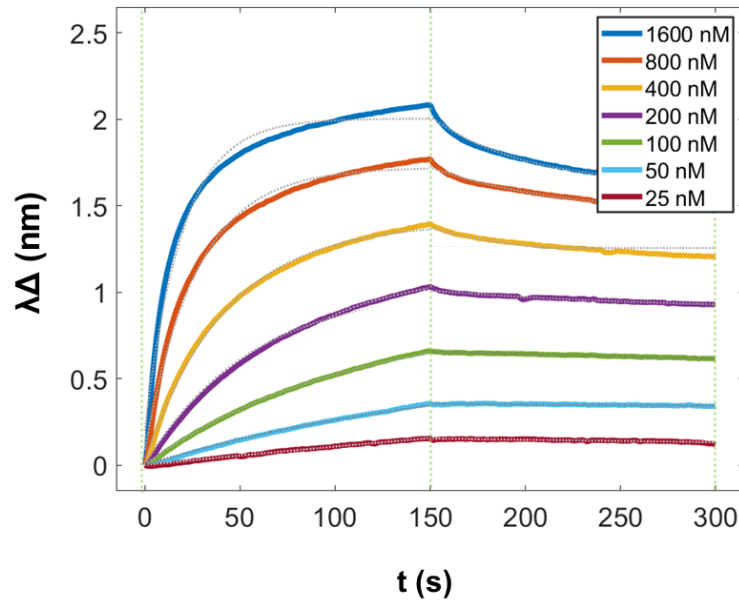


Figure 3.19. Time-courses of the reaction between HFt-PBM present in solution at decreasing concentrations and His-tagged CD71 loaded on the Ni-NTA biosensor tip. The vertical dashed lines indicate the time interval of the binding and of the dissociation steps, both lasting 150 s.

The results show a K_D value of 27.9 ± 0.4 nM for human ferritin. However, when the binding affinity between the wild-type ferritin protein and the transferrin receptor CD71 was tested, the K_D value was 9.42 ± 0.23 nM, meaning that HFt-WT shows higher affinity and it is consistent with the order of magnitude of the values reported in literature (Montemiglio et al., 2019). **Figure**

3.20 shows the experimental trends of the BLI experiment and **Table 5** gathers all the K_D measured in the three experiments.

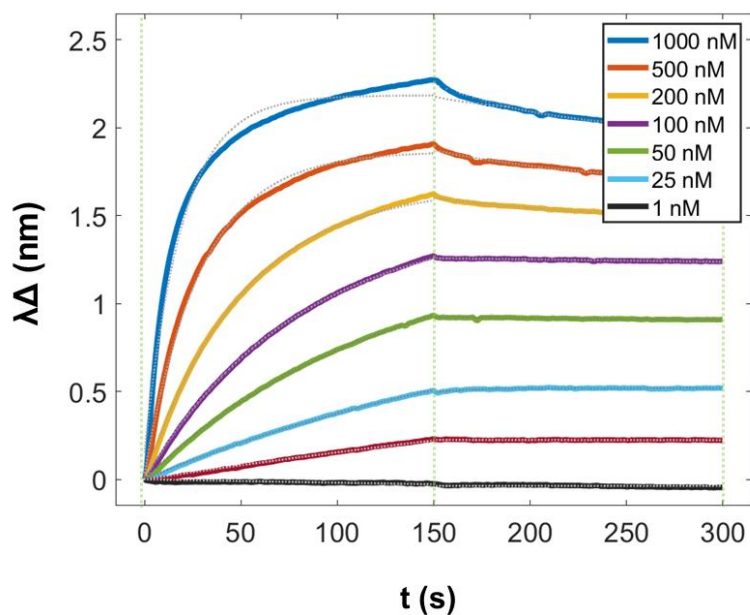


Figure 3.20. Time-courses of BLI measurements with CD71 loaded on Ni-NTA biosensors and HFt-WT in solution at different concentrations. Association and dissociation steps were set to 150 s.

Ferritin	ligand	K_D (nM)
HFt-PBM	PCSK9	0.66 ± 0.30
HFt-WT	CD71	9.42 ± 0.23
HFt-PBM	CD71	27.9 ± 0.4

Table 5. Values of K_D measured for all three combinations of interactions investigated. The ligand is the protein that was loaded on the biosensors, the ferritin column reports the analyte protein in solution.

These K_D values demonstrate that the engineered protein shows a greater affinity towards the PCSK9 protein compared to wild type ferritin, indicative of

the fact that PBM could indeed be exploited to interfere with PCSK9 metabolic pathway, but at the same time does not hinder the internalization of the “natural” human ferritin through the CD71 transferrin membrane receptors. This could be due to the fact that PCSK9, which has a weight of 75 kDa, strongly interacts with ferritin on multiple subunits simultaneously, effectively preventing its BC loop from freely interacting with Tfr1.

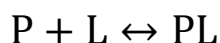
Chapter 4 Conclusion

The body of experimental evidence collected in the present dissertation serves primarily as a methodological workflow that addresses a key topic in protein science concerning the matching of innovative *in silico* methods with solution measurements dealing with interaction among proteins. As such, general conclusions can be drawn on protein-protein interactions together with specific examples of particular biological significance as commented in the following chapters.

4.1 Methods for investigating PPIs

At the basis of a greater understanding of the dense network of interactions that occur between proteins inside and outside the cellular environment there is continuous research and development of techniques that allow the characterization of these dynamics in a precise and least invasive way. Even before reaching *in vivo* testing, with the aid of predictions and simulations carried out *in silico* it is possible to characterize the propensity of a protein to interact with a partner by exploiting the physical and chemical aspects that define molecular systems. From a physics perspective, it is possible to comprehend these phenomena by employing a combination of quantum and statistical mechanics. Such an elucidation could be condensed into a collection of thermodynamic and kinetic relationships, which, in the end, account for the affinity and selectivity of a molecular interaction. From a chemical perspective, the aim is to understand the mechanism responsible for the free energy and kinetics of the binding. Various experimental techniques are available for quantifying the thermodynamics and kinetics of molecular binding. These methods can be categorized as biochemical (such as ELISA, enzymatic, and radioac-

tive assays) or biophysical (for example surface plasmon resonance, isothermal titration calorimetry, and FRET). All of these approaches yield empirical values for thermodynamic parameters (like K_D , IC_{50} , and EC_{50}) and kinetic constants (such as k_{off} and k_{on}). These experimental observables are quantitatively related to the free energy and the binding/unbinding kinetics. For instance, in a simple binding reaction between protein and ligand



the Gibbs binding free energy ΔG_{bind}^0 is directly related to the equilibrium concentration of bound ([PL]) and unbound ligand ([L]) and protein ([P]) complexes:

$$\Delta G_{bind}^0 = RT \ln K_D / C_0$$

where T is the temperature, R is the gas constant, C_0 is the standard state concentration of 1 mol/L, and K_D is the dissociation constant (previously defined in **Chapter 2**). In terms of equilibrium thermodynamics, the ergodic theorem then provides a suitable theoretical framework for linking the chemical world to the physical observables used to assess drug potency and efficacy. Specifically, for closed systems, the time-average of their properties equals the average across the entire system's configuration space. This offers insights into the statistical properties of a system in thermodynamic equilibrium. Molecular simulation can therefore integrate the microscopic and macroscopic approaches by estimating the time that the system spends in a certain microscopic state. This is becoming more doable thanks to efficient implementa-

tion of modern algorithms such as the Zernike algorithm, currently employed in the present dissertation.

Biolayer interferometry and nanoparticle enhanced turbidimetry techniques fit thoroughly into this context. In the experimental work exposed in the present dissertation, some examples of PPIs in solution studied with these innovative methods are therefore demonstrated, which, coupled with the structural data when available, provide a unique key to the interpretation of the major biological phenomena. BLI has proven to be a rapid and efficient technique with which it has been possible to study various molecular systems very different from each other in terms of the structure of the proteins in question, their dimensions, and the specificity and strength of the binding. The results relating to nanoparticle-enhanced turbidimetry are of particular interest, as the potential of this innovative technique is highlighted no longer in an immunoturbidimetry context, of which numerous examples can instead be found in the literature, but also as a mean to quantify the inhibitory potency of a protein, such as lactoferrin, in a binding process.

4.2 Lactoferrin binding properties

The interaction between SARS-CoV-2 Receptor-Binding Domain on the Spike protein and ACE2 human receptor represents a paradigmatic case of PPIs, especially when the molecular mechanism driving viral cell invasion are studied. In **Chapter 3.1** the inhibiting role of Lf in this interaction was investigated. The results have demonstrated that Lf shows a greater propensity to form binding with the ACE2 protein and on the contrary, it does not show any binding affinity towards the RBD protein up to the maximum physiological lactoferrin concentration range. Therefore, once the system with all

three proteins present had been analyzed experimentally, the detection of an inhibitory capacity in the formation of the RBD-ACE2 binding by Lf for concentrations greater than 1 μM was possible, bringing about a measurable, up to 300 -fold increase of the K_D value. These results are of particular relevance not only because they confirm the active role that lactoferrin can assume as an obstacle in viral entry in cells, but also because they experimentally show some characteristics of the protein from a biochemical point of view. In fact, in the literature the effects are often reported directly *in vivo* and rarely an in-depth analysis on the kinetics of molecular interaction involving this protein is present. Therefore, these results contribute to a more complete vision of the extra- and intra-cellular dynamics of lactoferrin.

4.3 The Zernike approach to the design of mutants

PPIs represent an intricate interplay involving many factors, comprehensive of electrostatic, hydrophobic, and geometric aspects. Among these, the long-range electrostatic component often serves as the main driving force. As distances decrease, the compatibility of shapes between interacting regions becomes fundamental, governing the stabilizing effect attributed to van der Waals interactions. Biological complexes, in particular, tend to display intermolecular interfaces characterized by a high degree of shape complementarity. Computational docking methods utilize measurements of shape complementarity as a guiding principle in their search algorithms. In this context, the 2D Zernike polynomial algorithm describes the shape of portions of the protein molecular surface by forming a complete basis in which any function of two variables defined in a unitary disc can be decomposed. The representation of the protein surface in terms of numerical vectors provided by Zernike

formalism, coupled with the possibility to pinpoint rotational invariant observables, allows a very thorough investigation of the shape and the complementary regions involved in molecular bindings.

Experimental results reported in **Chapter 3.2** are a representative example. An α -helix peptide belonging to the molecular surface on the interacting region of ACE2 was selected, to serve the scope as a viral inhibitor. Computational mutagenesis protocol was applied in order to identify the peptide sequence and to generate peptide mutants with a predicted increased or decreased binding affinity with respect to the wild type. Notably, both *in silico* and experimental *in vitro* tests confirm the effectiveness of the computational protocol in producing peptides with a controlled binding affinity toward its molecular partner. As such, the method can be used whenever a thermodynamic optimization of a PPI is desired, being it an antigen-antibody interaction or the complementarity between the couple of patches of a viral particle and its cellular receptor, allowing for a rapid and blind identification of the binding regions.

4.4 The multivalency of ferritin

The use of single peptides or small molecules for therapeutic purpose may bring with it few disadvantages, such as their susceptibility to rapid renal clearance and reduced bioavailability. To overcome these limitations, for the experiments carried out and exposed in **Chapters 3.2** and **3.3** a strategy was applied that consists in genetically fusing either ACE2 peptides or PCSK9 binding molecule at the N-terminus of each monomer of human H ferritin protein. The multivalent potency of ferritin, already vastly exploited in the field of drug delivery, has been also investigated in the context of SARS-

CoV-2 treatment or vaccine development and the fusion strategy helps benefiting from the multivalent effect conferred by the 24-meric ferritin structure. Indeed, it is important to highlight that ferritin presents several notable advantages when compared to monoclonal antibodies. Ferritin is a naturally occurring protein that is well-tolerated and metabolized by the human cells, in contrast to monoclonal antibodies, which can sometimes trigger adverse reactions. Also, the relatively small size of ferritin facilitates its internalization into cells and enables it to “find” its intracellular target effectively. Finally, the multifunctional nature of ferritin renders it a versatile platform that can be readily engineered to display various peptides or proteins. This versatility offers the potential to create multifunctional therapeutics capable of targeting multiple pathways, being them a viral entry process or pathways associated with hypercholesterolemia. In summary, the high affinity and the array of advantages associated with ferritin position it as a promising candidate for the development of innovative and effective therapies for managing a number of diseases.

Observing the results shown in **Chapter 3.2** , it is possible to infer that the strategy of fusing the high-affinity peptide HA1 to ferritin has led to a strong increase in the affinity with RBD, as measured by BLI. From the experiments carried out with the engineered ferritin HFt-PBM in **Chapter 3.3** , the conclusive results from the binding assays once again confirm the effectiveness of the fusion design protocol (a manuscript with the results is currently under writing). Altogether, the experimental work in the present dissertation shows the broad clinical applicability that the use of ferritin can ensure, and in general a very efficient procedure to hinder a very wide range of PPIs.

Bibliography

- Affatigato, L., Licciardi, M., Bonamore, A., Martorana, A., Incocciati, A., Boffi, A., & Militello, V. (2023). Ferritin-Coated SPIONs as New Cancer Cell Targeted Magnetic Nanocarrier. *Molecules*, 28(3), 1163. <https://doi.org/10.3390/molecules28031163>
- Ahamad, S., & Bhat, S. A. (2022). Recent Update on the Development of PCSK9 Inhibitors for Hypercholesterolemia Treatment. *Journal of Medicinal Chemistry*, 65(23), 15513–15539. <https://doi.org/10.1021/acs.jmedchem.2c01290>
- Anderson, B. F., Baker, H. M., Dodson, E. J., Norris, G. E., Rumball, S. V., Waters, J. M., & Baker, E. N. (1987). Structure of human lactoferrin at 3.2-Å resolution. *Proceedings of the National Academy of Sciences*, 84(7), 1769–1773. <https://doi.org/10.1073/pnas.84.7.1769>
- Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W., & Baker, E. N. (1989). Structure of human lactoferrin: Crystallographic structure analysis and refinement at 2.8 Å resolution. *Journal of Molecular Biology*, 209(4), 711–734. [https://doi.org/10.1016/0022-2836\(89\)90602-5](https://doi.org/10.1016/0022-2836(89)90602-5)
- Bai, C., Wang, J., Chen, G., Zhang, H., An, K., Xu, P., Du, Y., Ye, R. D., Saha, A., Zhang, A., & Warshel, A. (2021). Predicting Mutational Effects on Receptor Binding of the Spike Protein of SARS-CoV-2 Variants. *Journal of the American Chemical Society*, 143(42), 17646–17654. <https://doi.org/10.1021/jacs.1c07965>
- Baker, E. N., Anderson, B. F., Baker, H. M., Day, C. L., Haridas, M., Norris, G. E., Rumball, S. V., Smith, C. A., & Thomas, D. H. (1994). *Three-Dimensional Structure of Lactoferrin in Various Functional States* (pp. 1–12). https://doi.org/10.1007/978-1-4615-2548-6_1
- Baker, E. N., & Baker, H. M. (2005). Lactoferrin. *Cellular and Molecular Life Sciences*, 62(22), 2531–2539. <https://doi.org/10.1007/s00018-005-5368-9>

- Baker, H. M., & Baker, E. N. (2004). Lactoferrin and Iron: structural and dynamic aspects of binding and release. *BioMetals*, *17*(3), 209–216.
<https://doi.org/10.1023/B:BIOM.0000027694.40260.70>
- Barboza, M., Pinzon, J., Wickramasinghe, S., Froehlich, J. W., Moeller, I., Smilowitz, J. T., Ruhaak, L. R., Huang, J., Lönnerdal, B., German, J. B., Medrano, J. F., Weimer, B. C., & Lebrilla, C. B. (2012). Glycosylation of Human Milk Lactoferrin Exhibits Dynamic Changes During Early Lactation Enhancing Its Role in Pathogenic Bacteria-Host Interactions. *Molecular & Cellular Proteomics*, *11*(6), M111.015248.
<https://doi.org/10.1074/mcp.M111.015248>
- Barton, M. I., MacGowan, S. A., Kutuzov, M. A., Dushek, O., Barton, G. J., & van der Merwe, P. A. (2021). Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. *ELife*, *10*.
<https://doi.org/10.7554/eLife.70658>
- Berkel, P. H. C. van, Veen, H. A. van, Geerts, M. E. J., Boer, H. A. de, & Nuijens, J. H. (1996). Heterogeneity in utilization of N-glycosylation sites Asn624 and Asn138 in human lactoferrin: a study with glycosylation-site mutants. *Biochemical Journal*, *319*(1), 117–122.
<https://doi.org/10.1042/bj3190117>
- Berlov, M. N., Korableva, E. S., Andreeva, Yu. V., Ovchinnikova, T. V., & Kokryakov, V. N. (2007). Lactoferrin from canine neutrophils: Isolation and physicochemical and antimicrobial properties. *Biochemistry (Moscow)*, *72*(4), 445–451. <https://doi.org/10.1134/S0006297907040128>
- Brender, J. R., & Zhang, Y. (2015). Predicting the Effect of Mutations on Protein-Protein Binding Interactions through Structure-Based Interface Profiles. *PLOS Computational Biology*, *11*(10), e1004494.
<https://doi.org/10.1371/journal.pcbi.1004494>
- Calisti, L., Trabuco, M. C., Boffi, A., Testi, C., Montemiglio, L. C., des Georges, A., Benni, I., Ilari, A., Taciak, B., Białasek, M., Rygiel, T., Król, M., Baiocco, P., & Bonamore, A. (2018). Engineered ferritin for

lanthanide binding. *PLOS ONE*, *13*(8), e0201859.
<https://doi.org/10.1371/journal.pone.0201859>

- Campione, E., Cosio, T., Rosa, L., Lanna, C., Di Girolamo, S., Gaziano, R., Valenti, P., & Bianchi, L. (2020). Lactoferrin as protective natural barrier of respiratory and intestinal mucosa against coronavirus infection and inflammation. *International Journal of Molecular Sciences*, *21*(14), 4903.
- Campione, E., Lanna, C., Cosio, T., Rosa, L., Conte, M. P., Iacovelli, F., Romeo, A., Falconi, M., Del Vecchio, C., Franchin, E., Lia, M. S., Minieri, M., Chiamonte, C., Ciotti, M., Nuccetelli, M., Terrinoni, A., Iannuzzi, I., Coppeda, L., Magrini, A., ... Bianchi, L. (2021). Lactoferrin Against SARS-CoV-2: In Vitro and In Silico Evidences. *Frontiers in Pharmacology*, *12*. <https://doi.org/10.3389/fphar.2021.666600>
- Concepcion, J., Witte, K., Wartchow, C., Choo, S., Yao, D., Persson, H., Wei, J., Li, P., Heidecker, B., Ma, W., Varma, R., Zhao, L.-S., Perillat, D., Carricato, G., Recknor, M., Du, K., Ho, H., Ellis, T., Gamez, J., ... Tan, H. (2009). Label-Free Detection of Biomolecular Interactions Using BioLayer Interferometry for Kinetic Characterization. *Combinatorial Chemistry & High Throughput Screening*, *12*(8), 791–800.
<https://doi.org/10.2174/138620709789104915>
- Cunningham, D., Danley, D. E., Geoghegan, K. F., Griffor, M. C., Hawkins, J. L., Subashi, T. A., Varghese, A. H., Ammirati, M. J., Culp, J. S., Hoth, L. R., Mansour, M. N., McGrath, K. M., Seddon, A. P., Shenolikar, S., Stutzman-Engwall, K. J., Warren, L. C., Xia, D., & Qiu, X. (2007). Structural and biophysical studies of PCSK9 and its mutants linked to familial hypercholesterolemia. *Nature Structural & Molecular Biology*, *14*(5), 413–419. <https://doi.org/10.1038/nsmb1235>
- Daberduku, S., & Ferrari, C. (2018). Exploring the potential of 3D Zernike descriptors and SVM for protein–protein interface prediction. *BMC Bioinformatics*, *19*(1), 35. <https://doi.org/10.1186/s12859-018-2043-3>

- Daberdaku, S., & Ferrari, C. (2019). Antibody interface prediction with 3D Zernike descriptors and SVM. *Bioinformatics*, *35*(11), 1870–1876. <https://doi.org/10.1093/bioinformatics/bty918>
- de Campos, L. J., Palermo, N. Y., & Conda-Sheridan, M. (2021). Targeting SARS-CoV-2 Receptor Binding Domain with Stapled Peptides: An *In Silico* Study. *The Journal of Physical Chemistry B*, *125*(24), 6572–6586. <https://doi.org/10.1021/acs.jpcc.1c02398>
- De Lauro, A., Di Rienzo, L., Miotto, M., Olimpieri, P. P., Milanetti, E., & Ruocco, G. (2022). Shape Complementarity Optimization of Antibody–Antigen Interfaces: The Application to SARS-CoV-2 Spike Protein. *Frontiers in Molecular Biosciences*, *9*. <https://doi.org/10.3389/fmolb.2022.874296>
- Desantis, F., Miotto, M., Di Rienzo, L., Milanetti, E., & Ruocco, G. (2022). Spatial organization of hydrophobic and charged residues affects protein thermal stability and binding affinity. *Scientific Reports*, *12*(1), 12087. <https://doi.org/10.1038/s41598-022-16338-5>
- Di Rienzo, L., De Flaviis, L., Ruocco, G., Folli, V., & Milanetti, E. (2022). Binding site identification of G protein-coupled receptors through a 3D Zernike polynomials-based method: application to *C. elegans* olfactory receptors. *Journal of Computer-Aided Molecular Design*, *36*(1), 11–24. <https://doi.org/10.1007/s10822-021-00434-1>
- Di Rienzo, L., Milanetti, E., Testi, C., Montemiglio, L. C., Baiocco, P., Boffi, A., & Ruocco, G. (2020). A novel strategy for molecular interfaces optimization: The case of Ferritin-Transferrin receptor interaction. *Computational and Structural Biotechnology Journal*, *18*, 2678–2686. <https://doi.org/10.1016/j.csbj.2020.09.020>
- Di Rienzo, L., Miotto, M., Desantis, F., Grassmann, G., Ruocco, G., & Milanetti, E. (2023). Dynamical changes of SARS-CoV-2 spike variants in the highly immunogenic regions impact the viral antibodies escaping. *Proteins: Structure, Function, and Bioinformatics*. <https://doi.org/10.1002/prot.26497>

- Di Rienzo, L., Miotto, M., Milanetti, E., & Ruocco, G. (2023). Computational structural-based GPCR optimization for user-defined ligand: Implications for the development of biosensors. *Computational and Structural Biotechnology Journal*, *21*, 3002–3009. <https://doi.org/10.1016/j.csbj.2023.05.004>
- Di Rienzo, L., Monti, M., Milanetti, E., Miotto, M., Boffi, A., Tartaglia, G. G., & Ruocco, G. (2021). Computational optimization of angiotensin-converting enzyme 2 for SARS-CoV-2 Spike molecular recognition. *Computational and Structural Biotechnology Journal*, *19*, 3006–3014. <https://doi.org/10.1016/j.csbj.2021.05.016>
- Dzimianski, J. V., Lorig-Roach, N., O'Rourke, S. M., Alexander, D. L., Kimmey, J. M., & DuBois, R. M. (2020). Rapid and sensitive detection of SARS-CoV-2 antibodies by bilayer interferometry. *Scientific Reports*, *10*(1), 21738. <https://doi.org/10.1038/s41598-020-78895-x>
- Fan, K., Zhou, M., & Yan, X. (2017). Questions about horse spleen ferritin crossing the blood brain barrier via mouse transferrin receptor 1. *Protein & Cell*, *8*(11), 788–790. <https://doi.org/10.1007/s13238-017-0481-8>
- Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C. Y., Jia, M., De, T., Teague, J. W., Stratton, M. R., McDermott, U., & Campbell, P. J. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research*, *43*(D1), D805–D811. <https://doi.org/10.1093/nar/gku1075>
- Geng, C., Xue, L. C., Roel-Touris, J., & Bonvin, A. M. J. J. (2019). Finding the $\Delta\Delta G$ spot: Are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it? *WIREs Computational Molecular Science*, *9*(5). <https://doi.org/10.1002/wcms.1410>
- Gheblawi, M., Wang, K., Viveiros, A., Nguyen, Q., Zhong, J.-C., Turner, A. J., Raizada, M. K., Grant, M. B., & Oudit, G. Y. (2020). Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System. *Circulation Research*, *126*(10), 1456–1474. <https://doi.org/10.1161/CIRCRESAHA.120.317015>

- Ginex, T., Marco-Marín, C., Wieczór, M., Mata, C. P., Krieger, J., Ruiz-Rodriguez, P., López-Redondo, M. L., Francés-Gómez, C., Melero, R., Sánchez-Sorzano, C. Ó., Martínez, M., Gougéard, N., Forcada-Nadal, A., Zamora-Caballero, S., Gozalbo-Rovira, R., Sanz-Frasquet, C., Arranz, R., Bravo, J., Rubio, V., ... Carazo, J.-M. (2022). The structural role of SARS-CoV-2 genetic background in the emergence and success of spike mutations: The case of the spike A222V mutation. *PLOS Pathogens*, *18*(7), e1010631. <https://doi.org/10.1371/journal.ppat.1010631>
- Grishin, A. M., Dolgova, N. V., Landreth, S., Fisette, O., Pickering, I. J., George, G. N., Falzarano, D., & Cygler, M. (2022). Disulfide Bonds Play a Critical Role in the Structure and Function of the Receptor-binding Domain of the SARS-CoV-2 Spike Antigen. *Journal of Molecular Biology*, *434*(2), 167357. <https://doi.org/10.1016/j.jmb.2021.167357>
- Hamdy, M. E., El-Deeb, A. H., Hagag, N. M., Shahein, M. A., Alaidi, O., & Hussein, H. A. (2022). Mutations of the SARS-CoV-2 Spike Glycoprotein Detected in Cats and Their Effect on Its Structure and Function. *Frontiers in Cellular and Infection Microbiology*, *12*. <https://doi.org/10.3389/fcimb.2022.875123>
- Hampton, E. N., Knuth, M. W., Li, J., Harris, J. L., Lesley, S. A., & Spraggon, G. (2007). The self-inhibited structure of full-length PCSK9 at 1.9 Å reveals structural homology with resistin within the C-terminal domain. *Proceedings of the National Academy of Sciences*, *104*(37), 14604–14609. <https://doi.org/10.1073/pnas.0703402104>
- Haridas, M., Anderson, B. F., & Baker, E. N. (1995). Structure of human diferric lactoferrin refined at 2.2 Å resolution. *Acta Crystallographica Section D Biological Crystallography*, *51*(5), 629–646. <https://doi.org/10.1107/S0907444994013521>
- Harrison, P. M., & Arosio, P. (1996). The ferritins: molecular properties, iron storage function and cellular regulation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1275*(3), 161–203. [https://doi.org/10.1016/0005-2728\(96\)00022-9](https://doi.org/10.1016/0005-2728(96)00022-9)

- Honarmand Ebrahimi, K., Hagedoorn, P.-L., & Hagen, W. R. (2015). Unity in the Biochemistry of the Iron-Storage Proteins Ferritin and Bacterioferritin. *Chemical Reviews*, *115*(1), 295–326. <https://doi.org/10.1021/cr5004908>
- Huang, Y., Yang, C., Xu, X., Xu, W., & Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica*, *41*(9), 1141–1149. <https://doi.org/10.1038/s41401-020-0485-4>
- Incocciati, A., Bertuccini, L., Boffi, A., Macone, A., & Bonamore, A. (2022). Unlocking the Treasure Box: The Role of HEPES Buffer in Disassembling an Uncommon Ferritin Nanoparticle. *Separations*, *9*(8), 222. <https://doi.org/10.3390/separations9080222>
- Jeong, W., Bu, J., Kubiawicz, L. J., Chen, S. S., Kim, Y., & Hong, S. (2018). Peptide–nanoparticle conjugates: a next generation of diagnostic and therapeutic platforms? *Nano Convergence*, *5*(1), 38. <https://doi.org/10.1186/s40580-018-0170-1>
- Kalathiya, U., Padariya, M., Fahraeus, R., Chakraborti, S., & Hupp, T. R. (2021). Multivalent Display of SARS-CoV-2 Spike (RBD Domain) of COVID-19 to Nanomaterial, Protein Ferritin Nanocages. *Biomolecules*, *11*(2), 297. <https://doi.org/10.3390/biom11020297>
- Kamat, V., & Rafique, A. (2017). Designing binding kinetic assay on the biolayer interferometry (BLI) biosensor to characterize antibody-antigen interactions. *Analytical Biochemistry*, *536*, 16–31. <https://doi.org/10.1016/j.ab.2017.08.002>
- Karav, S., German, J., Rouquié, C., Le Parc, A., & Barile, D. (2017). Studying Lactoferrin N-Glycosylation. *International Journal of Molecular Sciences*, *18*(4), 870. <https://doi.org/10.3390/ijms18040870>
- Karthikeyan, S., Paramasivam, M., Yadav, S., Srinivasan, A., & Singh, T. P. (1999). Structure of buffalo lactoferrin at 2.5 Å resolution using crystals grown at 303 K shows different orientations of the N and C lobes. *Acta Crystallographica Section D Biological Crystallography*, *55*(11), 1805–1813. <https://doi.org/10.1107/S0907444999010951>

- Karthikeyan, S., Yadav, S., Paramasivam, M., Srinivasan, A., & Singh, T. P. (2000). Structure of buffalo lactoferrin at 3.3 Å resolution at 277 K. *Acta Crystallographica Section D Biological Crystallography*, 56(6), 684–689. <https://doi.org/10.1107/S0907444900005151>
- Kell, D. B., Heyden, E. L., & Pretorius, E. (2020). The Biology of Lactoferrin, an Iron-Binding Protein That Can Help Defend Against Viruses and Bacteria. *Frontiers in Immunology*, 11. <https://doi.org/10.3389/fimmu.2020.01221>
- Khan, J. A., Kumar, P., Paramasivam, M., Yadav, R. S., Sahani, M. S., Sharma, S., Srinivasan, A., & Singh, T. P. (2001). Camel Lactoferrin, a Transferrin-cum-Lactoferrin: Crystal Structure of Camel Apolactoferrin at 2.6Å Resolution and Structural Basis of its Dual Role. *Journal of Molecular Biology*, 309(3), 751–761. <https://doi.org/10.1006/jmbi.2001.4692>
- Khoshnejad, M., Parhiz, H., Shuvaev, V. V., Dmochowski, I. J., & Muzykantov, V. R. (2018). Ferritin-based drug delivery systems: Hybrid nanocarriers for vascular immunotargeting. *Journal of Controlled Release*, 282, 13–24. <https://doi.org/10.1016/j.jconrel.2018.02.042>
- Kim, J., Heu, W., Jeong, S., & Kim, H.-S. (2017). Genetically functionalized ferritin nanoparticles with a high-affinity protein binder for immunoassay and imaging. *Analytica Chimica Acta*, 988, 81–88. <https://doi.org/10.1016/j.aca.2017.07.060>
- Kim, S. A., Kim, S., Kim, G. B., Goo, J., Kim, N., Lee, Y., Nam, G.-H., Lim, S., Kim, T., Chang, K. H., Lee, T. G., Kim, I.-S., & Lee, E. J. (2022). A Multivalent Vaccine Based on Ferritin Nanocage Elicits Potent Protective Immune Responses against SARS-CoV-2 Mutations. *International Journal of Molecular Sciences*, 23(11), 6123. <https://doi.org/10.3390/ijms23116123>
- Kowalczyk, P., Kaczyńska, K., Kleczkowska, P., Bukowska-Ośko, I., Kramkowski, K., & Sulejczak, D. (2022). The Lactoferrin Phenomenon—A Miracle Molecule. *Molecules*, 27(9), 2941. <https://doi.org/10.3390/molecules27092941>

- Kumaraswamy, S., & Tobias, R. (2015). *Label-Free Kinetic Analysis of an Antibody–Antigen Interaction Using Biolayer Interferometry* (pp. 165–182). https://doi.org/10.1007/978-1-4939-2425-7_10
- Lambert, G., Charlton, F., Rye, K.-A., & Piper, D. E. (2009). Molecular basis of PCSK9 function. *Atherosclerosis*, *203*(1), 1–7. <https://doi.org/10.1016/j.atherosclerosis.2008.06.010>
- Lambert, L. A. (2012). Molecular evolution of the transferrin family and associated receptors. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1820*(3), 244–255. <https://doi.org/10.1016/j.bbagen.2011.06.002>
- Lambert, L. A., Perri, H., Halbrooks, P. J., & Mason, A. B. (2005). Evolution of the transferrin family: Conservation of residues associated with iron and anion binding. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *142*(2), 129–141. <https://doi.org/10.1016/j.cbpb.2005.07.007>
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*(7807), 215–220. <https://doi.org/10.1038/s41586-020-2180-5>
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., & Maglott, D. R. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research*, *44*(D1), D862–D868. <https://doi.org/10.1093/nar/gkv1222>
- Larue, R. C., Xing, E., Kenney, A. D., Zhang, Y., Tuazon, J. A., Li, J., Yount, J. S., Li, P.-K., & Sharma, A. (2021). Rationally Designed ACE2-Derived Peptides Inhibit SARS-CoV-2. *Bioconjugate Chemistry*, *32*(1), 215–223. <https://doi.org/10.1021/acs.bioconjchem.0c00664>
- Legrand, D., Mazurier, J., Colavizza, D., Montreuil, J., & Spik, G. (1990). Properties of the iron-binding site of the N-terminal lobe of human and bovine lactotransferrins. Importance of the glycan moiety and of the non-covalent interactions between the N- and C-terminal lobes in the

- stability of the iron-binding site. *The Biochemical Journal*, 266(2), 575–581.
- Levi, S., Luzzago, A., Franceschinelli, F., Santambrogio, P., Cesareni, G., & Arosio, P. (1989). Mutational analysis of the channel and loop sequences of human ferritin H-chain. *Biochemical Journal*, 264(2), 381–388. <https://doi.org/10.1042/bj2640381>
- Levi, S., Santambrogio, P., Cozzi, A., Rovida, E., Corsi, B., Tamborini, E., Spada, S., Albertini, A., & Arosio, P. (1994). The Role of the L-Chain in Ferritin Iron Incorporation. *Journal of Molecular Biology*, 238(5), 649–654. <https://doi.org/10.1006/jmbi.1994.1325>
- Li, Z., & Furmanski, P. (1995). Role of sialic acid residues in iron binding by human Lactoferrin - α . *Chinese Journal of Cancer Research*, 7(2), 79–85. <https://doi.org/10.1007/BF03014401>
- Liu, C., Chen, J., Chen, H., Zhang, T., He, D., Luo, Q., Chi, J., Hong, Z., Liao, Y., Zhang, S., Wu, Q., Cen, H., Chen, G., Li, J., & Wang, L. (2022). PCSK9 Inhibition: From Current Advances to Evolving Future. *Cells*, 11(19), 2972. <https://doi.org/10.3390/cells11192972>
- Malatesta, F. (2005). The study of bimolecular reactions under non-pseudo-first order conditions. *Biophysical Chemistry*, 116(3), 251–256. <https://doi.org/10.1016/j.bpc.2005.04.006>
- Mason, A. B., Halbrooks, P. J., James, N. G., Connolly, S. A., Larouche, J. R., Smith, V. C., MacGillivray, R. T. A., & Chasteen, N. D. (2005). Mutational Analysis of C-Lobe Ligands of Human Serum Transferrin: Insights into the Mechanism of Iron Release. *Biochemistry*, 44(22), 8013–8021. <https://doi.org/10.1021/bi050015f>
- Masson, P. L., Heremans, J. F., & Dive, C. H. (1966). An iron-binding protein common to many external secretions. *Clinica Chimica Acta*, 14(6), 735–739. [https://doi.org/10.1016/0009-8981\(66\)90004-0](https://doi.org/10.1016/0009-8981(66)90004-0)
- Milanetti, E., Miotto, M., Di Rienzo, L., Monti, M., Gosti, G., & Ruocco, G. (2021). 2D Zernike polynomial expansion: Finding the protein-protein

- binding regions. *Computational and Structural Biotechnology Journal*, 19, 29–36. <https://doi.org/10.1016/j.csbj.2020.11.051>
- Milanetti, E., Miotto, M., Di Rienzo, L., Nagaraj, M., Monti, M., Golbek, T. W., Gosti, G., Roeters, S. J., Weidner, T., Otzen, D. E., & Ruocco, G. (2021). In-Silico Evidence for a Two Receptor Based Strategy of SARS-CoV-2. *Frontiers in Molecular Biosciences*, 8. <https://doi.org/10.3389/fmolb.2021.690655>
- Miotto, M., Di Rienzo, L., Bò, L., Boffi, A., Ruocco, G., & Milanetti, E. (2021). Molecular Mechanisms Behind Anti SARS-CoV-2 Action of Lactoferrin. *Frontiers in Molecular Biosciences*, 8. <https://doi.org/10.3389/fmolb.2021.607443>
- Miotto, M., Di Rienzo, L., Gosti, G., Bo', L., Parisi, G., Piacentini, R., Boffi, A., Ruocco, G., & Milanetti, E. (2022). Inferring the stabilization effects of SARS-CoV-2 variants on the binding with ACE2 receptor. *Communications Biology*, 5(1), 20221. <https://doi.org/10.1038/s42003-021-02946-w>
- Miotto, M., Di Rienzo, L., Gosti, G., Milanetti, E., & Ruocco, G. (2021). Does blood type affect the COVID-19 infection pattern? *Plos One*, 16(5), e0251535.
- Montemiglio, L. C., Testi, C., Ceci, P., Falvo, E., Pitea, M., Savino, C., Arcovito, A., Peruzzi, G., Baiocco, P., Mancina, F., Boffi, A., des Georges, A., & Vallone, B. (2019). Cryo-EM structure of the human ferritin–transferrin receptor 1 complex. *Nature Communications*, 10(1), 1121. <https://doi.org/10.1038/s41467-019-09098-w>
- Moore, S. A., Anderson, B. F., Groom, C. R., Haridas, M., & Baker, E. N. (1997). Three-dimensional structure of diferric bovine lactoferrin at 2.8 Å resolution. *Journal of Molecular Biology*, 274(2), 222–236. <https://doi.org/10.1006/jmbi.1997.1386>
- Moretti, R., Fleishman, S. J., Agius, R., Torchala, M., Bates, P. A., Kastritis, P. L., Rodrigues, J. P. G. L. M., Trellet, M., Bonvin, A. M. J. J., Cui, M., Rومان, M., Gillis, D., Dehouck, Y., Moal, I., Romero-Durana, M., Perez-Cano, L., Pallara, C., Jimenez, B., Fernandez-Recio, J., ... Baker, D.

- (2013). Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. *Proteins: Structure, Function, and Bioinformatics*, 81(11), 1980–1987.
<https://doi.org/10.1002/prot.24356>
- Müller-Esparza, H., Osorio-Valeriano, M., Steube, N., Thanbichler, M., & Randau, L. (2020). Bio-Layer Interferometry Analysis of the Target Binding Activity of CRISPR-Cas Effector Complexes. *Frontiers in Molecular Biosciences*, 7. <https://doi.org/10.3389/fmolb.2020.00098>
- Palombarini, F., Di Fabio, E., Boffi, A., Macone, A., & Bonamore, A. (2020). Ferritin Nanocages for Protein Delivery to Tumor Cells. *Molecules*, 25(4), 825. <https://doi.org/10.3390/molecules25040825>
- Palombarini, F., Ghirga, F., Boffi, A., Macone, A., & Bonamore, A. (2019). Application of crossflow ultrafiltration for scaling up the purification of a recombinant ferritin. *Protein Expression and Purification*, 163, 105451. <https://doi.org/10.1016/j.pep.2019.105451>
- Papageorgiou, A. C., & Mohsin, I. (2020). The SARS-CoV-2 Spike Glycoprotein as a Drug and Vaccine Target: Structural Insights into Its Complexes with ACE2 and Antibodies. *Cells*, 9(11), 2343. <https://doi.org/10.3390/cells9112343>
- Park, I., Schaeffer, E., Sidoli, A., Baralle, F. E., Cohen, G. N., & Zakin, M. M. (1985). Organization of the human transferrin gene: direct evidence that it originated by gene duplication. *Proceedings of the National Academy of Sciences*, 82(10), 3149–3153. <https://doi.org/10.1073/pnas.82.10.3149>
- Petersen, R. (2017). Strategies Using Bio-Layer Interferometry Biosensor Technology for Vaccine Research and Development. *Biosensors*, 7(4), 49. <https://doi.org/10.3390/bios7040049>
- Piacentini, R., Centi, L., Miotto, M., Milanetti, E., Di Rienzo, L., Pitea, M., Piazza, P., Ruocco, G., Boffi, A., & Parisi, G. (2022). Lactoferrin Inhibition of the Complex Formation between ACE2 Receptor and SARS CoV-2 Recognition Binding Domain. *International Journal of Molecular Sciences*, 23(10). <https://doi.org/10.3390/ijms23105436>

- Rascón-Cruz, Q., Espinoza-Sánchez, E. A., Siqueiros-Cendón, T. S., Nakamura-Bencomo, S. I., Arévalo-Gallegos, S., & Iglesias-Figueroa, B. F. (2021). Lactoferrin: A Glycoprotein Involved in Immunomodulation, Anticancer, and Antimicrobial Processes. *Molecules*, *26*(1), 205. <https://doi.org/10.3390/molecules26010205>
- Rodrigues, C. H. M., Myung, Y., Pires, D. E. V., & Ascher, D. B. (2019). mCSM-PPI2: predicting the effects of mutations on protein–protein interactions. *Nucleic Acids Research*, *47*(W1), W338–W344. <https://doi.org/10.1093/nar/gkz383>
- Saponaro, F., Rutigliano, G., Sestito, S., Bandini, L., Storti, B., Bizzarri, R., & Zucchi, R. (2020). ACE2 in the Era of SARS-CoV-2: Controversies and Novel Perspectives. *Frontiers in Molecular Biosciences*, *7*. <https://doi.org/10.3389/fmolb.2020.588618>
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221–224. <https://doi.org/10.1038/s41586-020-2179-y>
- Sharma, A. K., Paramasivam, M., Srinivasan, A., Yadav, M. P., & Singh, T. P. (1999). Three-dimensional structure of mare diferric lactoferrin at 2.6 Å resolution. *Journal of Molecular Biology*, *289*(2), 303–317. <https://doi.org/10.1006/jmbi.1999.2767>
- Sinegubova, M. V., Orlova, N. A., Kovnir, S. V., Dayanova, L. K., & Vorobiev, I. I. (2021). High-level expression of the monomeric SARS-CoV-2 S protein RBD 320-537 in stably transfected CHO cells by the EEF1A1-based plasmid vector. *PLOS ONE*, *16*(2), e0242890. <https://doi.org/10.1371/journal.pone.0242890>
- Song, N., Zhang, J., Zhai, J., Hong, J., Yuan, C., & Liang, M. (2021). Ferritin: A Multifunctional Nanoplatfor for Biological Detection, Imaging Diagnosis, and Drug Delivery. *Accounts of Chemical Research*, *54*(17), 3313–3325. <https://doi.org/10.1021/acs.accounts.1c00267>
- Stefanini, S., Vecchini, P., & Chiancone, E. (1987). On the mechanism of horse spleen apoferritin assembly: a sedimentation velocity and circular

- dichroism study. *Biochemistry*, 26(7), 1831–1837.
<https://doi.org/10.1021/bi00381a007>
- Suzuki, Y. A., Lopez, V., & Lönnnerdal, B. (2005). Lactoferrin. *Cellular and Molecular Life Sciences*, 62(22), 2560–2575.
<https://doi.org/10.1007/s00018-005-5371-1>
- Tombling, B. J., Zhang, Y., Huang, Y.-H., Craik, D. J., & Wang, C. K. (2021). The emerging landscape of peptide-based inhibitors of PCSK9. *Atherosclerosis*, 330, 52–60.
<https://doi.org/10.1016/j.atherosclerosis.2021.06.903>
- Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Boons, G.-J., Bosch, B.-J., Rey, F. A., de Groot, R. J., & Veerler, D. (2019). Structural basis for human coronavirus attachment to sialic acid receptors. *Nature Structural & Molecular Biology*, 26(6), 481–489.
<https://doi.org/10.1038/s41594-019-0233-y>
- Tzotzos, S. (2022). Stapled peptides as potential inhibitors of SARS-CoV-2 binding to the hACE2 receptor. *Journal of Peptide Science*, 28(9).
<https://doi.org/10.1002/psc.3409>
- Uchida, M., Kang, S., Reichhardt, C., Harlen, K., & Douglas, T. (2010). The ferritin superfamily: Supramolecular templates for materials synthesis. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1800(8), 834–845. <https://doi.org/10.1016/j.bbagen.2009.12.005>
- Van Veen, H. A., Geerts, M. E. J., Van Berkel, P. H. C., & Nuijens, J. H. (2004). The role of N-linked glycosylation in the protection of human and bovine lactoferrin against tryptic proteolysis. *European Journal of Biochemistry*, 271(4), 678–684. <https://doi.org/10.1111/j.1432-1033.2003.03965.x>
- Vangone, A., & Bonvin, A. M. (2015). Contacts-based prediction of binding affinity in protein–protein complexes. *ELife*, 4.
<https://doi.org/10.7554/eLife.07454>

- Venkatraman, V., Yang, Y. D., Sael, L., & Kihara, D. (2009). Protein-protein docking using region-based 3D Zernike descriptors. *BMC Bioinformatics*, *10*(1), 407. <https://doi.org/10.1186/1471-2105-10-407>
- Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veerler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, *181*(2), 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
- Wang, R., Huang, H., Yu, C., Sun, C., Ma, J., Kong, D., Lin, Y., Zhao, D., Zhou, S., Lu, J., Cao, S., Zhang, Y., Luo, C., Li, X., Wang, Y., & Xie, L. (2022). A spike-trimer protein-based tetravalent COVID-19 vaccine elicits enhanced breadth of neutralization against SARS-CoV-2 Omicron subvariants and other variants. *Science China Life Sciences*. <https://doi.org/10.1007/s11427-022-2207-7>
- Wang, Y., Wang, P., Wang, H., Luo, Y., Wan, L., Jiang, M., & Chu, Y. (2020). Lactoferrin for the treatment of COVID-19. *Experimental and Therapeutic Medicine*, *20*(6), 1.
- Wei, Z., Nishimura, T., & Yoshida, S. (2000). Presence of a Glycan at a Potential N-Glycosylation Site, Asn-281, of Bovine Lactoferrin. *Journal of Dairy Science*, *83*(4), 683–689. [https://doi.org/10.3168/jds.S0022-0302\(00\)74929-0](https://doi.org/10.3168/jds.S0022-0302(00)74929-0)
- Wesener, D. A., Wangkanont, K., McBride, R., Song, X., Kraft, M. B., Hodges, H. L., Zarling, L. C., Splain, R. A., Smith, D. F., Cummings, R. D., Paulson, J. C., Forest, K. T., & Kiessling, L. L. (2015). Recognition of microbial glycans by human intelectin-1. *Nature Structural & Molecular Biology*, *22*(8), 603–610. <https://doi.org/10.1038/nsmb.3053>
- Yadav, R., Govindan, S., Daczkowski, C., Mesecar, A., Chakravarthy, S., & Noinaj, N. (2021). Structural insight into the dual function of LbpB in mediating Neisserial pathogenesis. *ELife*, *10*. <https://doi.org/10.7554/eLife.71683>
- Ye, X.-Y., Nishimura, T., & Yoshida, S. (1997). Characterization of the Protein and Glycan Moieties in Different Forms of Bovine Lactoferrin. *Bio-*

science, Biotechnology, and Biochemistry, 61(5), 782–786.
<https://doi.org/10.1271/bbb.61.782>

Zlatina, K., & Galuska, S. P. (2021). The *N*-glycans of lactoferrin: more than just a sweet decoration. *Biochemistry and Cell Biology*, 99(1), 117–127.
<https://doi.org/10.1139/bcb-2020-0106>

Zou, J., Yin, J., Fang, L., Yang, M., Wang, T., Wu, W., Bellucci, M. A., & Zhang, P. (2020). Computational Prediction of Mutational Effects on SARS-CoV-2 Binding by Relative Free Energy Calculations. *Journal of Chemical Information and Modeling*, 60(12), 5794–5802.
<https://doi.org/10.1021/acs.jcim.0c00679>

List of abbreviations

ACE2	Angiotensin-converting enzyme 2
BLI	Biolayer interferometry
Ft	Ferritin
HDL	High-density lipoprotein
HFt	Heavy-chain ferritin
LbpB	Lactoferrin-binding protein B
LDL	Low-density lipoprotein
LDLR	Low-density lipoprotein receptor
Lf	Lactoferrin
LFt	Light-chain ferritin
NP	Nanoparticle
PBM	PCSK9-binding molecule
PCSK9	Proprotein convertase subtilisin/kexin type 9
PFO	Pseudo-first order
PPI	Protein-protein interaction
RBD	Receptor-binding domain
RBM	Receptor-binding motif
SEM	Scanning electron microscopy
TEM	Transmission electron microscopy
Tf	Transferrin
TfR1	Transferrin receptor 1



Article

Lactoferrin Inhibition of the Complex Formation between ACE2 Receptor and SARS CoV-2 Recognition Binding Domain

Roberta Piacentini ^{1,2}, Laura Centi ¹, Mattia Miotto ^{2,3}, Edoardo Milanetti ^{2,3}, Lorenzo Di Rienzo ², Martina Pitea ^{2,4}, Paolo Piazza ⁵, Giancarlo Ruocco ^{2,3}, Alberto Boffi ¹ and Giacomo Parisi ^{2,*}

- ¹ Department of Biochemistry, Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Italy; roberta.piacentini@uniroma1.it (R.P.); centi.laura95@gmail.com (L.C.); alberto.boffi@uniroma1.it (A.B.)
- ² Center of Life Nano and Neuro Science, Institute of Italian Technology, Viale Regina Elena 291, 00181 Rome, Italy; miottomattia1@gmail.com (M.M.); edoardo.milanetti@uniroma1.it (E.M.); lorenzo.dirienzo@iit.it (L.D.R.); martina.pitea@iit.it (M.P.); giancarlo.ruocco@iit.it (G.R.)
- ³ Department of Physics, Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Italy
- ⁴ D-Tails s.r.l., Via di Torre Rossa 66, 00165 Rome, Italy
- ⁵ EDIF Instruments s.r.l., Via Ardeatina 132, 00147 Rome, Italy; paolopiazza@edifinstruments.com
- * Correspondence: giacomo.parisii@iit.it

Abstract: The present investigation focuses on the analysis of the interactions among human lactoferrin (LF), SARS-CoV-2 receptor-binding domain (RBD) and human angiotensin-converting enzyme 2 (ACE2) receptor in order to assess possible mutual interactions that could provide a molecular basis of the reported preventative effect of lactoferrin against CoV-2 infection. In particular, kinetic and thermodynamic parameters for the pairwise interactions among the three proteins were measured via two independent techniques, biolayer interferometry and latex nanoparticle-enhanced turbidimetry. The results obtained clearly indicate that LF is able to bind the ACE2 receptor ectodomain with significantly high affinity, whereas no binding to the RBD was observed up to the maximum “physiological” lactoferrin concentration range. Lactoferrin, above 1 μM concentration, thus appears to directly interfere with RBD-ACE2 binding, bringing about a measurable, up to 300-fold increase of the K_D value relative to RBD-ACE2 complex formation.

Keywords: SARS-CoV-2 receptor-binding domain (RBD); angiotensin-converting enzyme 2 (ACE2); lactoferrin; biolayer interferometry; nanoparticle enhanced turbidimetry; kinetic analysis



Citation: Piacentini, R.; Centi, L.; Miotto, M.; Milanetti, E.; Di Rienzo, L.; Pitea, M.; Piazza, P.; Ruocco, G.; Boffi, A.; Parisi, G. Lactoferrin Inhibition of the Complex Formation between ACE2 Receptor and SARS CoV-2 Recognition Binding Domain. *Int. J. Mol. Sci.* **2022**, *23*, 5436. <https://doi.org/10.3390/ijms23105436>

Academic Editor: Samuel De Visser

Received: 20 April 2022
Accepted: 11 May 2022
Published: 13 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Lactoferrin (LF) is an iron-binding glycoprotein composed of approximately 700 amino acids (molecular weight of about 80 kD, depending on glycosylation extent) folded into two globular lobes connected by an α -helix. LF can bind two ferric ions with a high-affinity metal binding site in each lobe; furthermore, it can also bind Cu^{+2} , Zn^{+2} and Mn^{+2} ions. All mammals can produce LF; secreted from cells of the epithelial mucosa within most exocrine fluids, including tears, saliva, nasal and bronchial secretions and, above all, milk, where it is the major iron-binding protein. Lactoferrin has been shown to play a key role in many biological functions related to innate immunity and more in general in the defense against pathogens [1–4]. In recent years, many experimental observations highlighted novel details within the reported biological activities, with particular focus on the broad antimicrobial action against bacteria, viruses and fungi. Thus, LF biological activities have been attributed only in part to the iron-sequestering activity [5], as they also involve receptor binding, signaling or even protein folding [6,7]. LF has been reported to interact directly with a variety of virus capsid proteins [8], thus suggesting an impairment of viral entry into target cells by blocking the recognition of host cell molecules that the virus itself uses as receptors or co-receptors. Direct binding to viral particles has been described for herpesvirus [9], polio- and rotaviruses [10] and, possibly, human immunodeficiency virus [11] according

to mechanisms reviewed by Berlutti et al. [12]. So far, several lines of evidence indicated that a protective role of LF might be also operating towards SARS-CoV-2 coronavirus infection [13–18]. In this case, the key cellular receptors of SARS-CoV-2, similarly to those of SARS-CoV and MERS-CoV, have been identified with certainty and have been considered as the main conduit for virus entry within the lower respiratory system [19]. In vivo experiments confirmed that SARS-CoV-2 virus cell entry is mediated by high-affinity interactions between the receptor-binding domain (RBD) of the virus spike glycoprotein and the human host angiotensin-converting enzyme 2 (ACE2) receptor [20]. The spike protein is a 1273-amino acid single-pass transmembrane protein with a short C-terminal tail on the interior of the virus, a transmembrane helix, and a large N-terminal ectodomain exposed on the virus external surface. The spike glycoprotein forms homotrimers in which the three subunits interact through their ectodomains. Each subunit contains two topologically and functionally distinct regions named S1 and S2. The S1 regions at the N-terminal end form the portion of the protein furthest from the viral surface within the assembled trimer, whereas the S2 regions form a flexible “stalk” harboring interfaces that hold the trimer in place. High-resolution cryo-EM studies have unveiled the structural determinants that govern the interaction of the spike protein and/or its RBD domain with ACE2, pointing out that the entry of the virus inside the host cells requires the occurrence of a sequence of molecular interactions involving sialic acid and/or heparan sulphate (HF) residues that mediate the attachment of the virion to the cell surface and ultimately lead to spike protein binding to ACE2 receptor and initiate the internalization process [21–24]. In this framework, the possible molecular mechanisms behind the suggested antiviral action of LF was investigated in a computational study [25,26] based on the search for a possible direct interaction between LF and spike protein or between LF and ACE2 receptor, that could inhibit the RBD/ACE2 complex formation. The results of the computational analysis indicated the presence of possible binding regions having a meaningful shape complementarity for both LF and spike protein and LF and ACE2 receptor. On this basis, the present investigation was dedicated to the direct experimental measurement of the computationally predicted interactions in solution on purified proteins. To this end, the purified ACE2 receptor protein interaction with the RBD domain and with lactoferrin were measured by means of biolayer interferometry (BLI) and latex nanoparticle-enhanced turbidimetry. Turbidimetric assays with latex nanoparticles are widely applied for the detection of biological analytes. The most common applications involve the study of immunoreactivity [27,28] or of the colloidal properties of polymers [29]. In general, biolayer interferometry has been applied in a wide variety of research and development environments for measuring kinetics parameters and for the quantitation of antibodies and proteins, whereas latex nanoparticle-enhanced turbidimetry that exploits molecules adsorption properties to a surface is largely unexplored in non-clinical applications. Nonetheless, it was demonstrated to be a very useful complement to interferometry in order to quantify the interaction between proteins with sufficient reciprocal affinity. Overall, the data indicate that the inhibitory power of lactoferrin on the formation of the ACE2/RBD complex is manifested in its binding to the ACE2 receptor rather than to the RBD domain of the spike protein.

2. Results

The ability of LF to bind to specific sites on either the human receptor ACE2 and/or COVID-19 spike's RBD region was investigated by means of interferometric (BLI) and latex nanoparticle-enhanced turbidimetry measurements. The two independent methodologies were used to assess the affinities and rates of LF binding to ACE2 human receptor and to COVID-19 RBD (see Supplementary Materials section for further explanations of the techniques used).

2.1. Binding of Lactoferrin to ACE2 and RBD

The molecular interaction of LF with ACE2 and RBD and related affinities were investigated independently, as depicted in Figure 1. Panels (a) and (b) report the time-

courses of the interaction between LF and ACE2, whereas panels (c) and (d) refer to the interaction between LF and RBD.

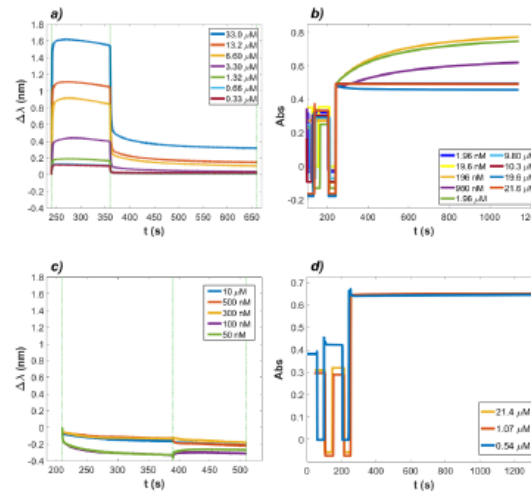


Figure 1. Time–courses of the binding between LF and ACE2 and LF and RBD. (a) BLI signals of LF as the analyte protein at different concentrations with ACE2 loaded on the biosensor (ProA). The vertical dashed lines indicate the duration of the binding step (120 s) and of the dissociation step (300 s). The value of K_D obtained from the BLItz software was $K_D = 27.64 \mu\text{M}$. (b) Turbidimetry assay performed with LF as the analyte protein at variable concentrations in solution with latex nanospheres coated with ACE2. From the data analysis performed using Equations (S3)–(S5), the value of K_D obtained was $46.12 \mu\text{M}$. (c) BLI signals of LF as the analyte protein at different concentrations with RBD loaded on the biosensor (HIS2). The binding step was 180 s, the dissociation step was 120 s. (d) Turbidimetry assay performed with LF as the analyte protein at variable concentrations in solution with latex nanospheres coated with RBD. No association of LF to RBD was recorded in both cases.

For the BLI assay (Figure 1a), ACE2 was immobilized on ProA biosensors, and LF was added in 120 s association steps at decreasing concentrations ranging from $33 \mu\text{M}$ to $0.33 \mu\text{M}$, obtained as consecutive dilutions in 1X kinetic buffer. The dissociation rates were measured in a 300 s time range. The time-courses relative to the turbidimetric assay are shown in Figure 1b. Nanospheres with a suitable diameter of 103 nm were coated with ACE2 as described in Materials and Methods. LF was added to the cuvette at concentrations ranging from $21.6 \mu\text{M}$ to 1.96 nM , and the increase in light absorbance at 340 nm was measured. The same experimental set up was applied to study the binding between LF and RBD, where LF concentration ranged from $10 \mu\text{M}$ to 50 nM for BLI and from $21.6 \mu\text{M}$ to 540 nM for turbidimetry. Figure 1c,d show the BLI and turbidimetric analysis for the association of LF to RBD. A clear signal, corresponding to binding or aggregation, was observed between LF and ACE2 ectodomain using both techniques, whereas no signal could be detected, in the same concentration range, between LF and RBD. Table 1 displays the kinetic parameters of LF–ACE2 interaction. The corresponding K_D values were $27.64 \mu\text{M}$ for the interferometric experiment and $46.12 \mu\text{M}$ in the turbidimetric measurement. These results clearly indicated that LF showed a quantifiable interaction with the ACE2 protein but no binding at all with the RBD protein within the observed concentration range.

Table 1. Kinetic parameters relative to LF and ACE2 interaction. The k_{on} and k_{off} parameters were estimated by single exponential fit of the curves (see Equations (S3)–(S5) in Supplementary material), and the corresponding K_D values were calculated according to Equation (S2).

	k_{on} ($M^{-1} s^{-1}$)	k_{off} (s^{-1})	K_D (μM)
INTERFEROMETRY	$(166.90 \pm 4.79) 10^2$	0.461 ± 0.007	27.64 ± 0.91
TURBIDIMETRY	33.41 ± 5.41	$(1.54 \pm 0.39) 10^{-3}$	46.12 ± 12.12

2.2. Interaction between RBD and ACE2 in the Presence of Lactoferrin

Since LF was effectively demonstrated to bind the ACE2 ectodomain within a μM concentration range whereas no binding to RBD was observed in the same concentration range, the ability of LF to effectively inhibit the interaction between ACE2 and RBD was further investigated by means of interferometric and turbidimetric measurements. The measured affinity between RBD and ACE2 was verified to fall in the nanomolar range, in agreement with reports by Saponaro et al. [30]. In Figure 2a,b, the signals acquired with both BLI and turbidimetric techniques are shown. BLI analysis is presented in Figure 2a and was performed by loading ACE2 on ProA biosensors and allowing the association/dissociation of RBD at concentrations ranging from 1.67 μM to 10 nM. From data fitting analysis performed by BLItz software, the resulting affinity constant was 27.06 nM, in good agreement with previously observed values. The concentrations considered for the analyte protein in the assay were such that one of the proteins was in excess with respect to the other, so the pseudo-first order (PFO) conditions were satisfied. In Figure 2b, aggregation time-courses from turbidimetric assays are shown. The graphs refer to acquisitions of latex nanospheres coated with ACE2 protein and mixed in solution with RBD at decreasing concentrations. The resulting affinity constant was 18.15 nM. Finally, LF was added to a solution with RBD, and the interaction with ACE2 was investigated. In Figure 2c,d, it is possible to observe data acquisition via BLI and turbidimetry of the protein system, where ACE2 was the receptor (on the sensor tip or on the nanoparticles' surface, respectively), RBD was maintained constant, and LF was present at variable concentrations. The LF effect on the ACE2–RBD complex formation was first investigated by BLI analysis, as shown in Figure 2c. ACE2 was again loaded on ProA biosensors, while RBD was presented to the biosensors in the association step at a fixed concentration of 452.5 nM and then mixed with decreasing concentrations of LF. As a result, both LF and RBD bound to ACE2 fixed to the biosensors, resulting in an apparent K_{obs} in which the binding affinities of both proteins to the receptor were merged. Interestingly, when LF was present in solution at a concentration as low as 45.2 nM, the association signal recorded for the mix was significantly lower than that obtained in the presence of RBD alone, thus suggesting a strong inhibiting effect of LF. The value of K_D obtained from the BLItz software was $K_D = 101.30$ nM. The same mixing approach was then applied to the turbidimetric methodology, by mixing in the cuvette nanospheres coated with ACE2 as receptor and a solution at a constant concentration of RBD (19.6 nM) in the presence of decreasing concentrations of LF. An inhibitory effect on the aggregation of ACE2 with RBD was observed, as shown in Figure 2d.

To further analyze the data shown in Figure 2d, curve fitting of the saturation curve (Figure 3a) and of the observed rates (Figure 3b) was executed. According to Equation (S7), the saturation of binding sites expressed in terms of ΔAbs exhibited a hyperbolic dependence on the concentration of the ligand protein, i.e., LF. When expressing protein concentration in logarithmic form, the curve appeared sigmoidal, and the flex point corresponded to an "apparent" dissociation constant K_{obs} of the reaction. In Figure 3a, the values of ΔAbs were plotted against $\log_{10}(LF)$. The difference in absorbance for each concentration of LF was calculated between the start of step 4 of the turbidimetric assay (see Figure S1b) and 600 s after it. In Figure 3b, the initial rates, measured on the same experimental curves, are reported.

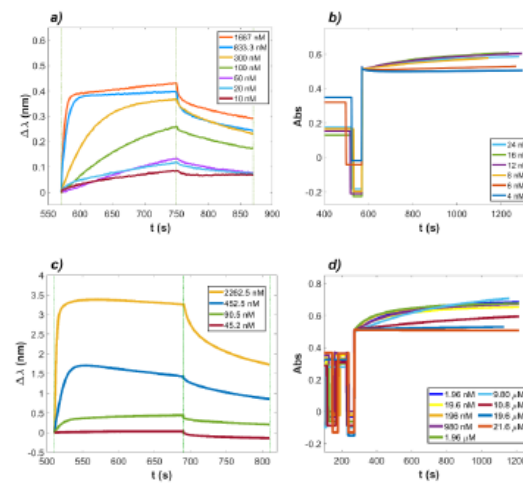


Figure 2. Time-courses of the reaction between RBD in solution with ACE2 in the absence (panels a and b) and in the presence (panels c and d) of human lactoferrin (LF). (a) Signals of the binding and dissociation experiment performed via BLI with the protein system RBD and ACE2. ACE2 was loaded on ProA biosensors, and RBD was in solution at decreasing concentrations. The vertical dashed lines indicate the time interval of the binding step (180 s) and of the dissociation step (120 s). (b) Aggregation signals from turbidimetric assays. Latex nanospheres were coated with the ACE2 protein and mixed in solution with RBD at decreasing concentrations. Panels (c,d): Time-courses of the reaction with RBD and ACE2 in the presence of LF in solution. (c) BLI signals of LF in solution with RBD as the analyte protein at different LF concentrations, with ACE2 loaded on ProA biosensors. The vertical dashed lines indicate the start of the binding step (180 s) and of the dissociation step (120 s). (d) Turbidimetry assay performed with LF at variable concentrations and RBD present at a fixed concentration, in solution with latex nanospheres coated with ACE2.

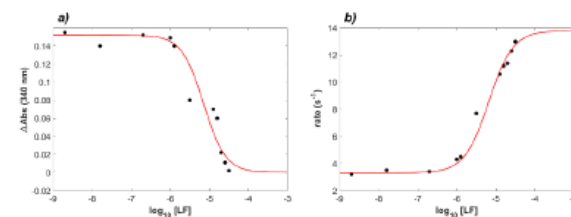


Figure 3. Inhibition of RBD-ACE2 complex formation in the presence of increasing concentrations of lactoferrin. The turbidimetric data of Figure 2d were used to analyze the effect of increasing lactoferrin concentration (in log scale). The absorbance amplitudes of the reactions are shown in panel (a), whereas the initial rates of the same curves are depicted in panel (b). Interpolating red curves represent the best fit to the data obtained by Equation (S8). The data analysis, performed via custom MATLAB program, yielded apparent K_{obs} values of $9.5 \pm 1.5 \mu\text{M}$ (a) and $6.3 \pm 1.2 \mu\text{M}$ (b), respectively, with cooperativity coefficient n of 1.44 ± 0.11 and 1.54 ± 0.08 , respectively.

2.3. Computational Recognition of the Binding Regions of ACE2 to LF

To investigate the possible binding mechanism between ACE2 and LF, the molecular structures of human ACE2 (PDB id: 1R42) and human holo lactoferrin (PDB id: 1LFG) were inspected for portions of the molecular surfaces with high shape complementarity, which can indicate possible binding site candidates (see Methods and [25]). In [26], a computational approach based on the Zernike formalism was applied in order to investigate the possible role of LF in inhibiting SARS-CoV-2 attachment/entrance to the host cells. In particular, the study probed the shape complementarity between LF and ACE2 focusing on the region where the latter binds to SARS-CoV2 spike protein. Here, we extended the analysis looking for possible bindings between LF and ACE2 receptor considering their whole surfaces. In particular, possible binding between LF and ACE2 receptor was studied, and a set of ACE2 receptor regions with high binding propensity were identified, making these possible candidates. Figure 4a shows the residues of ACE2 having binding propensity scores higher than 0.85 with human LF. There are three regions with marked complementarity. Panel b,c and d of Figure 4 show the identified regions on the molecular surface of human ACE2. As widely discussed in [26], the region with the highest probability of interaction is located in the part of ACE2 that in physiological conditions is difficult to access, as it faces the cell membrane (see Figure 4c). However, there are other regions, which are more exposed, characterized by good shape complementarity with other molecular surface patches of LF that may mediate the interaction (see Figure 4b,d).

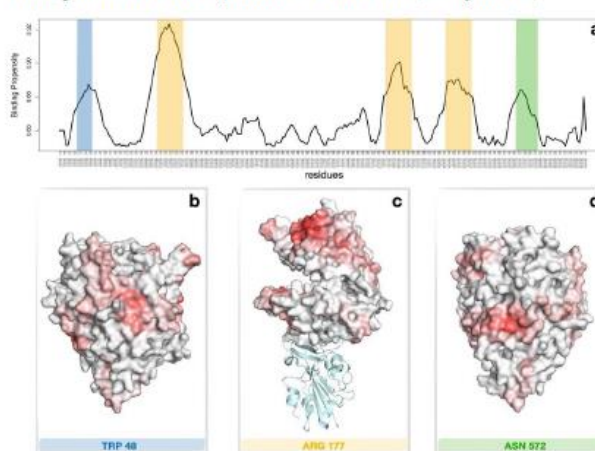


Figure 4. Regions of high binding propensity of ACE2 to LF. (a) Binding propensity of human ACE2 residues for human lactoferrin obtained on the basis of local shape complementarity of the molecular surfaces [7]. Only residues whose binding propensity is higher than 0.85 are reported. Blue, orange, and green bands highlight three different portions of the molecular surfaces characterized by high binding propensity. (b) Molecular surface of the extracellular region of human ACE2 colored according to the Zernike binding propensity score. The color turns from white to red as the local binding propensity increases. The surface was oriented to show the region around residue Tpr48, which is comprised in the blue band in panel (a). (c) Same as in (b) but displaying the region around residue Arg177 and marked with the orange bands; a cartoon representation of the RDB of the spike protein bound to ACE2 is also shown. (d) Same as in (b), but for the green band of the panel (a) marking the region around residue Asn572.

3. Discussion

In the present work, the effect of human lactoferrin (LF) on the interaction between the SARS-CoV-2 RBD domain and the human ACE2 receptor was investigated in solution by means of two independent experimental methodologies, namely, biolayer interferometry and latex nanoparticle-enhanced turbidimetry. The rationale of this experimental set up was based on previous proposals that suggested possible mechanisms of antiviral action of lactoferrin [13,14,31,32] and on computational studies [26,33–35] that revealed significant surface complementarity among lactoferrin and two binding regions on the RBD domain and several possible binding sites to ACE2. Thus far, solution measurements were carried out in order to single out kinetics and thermodynamics parameters for the interaction between lactoferrin and RBD or ACE2, separately. The experimental results in solution pointed out clearly that LF binds to the ACE2 protein ectodomain ($K_D = 27.6 \mu\text{M}$ in BLI experiments, $K_D = 46.1 \mu\text{M}$ in turbidimetric experiments), whereas no thermodynamically significant interaction could be detected for LF binding to RBD.

In a second set of experiments, the inhibitory effect of lactoferrin on the complex formation between RBD and ACE2 was investigated, starting from a reassessment of RBD–ACE2 complex formation by means of interferometry and turbidimetry methods. The reported K_D values for ACE2–RBD interaction, estimated by either BLI or Surface Plasmon Resonance (SPR), ranged between 1.2 nM to 133.3 nM, as reviewed by Saponaro et al. [30]. In the current experimental set up, at 25 °C and in PBS buffer pH 7.4, the calculated K_D values were 27.06 nM (BLI) and 18.15 nM (turbidimetry), in good agreement with reported measurements. Thereafter, the effect of increasing concentrations of LF on the formation of the RBD–ACE2 complex was investigated in experiments in which ACE2 was bound either to the biosensor surface (BLI measurements) or to the latex nanoparticle surface (turbidimetric measurements). Overall, the body of the *in vitro* experimental results thus obtained indicated that LF is indeed capable of inhibiting RBD–ACE2 interaction in solution at concentrations close to the physiological one in human milk (1–10 μM). In particular, as shown in Figures 2 and 3, the apparent dissociation constant (K_{obs}) of the RBD–ACE2 complex increased up to 300-fold as a function of LF concentration in solution. A full thermodynamic model for the complex ternary interactions, however, could not be worked out at the present stage, and the values of K_{obs} rest on a phenomenological basis.

The overall picture emerging from the present investigation is that human lactoferrin can be active in directly impairing spike protein binding to the ACE2 receptor and thus limit viral entry into epithelial cells. On the basis of previous computational studies, a possible structural model for ACE2–LF interaction was proposed, as depicted in Figure 4. The model suggested that the region of binding of LF does not strictly overlap with the well-known ACE2–RBD binding interface. As such, two different inhibition scenarios can be hypothesized: on one side, LF binding on an ACE2 surface region far from the interaction surface with RBD (such as in Figure 4c) can lead to structural rearrangements on the ACE2 ectodomain that could effectively prevent spike attachment to host cells; on the other hand, LF attachment on regions highlighted in Figure 4b,d could hamper directly ACE2–RBD interaction. Altogether, our results vouch for a future deeper investigation on the possible bound conformations and their relative stabilities. In fact, the found binding propensities may help guide docking algorithms and subsequent tailored molecular dynamics simulations [36–38] which could identify the effective binding regions. As an additional remark, BLI and nanoparticle-enhanced turbidimetry proved to be reliable, independent and efficient techniques for the detection and analysis of molecular interactions and binding; hopefully, the present work can be considered for future applications of these techniques.

4. Materials and Methods

4.1. Kinetic Analysis of the LF Inhibition Effect on the ACE2–RBD Complex

Human milk lactoferrin (LF) protein was from Sigma-Aldrich, with molecular weight between 82.4 and 87 kD; the estimated extinction coefficient was $\epsilon_{\text{mM}} = 110.96 \text{ mM}^{-1} \text{ cm}^{-1}$

(280 nm), and the pI was 8.7. The storage buffer was 10 mM phosphate buffered saline at pH 7.4; therefore, all dilutions were executed in the same buffer.

The wild-type receptor-binding domain (RBD) of SARS-CoV-2 spike protein was provided by GenScript. The protein is a C-term HIS-tagged recombinant protein with a predicted molecular weight of 30 kD. The storage buffer was phosphate-buffered saline at pH 7.2, and the calculated pI was 8.91; the stock concentration was 0.89 mg/mL according to the manufacturer. The angiotensin-converting enzyme 2 (ACE2) was also provided by GenScript. The ACE2 ectodomain is a C-term Fc-tagged recombinant protein with predicted molecular weight of 110.5 kD at a concentration of 1.34 mg/mL, in a storage buffer consisting of 20 mM Tris-HCl, 300 mM NaCl, 1 mM ZnCl₂ and 10% glycerol, pH 7.4. The protein pI is 5.49.

Biolayer interferometry is a label-free technology for measuring biomolecular interactions. It is an optical analytical technique that analyzes the interference pattern of white light reflected from two surfaces: a layer of immobilized protein and an internal reference layer. The binding between a ligand immobilized on the biosensor tip surface and an analyte in solution produces an increase in optical thickness at the biosensor tip, which results in a shift in the interference pattern measured in nanometers. This interaction was measured in real time. BLI has been widely used in the study of biomolecular antibody–antigen interactions [39–41], with particular focus, recently, on SARS-CoV-2 antibodies investigation [42].

BLI assays were performed by means of the BLItz system (Sartorius). A description of a BLI assay is presented in Supplementary material (Figure S1a). Protein-A (ProA), anti-His tag (HIS2) and nickel nitriloacetic acid (Ni-NTA) were selected as biosensors and provided by Sartorius. ACE2 Fc-tagged was immobilized on ProA biosensors, with no aspecific binding to all other experimental components, whereas RBD His-tagged was immobilized on either HIS2 or NTA biosensors, the latter being more susceptible to aspecific binding of non-his-tagged proteins. The biosensors were first equilibrated 10 min in 1X kinetic buffer (Sartorius) consisting of PBS with 0.02% Tween20, 0.1% BSA and 0.05% Na₂S₂O₃. Afterwards, depending on the assay, either ACE2 or RBD was loaded on the corresponding biosensor at a concentration of 50 µg/mL (corresponding to C_{ACE2} = 452.5 nM and C_{RBD} = 1.67 µM), as indicated in Sartorius biosensors' datasheets, for an appropriate time interval. To reach the maximum binding capacity in every experiment, the time length of each experimental step was tested, and each experiment was designed accordingly. The concentration range for each associating protein was chosen, when possible, based on the K_D value available from the literature or experimentally determined for every different scenario where K_D values were unknown. The data recorded were analyzed by means of BLItz software and MATLAB to extrapolate the kinetic parameters. All association and dissociation curves were fitted by a single exponential function. A description of the instrument and of the analytical model of the system is reported in the Supplementary Material. Pseudo-first order (PFO) conditions are met when the initial concentration of one of the two reagents is in large excess with respect to the other (between 50- and 100-fold) [43–45]. Each acquisition was repeated twice to confirm reproducibility.

The method for the turbidimetric assay consisted in performing a time-course acquisition of light absorption signals with a spectrophotometer (Jasco V-750 UV-Visible/NIR) at fixed wavelength (340 nm) and bandwidth (10 nm) for 1300 s. For the turbidimetric assays, chloromethylated (-CH₂-Cl) polystyrene nanospheres (furnished by Ikerlat) were used, with a diameter of 103 nm according to the manufacturer. The nanosphere stock concentration was 0.1 g/mL (10%). The parking area was 95 Å²/group. Before the assay, the nanospheres were coated with the receptor protein in 10 mM phosphate buffer (PB) at pH 7.4 in the case of ACE2 protein. A total of 2 mL of the solution of ACE2-coated nanospheres was obtained by adding 4 µL of Ikerlat nanospheres to 1.92 mL of PB. Such an amount of nanospheres guarantees 0.02% of solids in the final solution. Then, 26 µL of ACE2 (stock concentration 1.34 mg/mL) was added to the nanospheres solution (final ACE2 concentration 17.42 µg/mL = 157.65 nM), and the solution was kept in gentle agita-

tion on a tilting platform at room temperature for 3 h. Finally, 52 μL of blocking buffer was added. Such buffer consisted of goat serum (Millipore) at a concentration of 20 mL/L and ProClin 300 biocide (Sigma-Aldrich) at 0.3 mL/L in PBS at pH 7.4, to fully block potentially unbound chloromethyl groups. This reaction was incubated again on a tilting surface, gently mixing overnight. In the case of the RBD protein, the coating buffer was 20 mM sodium carbonate at pH 9.2 containing 150 mM NaCl. For the coating step of nanospheres with RBD as the receptor protein, the procedure was the same but with different quantities: 1 μL of Ikerlat nanospheres was added to 0.5 mL of bicarbonate and guaranteed 0.02% of solids in the final solution (the stock concentration was 0.1 g/mL). The concentration of RBD used for the coating procedure was 174.2 $\mu\text{g}/\text{mL} = 5.8 \mu\text{M}$.

For spectra acquisition, polymethyl methacrylate (PMMA) cuvettes with 500 μL capacity and 0.6 cm optical path were used. Figure S1b shows an example of data acquisition in the turbidimetric assays performed. Nanoparticle aggregation results in an increasing signal of absorbance attributed to RBD-ACE2 complex formation and was analyzed in phenomenological terms as a saturation curve (see Supplementary Material for a brief discussion). Thus, a turbidimetric assay was performed with nanospheres coated with fixed concentrations of either ACE2 or RBD, while LF was present at different concentrations. The analysis was carried out in phosphate-buffered saline at pH 7.25 with 0.1% glycine, 0.1% NaN_3 and polyethylene glycol (PEG) 6000 at a concentration of 5%. The analysis of the turbidimetric data was performed by means of a MATLAB custom program.

4.2. Computational Recognition of Binding Regions of ACE2 to LF

The computational analysis was performed starting from the crystallographic structures of human lactoferrin in the holo form (PDB id: 1LFG) and that of human ACE2 (PDB id: 1R42). For both protein structures, we used DMS [46] to compute the solvent-accessible surface, using a density of 5 points per \AA^2 and a water probe radius of 1.4 \AA . The unit normal vector, for each point of the surface, was calculated using the flag $-n$. The DMS software returns a discretized version of the molecular surface, that is represented by a set of points in the three-dimensional space. To find possible interacting regions, we looked for regions of molecular surfaces with high shape complementarity. Operatively, we compared all possible regions of ACE2 with all possible regions of LF. We defined surface regions, i.e., patches, as the group of points that fell within a sphere of radius $R_s = 6 \text{\AA}$, centered on each of the points of the surfaces. Once the patch was selected, we described it on the basis of the Zernike polynomials [26], i.e., we associated to each patch a set of coefficients that represented the patch in the Zernike basis. Once a patch was represented in terms of its Zernike descriptors, the complementarity between that patch and another one could be simply measured as the Euclidean distance between the invariant vectors. We thus associated to each point of the two surfaces the minimum distance value observed—the binding propensity—between the considered point and all points of the other surface. After all surface points were associated with their binding propensity, we performed a smoothing process to highlight signals in specific regions characterized mostly by low-distance values. In this process, each point was associated with the mean value of the points in its neighborhood: the basic idea is that the interacting region should be mostly made up of elements with high complementarity and, therefore, a high average value of binding propensity (see [26] for further details). For both the patch definition and the smoothing process, we adopted a sphere radius of 6 \AA .

5. Conclusions

In the present work, the inhibiting effect of human lactoferrin on the complex formation between the spike protein binding domain RBD and the ACE2 receptor, was demonstrated directly for the first time. In detail, LF showed a remarkable binding propensity towards the ACE2 receptor, whereas no affinity to RBD was established. Moreover, when LF was mixed at a concentration of 1 μM with RBD, a strong decrease in binding affinity between ACE2 and wild type RBD was observed, thus suggesting the ability of

LF to inhibit, in vitro, the formation of such complex. Three distinct binding sites were computationally identified for ACE2-LF interaction, enlightening two different inhibition mechanisms through which lactoferrin could exert its inhibiting properties. The detailed mechanism of LF inhibition of the complex formation remains unclear at the moment. Considering the present lack of a preventative regimen established for COVID-19, the use of LF as a local protection might be increasingly desirable as an adjunct protection against SARS-CoV-2 infection.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23105436/s1>.

Author Contributions: Conceptualization, E.M., P.P., A.B. and G.P.; Data Curation, R.P.; Formal analysis, R.P. and L.D.R.; Funding acquisition, G.R., A.B.; Investigation, R.P., L.C., M.P.; Methodology, M.M., L.D.R., P.P. and M.P.; Project administration, G.R. and A.B.; Software, M.M., E.M. and L.D.R.; Supervision, A.B. and G.P.; Validation, E.M.; Visualization, M.M. and E.M.; Writing—original Draft, R.P.; Writing—review & editing, G.P. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to these results was also supported by a European Research Council Synergy grant ASTRA (n. 855923).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Redwan, E.M.; Uversky, V.N.; El-Fakharany, E.M.; Al-Mehdar, H. Potential lactoferrin activity against pathogenic viruses. *Comptes Rendus Biol.* **2014**, *337*, 581–595. [CrossRef] [PubMed]
- Baker, E.; Baker, H.M. Molecular structure, binding properties and dynamics of Lactoferrin. *Cell. Mol. Life Sci.* **2005**, *62*, 2531–2539. [CrossRef] [PubMed]
- Giansanti, E.; Panella, G.; Leboffe, L.; Antonini, G. Lactoferrin from Milk: Nutraceutical and Pharmacological Properties. *Pharmaceuticals* **2016**, *9*, 61. [CrossRef] [PubMed]
- Rosa, L.; Cutone, A.; Lepanto, M.S.; Paesano, R.; Valenti, P. Lactoferrin: A Natural Glycoprotein Involved in Iron and Inflammatory Homeostasis. *Int. J. Mol. Sci.* **2017**, *18*, 1985. [CrossRef] [PubMed]
- Varki, A. Biological roles of glycans. *Glycobiology* **2017**, *27*, 3–49. [CrossRef] [PubMed]
- Bishop, J.R.; Gagneux, P. Evolution of carbohydrate antigens—microbial forces shaping host glycomes? *Glycobiology* **2007**, *17*, 23R–34R. [CrossRef]
- Miotto, M.; Di Rienzo, L.; Gosti, G.; Milanetti, E.; Ruocco, G. Does blood type affect the COVID-19 infection pattern? *PLoS ONE* **2021**, *16*, e0251535. [CrossRef]
- Yi, M.; Kaneko, S.; Yu, D.Y.; Murakami, S. Hepatitis C virus envelope proteins bind lactoferrin. *J. Virol.* **1997**, *71*, 5997–6002. [CrossRef]
- Harmsen, M.; Swart, P.J.; De Bèthune, M.-P.; Pauwels, R.; De Clercq, E.; The, T.B.; Meijer, D.K.F. Antiviral Effects of Plasma and Milk Proteins: Lactoferrin Shows Potent Activity against Both Human Immunodeficiency Virus and Human Cytomegalovirus Replication In Vitro. *J. Infect. Dis.* **1995**, *172*, 380–388. [CrossRef]
- Superti, E.; Siciliano, R.; Rega, B.; Giansanti, E.; Valenti, P.; Antonini, G. Involvement of bovine lactoferrin metal saturation, sialic acid and protein fragments in the inhibition of rotavirus infection. *Biochim. Biophys. Acta* **2001**, *1528*, 107–115. [CrossRef]
- Puddu, P.; Borghi, P.; Gessani, S.; Valenti, P.; Belardelli, E.; Seganti, L. Antiviral effect of bovine lactoferrin saturated with metal ions on early steps of human immunodeficiency virus type 1 infection. *Int. J. Biochem. Cell Biol.* **1998**, *30*, 1055–1063. [CrossRef]
- Berlutti, E.; Pantanella, E.; Natalizi, T.; Frioni, A.; Paesano, R.; Polimeni, A.; Valenti, P. Antiviral Properties of Lactoferrin—A Natural Immunity Molecule. *Molecules* **2011**, *16*, 6992–7018. [CrossRef] [PubMed]
- Campione, E.; Cosio, T.; Rosa, L.; Lanna, C.; Di Girolamo, S.; Gaziano, R.; Valenti, P.; Bianchi, L. Lactoferrin as Protective Natural Barrier of Respiratory and Intestinal Mucosa against Coronavirus Infection and Inflammation. *Int. J. Mol. Sci.* **2020**, *21*, 4903. [CrossRef] [PubMed]
- Wang, Y.; Wang, P.; Wang, H.; Luo, Y.; Wan, L.; Jiang, M.; Chu, Y. Lactoferrin for the treatment of COVID-19 (Review). *Exp. Ther. Med.* **2020**, *20*, 272. [CrossRef] [PubMed]

15. Rosa, L.; Tripepi, G.; Naldi, E.; Aimati, M.; Santangeli, S.; Venditto, E.; Caldarelli, M.; Valenti, P. Ambulatory COVID-19 Patients Treated with Lactoferrin as a Supplementary Antiviral Agent: A Preliminary Study. *J. Clin. Med.* **2021**, *10*, 4276. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Chang, R.; Ng, T.B.; Sun, W.-Z. Lactoferrin as potential preventative and adjunct treatment for COVID-19. *Int. J. Antimicrob. Agents* **2020**, *56*, 106118. [\[CrossRef\]](#)
17. Nedyalkova, M.; Vasighi, M.; Sappati, S.; Kumar, A.; Madurga, S.; Simeonov, V. Inhibition Ability of Natural Compounds on Receptor-Binding Domain of SARS-CoV2: An In Silico Approach. *Pharmaceuticals* **2021**, *14*, 1328. [\[CrossRef\]](#)
18. Odolczyk, N.; Marzec, E.; Winiewska-Szajewska, M.; Poznański, J.; Zielenkiewicz, P. Native Structure-Based Peptides as Potential Protein-Protein Interaction Inhibitors of SARS-CoV-2 Spike Protein and Human ACE2 Receptor. *Molecules* **2021**, *26*, 2157. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [\[CrossRef\]](#)
20. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [\[CrossRef\]](#)
21. Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367*, 1444–1448. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Lang, J.; Yang, N.; Deng, J.; Liu, K.; Yang, P.; Zhang, G.; Jiang, C. Inhibition of SARS Pseudovirus Cell Entry by Lactoferrin Binding to Heparan Sulfate Proteoglycans. *PLoS ONE* **2011**, *6*, e23710. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Langford-Smith, A.; Day, A.J.; Bishop, P.; Clark, S.J. Complementing the Sugar Code: Role of GAGs and Sialic Acid in Complement Regulation. *Front. Immunol.* **2015**, *6*, 25. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Hulswit, R.J.G.; Lang, Y.; Bakkers, M.J.G.; Li, W.; Li, Z.; Schouten, A.; Ophorst, B.; van Kuppeveld, E.J.M.; Boons, G.-J.; Bosch, B.-J.; et al. Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 2681–2690. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Milanetti, E.; Miotto, M.; Di Rienzo, L.; Monti, M.; Gosti, G.; Ruocco, G. 2D Zernike polynomial expansion: Finding the protein-protein binding regions. *Comput. Struct. Biotechnol. J.* **2020**, *19*, 29–36. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Miotto, M.; Di Rienzo, L.; Bo, L.; Boffi, A.; Ruocco, G.; Milanetti, E. Molecular Mechanisms Behind Anti SARS-CoV-2 Action of Lactoferrin. *Front. Mol. Biosci.* **2021**, *8*, 607443. [\[CrossRef\]](#)
27. Cölfen, H.; Völkel, A.; Eda, S.; Kobold, U.; Kaufmann, J.; Puhlmann, A.; Göltner, A.C.; Wachernig, H. Mechanism of Nanoparticle-Enhanced Turbidimetric Assays Applying Nanoparticles of Different Size and Immunoreactivity. *Langmuir* **2002**, *18*, 7623–7628. [\[CrossRef\]](#)
28. Coletta, G.; Amendola, V. Numerical Modelling of the Optical Properties of Plasmonic and Latex Nanoparticles to Improve the Detection Limit of Immuno-Turbidimetric Assays. *Nanomaterials* **2021**, *11*, 1147. [\[CrossRef\]](#)
29. Gurmani, P.; Lunn, A.M.; Perrier, S. Synthesis of mannosylated and PEGylated nanoparticles via RAFT emulsion polymerisation, and investigation of particle-lectin aggregation using turbidimetric and DLS techniques. *Polymer* **2016**, *106*, 229–237. [\[CrossRef\]](#)
30. Saponaro, E.; Rutigliano, G.; Sestito, S.; Bandini, L.; Storti, B.; Bizzarri, R.; Zucchi, R. ACE2 in the Era of SARS-CoV-2: Controversies and Novel Perspectives. *Front. Mol. Biosci.* **2020**, *7*, 588618. [\[CrossRef\]](#)
31. Rascón-Cruz, Q.; Espinoza-Sánchez, E.A.; Siqueiros-Cendón, T.S.; Nakamura-Bencomo, S.I.; Arévalo-Gallegos, S.; Iglesias-Figueroa, B.F. Lactoferrin: A Glycoprotein Involved in Immunomodulation, Anticancer, and Antimicrobial Processes. *Molecules* **2021**, *26*, 205. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Kell, D.B.; Heyden, E.L.; Pretorius, E. The Biology of Lactoferrin, an Iron-Binding Protein That Can Help Defend Against Viruses and Bacteria. *Front. Immunol.* **2020**, *11*, 1221. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Tortorici, M.A.; Walls, A.C.; Lang, Y.; Wang, C.; Li, Z.; Koerhuis, D.; Boons, G.-J.; Bosch, B.-J.; Rey, F.A.; de Groot, R.J.; et al. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* **2019**, *26*, 481–489. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Milanetti, E.; Miotto, M.; Di Rienzo, L.; Nagaraj, M.; Monti, M.; Golbek, T.W.; Gosti, G.; Roeters, S.J.; Weidner, T.; Otzen, D.E.; et al. In-Silico Evidence for a Two Receptor Based Strategy of SARS-CoV-2. *Front. Mol. Biosci.* **2021**, *8*, 690655. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Campione, E.; Lanna, C.; Cosio, T.; Rosa, L.; Conte, M.P.; Iacovelli, E.; Romeo, A.; Falconi, M.; Del Vecchio, C.; Franchin, E.; et al. Lactoferrin Against SARS-CoV-2: In Vitro and In Silico Evidences. *Front. Pharmacol.* **2021**, *12*, 666600. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Martínez, M.; Cooper, C.D.; Poma, A.B.; Guzman, H.V. Free Energies of the Disassembly of Viral Capsids from a Multiscale Molecular Simulation Approach. *J. Chem. Inf. Model.* **2020**, *60*, 974–981. [\[CrossRef\]](#)
37. Mahmood, I.; Poma, A.B.; Okazaki, K.-I. Optimizing Go-MARTINI Coarse-Grained Model for F-BAR Protein on Lipid Membrane. *Front. Mol. Biosci.* **2021**, *8*, 619381. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Bo, L.; Miotto, M.; Di Rienzo, L.; Milanetti, E.; Ruocco, G. Exploring the Association Between Sialic Acid and SARS-CoV-2 Spike Protein Through a Molecular Dynamics-Based Approach. *Front. Med Technol.* **2021**, *2*, 614652. [\[CrossRef\]](#)
39. Concepcion, J.; Witte, K.; Wartchow, C.; Choo, S.; Yao, D.; Persson, H.; Wei, J.; Li, P.; Heidecker, B.; Ma, W.; et al. Label-Free Detection of Biomolecular Interactions Using BioLayer Interferometry for Kinetic Characterization. *Comb. Chem. High Throughput Screen.* **2009**, *12*, 791–800. [\[CrossRef\]](#)
40. Kamat, V.; Rafique, A. Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. *Anal. Biochem.* **2017**, *536*, 16–31. [\[CrossRef\]](#)

41. Petersen, R.I. Strategies Using Bio-Layer Interferometry Biosensor Technology for Vaccine Research and Development. *Biosensors* **2017**, *7*, 49. [[CrossRef](#)] [[PubMed](#)]
42. Dzimianski, J.V.; Lorig-Roach, N.; O'Rourke, S.M.; Alexander, D.L.; Kimmey, J.M.; DuBois, R.M. Rapid and sensitive detection of SARS-CoV-2 antibodies by biolayer interferometry. *Sci. Rep.* **2020**, *10*, 21738. [[CrossRef](#)] [[PubMed](#)]
43. Malatesta, F. The study of bimolecular reactions under non-pseudo-first order conditions. *Biophys. Chem.* **2005**, *116*, 251–256. [[CrossRef](#)] [[PubMed](#)]
44. Kumaraswamy, S.; Tobias, R. Label-Free Kinetic Analysis of an Antibody–Antigen Interaction Using Biolayer Interferometry. In *Protein-Protein Interactions. Methods in Molecular Biology*; Meyerkord, C., Fu, H., Eds.; Humana Press: New York, NY, USA, 2015; Volume 1278, pp. 165–182. [[CrossRef](#)]
45. Müller-Esparza, H.; Osorio-Valeriano, M.; Steube, N.; Thanbichler, M.; Randau, L. Bio-Layer Interferometry Analysis of the Target Binding Activity of CRISPR-Cas Effector Complexes. *Front. Mol. Biosci.* **2020**, *7*, 98. [[CrossRef](#)] [[PubMed](#)]
46. Richards, F.M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176. [[CrossRef](#)]

Appendix II

communications biology

ARTICLE



<https://doi.org/10.1038/s42003-021-02946-w> OPEN

Inferring the stabilization effects of SARS-CoV-2 variants on the binding with ACE2 receptor

Mattia Miotto¹, Lorenzo Di Rienzo¹, Giorgio Gosti¹, Leonardo Bo¹, Giacomo Parisi¹, Roberta Piacentini^{1,2}, Alberto Boffi^{1,2}, Giancarlo Ruocco^{1,3} & Edoardo Milanetti^{1,3}✉

As the SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) pandemic continues to spread, several variants of the virus, with mutations distributed all over the viral genome, are emerging. While most of the variants present mutations having little to no effects at the phenotypic level, some of these variants are spreading at a rate that suggests they may present a selective advantage. In particular, these rapidly spreading variants present specific mutations on the spike protein. These observations call for an urgent need to characterize the effects of these variants' mutations on phenotype features like contagiousness and antigenicity. With this aim, we performed molecular dynamics simulations on a selected set of possible spike variants in order to assess the stabilizing effect of particular amino acid substitutions on the molecular complex. We specifically focused on the mutations that are both characteristic of the top three most worrying variants at the moment, i.e. the English, South African, and Amazonian ones, and that occur at the molecular interface between SARS-CoV-2 spike protein and its human ACE2 receptor. We characterize these variants' effect in terms of (i) residue mobility, (ii) compactness, studying the network of interactions at the interface, and (iii) variation of shape complementarity via expanding the molecular surfaces in the Zernike basis. Overall, our analyses highlighted greater stability of the three variant complexes with respect to both the wild type and two negative control systems, especially for the English and Amazonian variants. In addition, in the three variants, we investigate the effects a not-yet observed mutation in position 501 could provoke on complex stability. We found that a phenylalanine mutation behaves similarly to the English variant and may cooperate in further increasing the stability of the South African one, hinting at the need for careful surveillance for the emergence of these mutations in the population. Ultimately, we show that the proposed observables describe key features for the stability of the ACE2-spike complex and can help to monitor further possible spike variants.

¹Center for Life Nano & Neuroscience, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161 Rome, Italy. ²Department of Biochemical Sciences "Alessandro Rossi Fanelli", Sapienza University of Rome, P.Le A. Moro 5, 00185 Rome, Italy. ³Department of Physics, Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Italy. ✉email: edoardo.milanetti@uniroma1.it

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection was firstly observed in late 2019^{1,2}. In the subsequent months of epidemic spreading, many SARS-CoV-2 variants, viral sequences characterized by at least one mutation with respect to the original one, have been detected worldwide³.

In coronaviruses, mutations naturally occur during viral replication, and despite the fact that coronaviruses encode for an enzyme that corrects the errors, some of these mutations are preserved, originating new variants. As for all biological systems, the action of natural selection eventually tends to fix in the genome the mutations characteristic of those variants that present an increase of the fitness, and it has been registered that in these months the rate of emergence of new SARS-CoV-2 variants is about two variants per month^{4,5}.

This rapid proliferation of variants poses a further threat for the community as the virus can acquire different phenotypes. For instance, the main mutation of the B line, involving the amino acid substitution D614G in the spike protein, is established since March 2020 and it is now largely dominant in patients⁶⁻⁸. This mutation would allow the receptor-binding domain (RBD) of the spike to assume a conformation more suitable to bind Angiotensin-converting enzyme 2 (ACE2), and it could be responsible for the increased viral action^{9,9}.

More recently, on December 14, 2020, authorities in the United Kingdom of Great Britain and Northern Ireland reported to World Health Organization that a new SARS-CoV-2 variant, named B.1.1.7, and commonly known as the English variant, has been identified via viral genomic sequencing^{10,11}. The variant is defined by the presence of a range of 14 mutations involving amino acid modifications and three deletions, including the spike D614G mutation. Even if investigations are underway to determine whether this variant is associated with any changes in antibody response or vaccine efficacy, it seems to be characterized by an increased transmissibility¹² and lethality¹³, also because it is spreading with very high speed all over the world.

Among the mutations affecting the spike protein, like 69-70del or P681H, mutation N501Y is located in the region that directly contacts the ACE2 receptor. Therefore, it is possible that this mutation could have a direct effect on the binding affinity between the two proteins¹⁴.

Furthermore, the importance of this amino acid substitution is confirmed by its presence in two other rapidly spreading variants, B.1.351¹⁵ and P.1¹⁶ (commonly referred to as South African and Amazonian variants, respectively).

These two variants both include other amino acid substitutions in the spike binding site, making clear that this region is under severe evolutionary pressure. Indeed, selectivity and affinity of the spike toward its main receptor, ACE2, remain the crucial factors determining SARS-CoV-2 contagiousness and virulence.

Even if, it has been demonstrated that SARS-CoV-2, similarly to other coronaviruses^{17,18}, is able to bind to sialoglycan-based receptors in the spike's N-terminal domain¹⁹⁻²¹.

From a molecular point of view, the spike protein of CoVs, protruding from the viral membrane, not only plays a crucial role as a fundamental structural protein, but it also is essential for the interaction between CoV systems and host cells²². Structurally, the spike protein is found in the trimeric complex, each chain composed of two sub-units: S1 and S2. The Receptor Binding Domain (RBD), located in the S1 domain, is responsible for viruses' interaction with receptors on the host cell surface²³. On the other hand, the S2 subunit is responsible for the fusion between the virus and host membrane, causing the viral genome to penetrate the host cell's cytoplasm²⁴.

Interestingly, the interaction with ACE2 involves the C-terminal domain of SARS-CoV-2 spike protein, whose amino

acid sequence is well conserved with respect to SARS-CoV homologous one²⁵. Conversely, the N-terminal domain presents some insertions, and these additional surface regions could be used by the virus to bind other cell receptors, so constituting an additional cell entry mechanism²⁶.

Here, we perform a set of molecular dynamics simulations on different spike-receptor complexes, involving mutations at the spike protein interface. In particular, we both consider single mutations belonging to the binding site of the spike protein for which the binding affinity experimental data is known²⁷, and some of the variants of SARS-CoV-2 are currently most widespread in the world. We analyze the mutations for which binding affinity is known in order to detect the dynamic-structural properties of the ACE2-spike complex that give the spike protein a greater propensity to interact with the molecular partner. Only by considering the RBD of the spike protein and the extra-membrane domain of the ACE2 receptor, we analyze the dynamic properties of the residues belonging to the two interfaces. Furthermore, an analysis based on graph theory was performed in order to investigate the stability of the contacts during molecular dynamics simulations. Finally, we analyze the shape complementarity of the two interfaces over time, using Zernike polynomials to characterize the shape of each portion of the molecular surface, establishing a complementarity value between each pair of surface²⁸⁻³¹. Also, in this case, we investigate the stability of the geometrical matching between the two interfaces during the molecular dynamics simulation because stable binding to the host receptor plays a crucial role for virus entry mechanisms³².

Overall, our study shows that the English variant and the South African variant have structural and dynamic properties similar to the N501F single-mutation variant, which is experimentally known to be the one with the highest binding affinity with ACE2²⁷. This finding invites us to explore the possible cooperative effect of mutations in the binding region. Therefore, we substituted the F amino acid (Phenylalanine) at position 501 of the spike protein for both the Amazonian and South African variants. The results obtained from our fully computational approach show that the presence of the F amino acid in the Amazonian variant would worsen the binding affinity. On the other hand, the same mutation carried out in addition to the mutations of the South African variant would increase the stability of the spike-ACE2 complex, suggesting these mutations as a possible worrying variant.

Results and discussion

The role of spike mutations in the emerging variants. Although mutations are showing up all over the SARS-CoV-2 viral genome, those taking place on the spike protein are under intense scrutiny as they are expected to directly impact viral entrance in the host cells, thus on transmissibility and infectivity. The first notable mutation that has been observed involved residue 614 which changed from D to G⁶. This mutation even if localized in a region distant from the binding site, is rapidly fixed in the population suggesting an indirect effect on the phenotype. Indeed, a recent computation study highlighted a conformational change driven by such mutation that favors ACE2 receptor binding, thus explaining the phenotypic advantage⁸.

With the huge spread of the epidemics, other mutations accumulated over the SARS-CoV-2 genome. In particular, a variant of the virus with six mutated amino acids and two deletions in the spike protein emerged in England in late 2020. The English variant has the D614G mutation together with mutations/deletions belonging to spike protein: 69-70 HV deletion, 144Y deletion, N501Y, A570D, P681H, T716I, S982A,

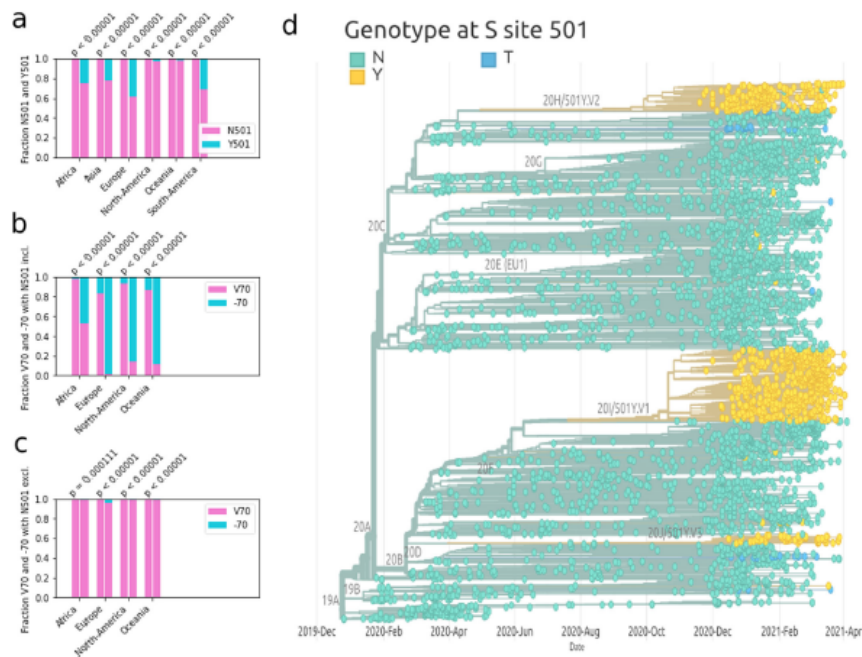


Fig. 1 Sequence analysis of the English variant mutations. **a** Fractions of sequenced viral samples having amino acid N in position 501 of the spike protein (orange) vs those having Y (blue) for two different time intervals and in different countries. **b** Same as panel (a), but comparing samples with residue V70 (orange) and samples for which deletion of residue 70 took place (blue). **c** Same as in panel (b) but considering only sequences where residue 501 is an N. **d** Phylogenetic tree of the SARS-CoV-2 variants. Leaves are colored in light green if the corresponding spike sequence has an N in position 501, in dark blue in case of a T, and in yellow for a Y.

and D1118H. Specifically, mutation N501Y and deletions 69–70 attracted much attention since the mutation falls in the receptor-binding domain (RBD) while the double deletion interests the N-terminal domain, which has been demonstrated to bind sialic acid-rich attachment factors^{19,21}.

To begin with, we looked at genomic data taken from infected individuals divided by geographical location—which is provided by the regularly updated public website (<https://cov.lanl.gov>)—to assess the variance invasion potential. In particular, analyses are based on 823,121 spike alignment sequences taken at different times and different geographical areas using the consistency and significance test presented by Korber et al.³³ First, we compared sequences carrying the “wild type” residue in position 501 (i.e., N501) with the ones carrying the 501Y mutation. Note that, in this work, we considered the spike protein first detected in Wuhan at the beginning of the epidemics as the “wild type” (WT).

Figure 1a displays the changes of relative frequencies for N501 and Y501 for two-time points separated by 2 weeks. The first time point represents all sequences up to the onset day, and the second time point includes all the sequences acquired at least 2 weeks after the onset date (at least 15 sequences are required for significance). One can see that in each country, the variant 501Y is rapidly spreading.

The significance of the signal is obtained by testing the null hypothesis that the mutation does not affect the variant fitness and that consequently, the relative frequency shift must randomly go with equal probability in each direction. The *p* values confirm that in all continents we observe a systematic increase of the relative frequency of Y501. Next, we moved to consider sequences carrying the deletion V70 (taking place on the glycan-binding domain). Figure 1b, c show respectively the change in the relative frequency of the deletion at position 70 constraining respectively for the N501 and Y501 variant; this shows that the deletion of the 70th position increases in relative frequencies at a higher rate in the Y501 variant. This suggests that deletion 70 is driven by the rapid spread of the English variant, and/or that the Y501 variant gives a stronger relative change in fitness with respect to the deletion.

Finally, looking at all infection data, we compare the invasive potential of the different 501 mutations. The phylogenetic tree in Fig. 1d was obtained using the publicly available interactive visualization platform Nextstrain³⁴ and it is based on the contribution of 4025 genomes sampled between December 2019 and April 2021, stored and elaborated in the Nextstrain database and bioinformatics pipeline for phylodynamics analysis³⁴. A full list of the sequence authors is available at <https://nextstrain.org/sars-cov-2>. Light green dots represent sampled genomes with

Table 1 Variants considered in the present work.

Wild type (WT)	N501N
English (UK)	N501Y
SARS-03 like	N501T
Negative control 1	N501K
Negative control 2	N501D
Positive control 1	N501F
South African	K417N-E484K-N501Y
Amazonian	K417T-E484K-N501Y

N501 sequences, yellow dots Y501 and blue T501. One can see that the N501Y mutation independently emerged in different branches of the phylogenetic tree and that in each case it rapidly spreads. Furthermore, up to now the N501T mutation emerged and spread only from a single branch.

Interestingly, experimental measurements of binding affinity upon single mutations found that changing residue 501 from N to both T and Y results in an increase of the affinity, while most of the other possible mutations lead to its decrease²⁷.

In the following, we propose to investigate in greater detail what are the molecular features responsible for the increase/decrease of complex stability upon mutations.

Mutational protocol. In order to analyze the role of amino acid substitutions on the binding between the RBD of the SARS-CoV-2 spike protein and the ACE2 receptor, we performed molecular dynamics simulations on a set of spike variants in complex with human ACE2 and characterized the effects of the different mutations on the stability of the protein complex. Starting from the structural complex comprising the RBD of the WT spike and the human ACE2 receptor (PDB id: 6M0J), we obtained the RBD of variant B.1.1.7 (the English variant) mutating residue 501 from amino acid N to Y.

In addition, we also considered four other single-mutated complexes: N501T and N501F, which are expected to display higher affinity; N501D and N501K, which should exhibit lower stability with respect to the wild type one²⁷.

Then we move to consider cases in which three mutations are present in the RBD, i.e. we analyze the South African and Amazonian variants carrying the mutations K417N-E484K-N501Y and K417T-E484K-N501Y, respectively. All considered systems are reported in Table 1. As experimentally-resolved structures are not available for all the considered systems, mutations were obtained using the PyMol software⁴⁵ (see Methods) and a 500 ns-long molecular dynamics simulation was conducted to equilibrate each system. To validate the mutational procedure, we used available experimental structures, performing two different analyses. First, we retrieved the experimental complexes of the SARS-CoV-2's RBD bound to human ACE2 for the English (PDB id: 7MJN) and Amazonian (PDB id: 7NXC) variants. Then, we carried two additional 500 ns-long molecular dynamics simulations and verify that the configurations explored by the two simulations overlap with those sampled during the MD of the computationally mutated ones. In practice, we compared the structures obtained for both simulations through a principal component analysis (PCA). Indeed, a set of configurations sampled from the simulation of the experimental structure was projected into the essential space, defined by the two principal components of the covariance matrix obtained from the trajectory of the computationally mutated complexes (see Supplementary Figs. 4, 5). Figure 2a, b shows an overlap between the two sets of structures, which indeed have a high degree of similarity in terms of the backbone conformation.

Next, we considered the South African variant's complex, whose experimental structure is not available. In this case, we adopted a different test strategy, starting from the structure of the spike in the unbound form of both the wild type (PDB id:7KJ5) and South African variant (PDB id: 7LYN). In particular, we performed three additional, 500 ns-long molecular dynamics simulations starting from (i) the experimental APO form of the RBD of the South African variant; (ii) a computationally-obtained APO conformation extracted from the experimental WT APO conformation of the spike trimer, computationally mutated into the South African variant, and (iii) the spike RBD bound conformation (obtained from the computationally-mutated complex but removing ACE2).

Then, we compared the dynamical behavior of the spike protein in all three cases with respect to the complex simulation. In particular, Fig. 2c shows the distribution of root mean square displacement (RMSD) among a sampled set of complex-spike conformations (purple curve) and between the complex configurations and a set of configurations extracted from each of the three free-spike dynamics. As one can see, all distributions partially overlap in the region around 1.5–2.0 Å, meaning that they can explore a similar configuration space. As we may expect, the mean RMSD progressively increases from the complex distribution to the computationally-mutated apo spike. Complete RMSD data are shown in Supplementary Fig. 1.

Finally, identifying the centroids of the clustering analysis of the sampled configurations, we performed a molecular docking with the ACE2 receptor structure.

According to silhouette analysis (see Supplementary Fig. 2), the frames of the entire trajectory can be grouped into three different groups, the centroids of which are the most representative structures of the entire trajectory. A docking analysis between ACE2 receptor structure and each of the centroid spike configurations was performed using Hdock web-server³⁶, where the two interacting regions were constrained (see Method section for more details). The analysis did not show important changes for the three analyzed configurations, therefore only the results related to the centroid belonging to the most populated cluster are reported. The 500 structures sampled from the trajectory and the first 100 docking poses were chained together and a Principal Component Analysis was performed. Each structure was then projected onto the essential plane defined by the first two main eigenvectors (whose eigenvalues explain the DD of the total variance). As expected, the structures of the trajectory take place in a restricted portion of the essential plane, while the structures coming from the docking occupy a greater space (see Supplementary Fig. 3). However, some of the docking structures are very similar to the structures sampled during molecular dynamics. To better appreciate this structural similarity, we consider 16 molecular dynamics structures (which are temporally equidistant) and the 16 docking structures closest to the dynamics structures. The contour lines in Fig. 2d show the probability density of finding a structure in a given portion of the essential plane. In fact, as expected, the region is less dense where the docking poses are projected (orange points) with respect to molecular dynamics frames. To better visualize the molecular docking results, we have extracted the two most representative conformations among the docking poses, one closest to the structures sampled in molecular dynamics and one further away from the trajectory frames and closer to the other docking poses. The results, in terms of structural visualization, are shown in Fig. 2e, f. Note that even in the case of the most distant poses (in blue), the binding conformation is very similar to the one most representative of molecular dynamics (in gray).

Overall, all the performed analyses confirmed that the obtained results are not significantly affected by the choice of a different

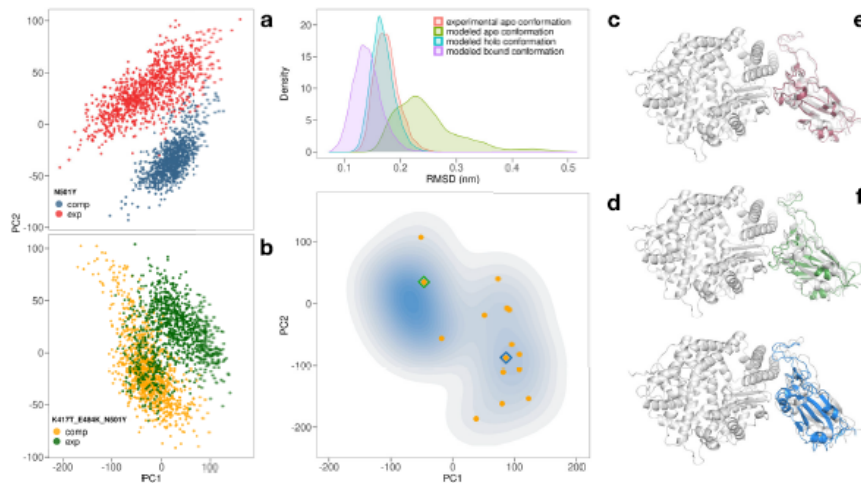


Fig. 2 Comparison between experimental and modeled SARS-CoV-2 Spike-Ace2 complexes. **a** Comparison between the molecular dynamics simulation of the experimental and computationally-obtained complex of the English variant of SARS-CoV-2 spike protein bound to ACE2 receptor. Snapshots of the spike protein are projected in the plane of the two major components of the covariance matrix obtained from a principal component analysis (PCA) over the position covariance matrix. **b** Same as in panel (a) but for the SARS-CoV-2's Amazonian variant. **c** Distribution of the root mean square deviation (RMSD) of the spike protein of the SARS-CoV-2 South African variant in apo and holo (bound to ACE2 receptor) forms. Fifty sampled frames of the spike protein from the computationally-obtained complex were used as references to evaluate the RMSD. **d** Probability density of finding a configuration of the apo spike protein in the plane identified by the two principal components of the covariance matrix. Colored dots represent the position of the configuration produced by the docking procedure when projected in the plane. **e** Cartoon representation of SARS-CoV2 spike frame extracted from the apo simulation having smallest RMSD. **f** Cartoon representation of two SARS-CoV2 spike-ACE2 complexes obtained from the docking procedure.

simulation starting point, being it either an experimental structure or a computationally mutated one. Indeed, this is testified both by the comparison between mutated vs experimental complexes of the English and Amazonian variants and by the analyses on the MD of the South African variant's RBD in apo and mutated-holo forms. We might speculate that the peculiar stability of the RBD fold is motivated by the fact that the spike protein has to continuously optimize its binding site to infect different hosts in different conditions; thus it could be optimal to have a stable fold upon single/few point mutations.

To ensure a complete equilibration of the systems, only configurations after 250 ns are used in the analyses.

Fluctuation of interface residues for the different variants. The first observable, we consider to evaluate the stability of each RBD-ACE2 complex, was the root mean square fluctuation (RMSF). In Fig. 3a, we show for each residue, both for spike protein and for the ACE2 receptor, the RMSF value obtained considering each molecular dynamics simulation at equilibrium. The average of the RMSF of all contact residues provides information on the overall mobility of the interface. The binding residues of ACE2 and spike have a comparable mean fluctuation value between them (of about 1.25 Å, as the average of all systems). The negative control 1 and 2 systems (whose mutations are N501K and N501D, respectively) are characterized by a higher binding affinity (quantified by the dissociation constant, K_d) than the WT form of the spike. This is clearly evident in terms of atomic fluctuation, given that these systems show greater mobility of the interface residues, with an RMSF average of 1.53 Å and 1.31 Å,

respectively. The system characterized only by N501Y mutation (UK variant) is the most stable of all, with an average of 1.01 Å. The Amazonian variant and the South African variant also have a low average fluctuation compared to the other systems, of 1.14 ± 0.23 and 1.22 ± 0.20 Å, respectively. A more detailed analysis was performed on the atomic fluctuation of residue 501. In all systems, except for the South African variant, residue 501 of the spike protein has a lower average fluctuation than the average of all residues belonging to the same interface. In particular, the English variant, the Amazonian variant, and the positive control mutation (N501F) show a fluctuation of the residue 501 lower than the average of the RMSF values. Indeed, the RMSF value for residue 501 in these three systems is 1.00, 1.00, and 1.20, while the average values of RMSF for the three systems are 1.01 ± 0.16 , 1.14 ± 0.23 , and 1.25 ± 0.20 , respectively. Therefore, for these three systems, residue 501 is particularly stable with respect to the other residues. On the contrary, the other systems have a more pronounced fluctuation of residue 501. Among all, the two negative control systems, N501K and N501D mutations are characterized by a very high average RMSF value: 1.83 and 2.03 respectively. The results of the analysis of atomic fluctuation over time of the interface residues highlight the stability of systems with higher binding affinity, including the South African and Amazonian variants for which we have not considered experimental data (since they are not present in ref. ²⁷). In order to investigate the stability of intermolecular interactions during molecular dynamics, we calculated the contact frequency matrix for each system. Therefore, we define contact between two residues if their α -carbon atoms have a distance less than 9 Å. For

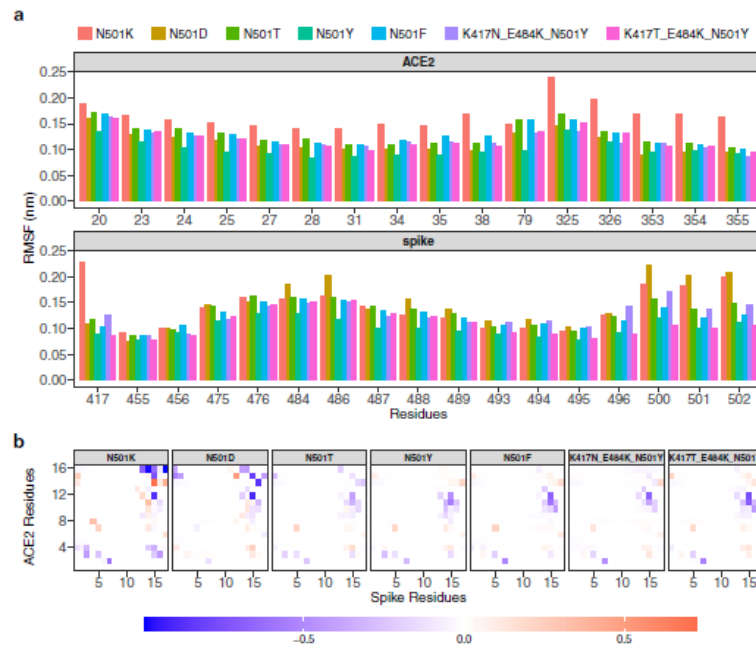


Fig. 3 Analysis of the motion of the residues of the binding region. **a** Root mean square fluctuations (RMSF) of ACE2 (top) and SARS-CoV-2 spike (bottom) protein residues found in interaction during the dynamics. Different colors corresponds to different spike variants (see Table 1). **b** Difference between the contact probability matrices of the interacting residues between each spike variant and the Wuhan (WT) one (see Table 1). Residues are considered as interacting if the distance between their α -carbons is lower than 9 Å.

Table 2 Correlation matrix between couples of variants. Correlation values (resp. p values) are reported in the upper (resp. lower) triangular matrix.

	K417N E484K N501Y	K417T E484K N501Y	N501D	N501F	N501K	N501T	N501Y
K417N-E484K - N501Y	-	0.91	0.44	0.92	0.38	0.67	0.77
K417T-E484K - N501Y	10^{-14}	-	0.20	0.96	0.31	0.61	0.87
N501D	3.6×10^{-14}	8.1×10^{-4}	-	0.23	0.35	0.52	0.17
N501F	10^{-14}	10^{-14}	9.8×10^{-5}	-	0.34	0.67	0.85
N501K	8.7×10^{-11}	2.6×10^{-7}	4.6×10^{-9}	1.4×10^{-8}	-	0.28	0.26
N501T	10^{-14}	10^{-14}	10^{-14}	10^{-14}	2.4×10^{-6}	-	0.66
N501Y	10^{-14}	10^{-14}	5.1×10^{-3}	10^{-14}	9.8×10^{-6}	10^{-14}	-

each matrix element, we then report the contact frequency between each residue of the spike protein and each other of the ACE2 receptor, then subtracting (to facilitate the comparison) each of these matrices with that obtained for the WT system (see Methods section). As shown in Fig. 3b, we notice similarities between some matrices. For example, the matrix relating to the N501Y system is particularly similar to the N501F system, which is characterized by the highest experimental binding affinity value. An analysis of the Pearson correlation between each pair of matrices allows us to quantify this evidence. In Table 2 the

correlation values between all the contact matrices are shown. Interestingly, the positive control system (characterized by the N501F mutation) has a mean contact map highly correlated with that of the UK, Amazon, and South African variants, with a Pearson correlation value of 0.96, 0.92, and 0.85 respectively. The two systems formed by the two mutations with low K_d (the two negative controls), on the other hand, show lower correlation values with any other system, with an average of the Pearson coefficient of 0.34 and 0.32 respectively. These results represent the first level of classification of the mutated systems of this study

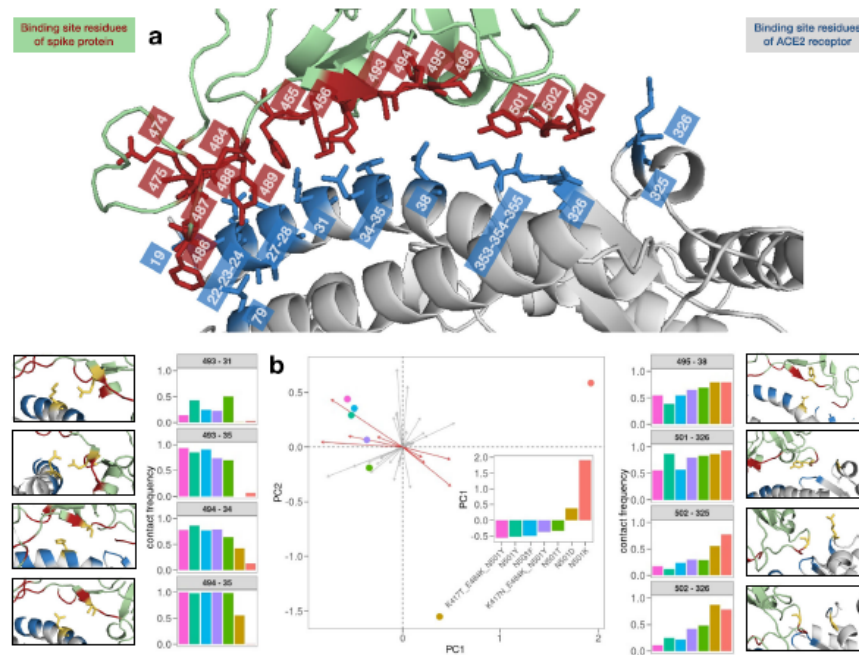


Fig. 4 Analysis of the covariance in the motion of the interacting residues. **a** Ribbon representation of the binding site of the SARS-CoV-2 spike protein bound to human ACE2 (PDB id: 6M0J). Interacting residues, i.e., the ones whose α -carbons have a distance lower than 9 Å, are represented with red sticks for the spike protein and with blue ones for the ACE2 receptor. **b** Projections on the two principal components of the analyzed variants (colored dots). Principal component analysis was performed over the contact probabilities of each couple of interacting residues reported in panel (a). The inset shows the projections on the first component. Side plots display the contact probability values for the eight couples of residues that contributed the most in differentiating the variants in the principal component analysis.

and allow us to identify the properties of mean atomic fluctuation at a single residue level.

Principal component analysis of contact frequency. In order to obtain a more exhaustive overview of the binding between ACE2 and spike protein, we investigate the frequency of the contacts for each analyzed system. To this end, we evaluated the intermolecular contact pairs for each complex, i.e., we consider the percentage of contacts each residue of the spike binding region forms with ACE2 ones during the simulation. The binding region is composed of 17 residues of the spike protein and 16 belonging to the ACE2 receptor (see Fig. 4a and Methods). Therefore, each system can be described by a vector of 272 contact frequencies. To compactly compare those vectors, we performed a principal component analysis of the vectors' covariance. The projection of each system on the essential plane of the first two principal components (which explain the 54.4 and 35.3% of the total variance, respectively) allows to clearly distinguish, completely unsupervised, the mutations with high affinity from those with low binding affinity. In particular, the projection along the first principal component shows a very interesting trend (as shown in Fig. 4b): the two mutations of the negative control have a

positive value on the first component (distinguishing themselves from any other). On the other hand, the first three systems along this component are the Amazonian variant, the English variant, and the positive control (N501F). We also analyze the loading of each interacting residue pair on the first principal component. In Fig. 4 we show the considered pairs, given that their projection along the axis of the first component is high compared to the others (and the trend of the contact frequency is not trivial between the different systems). In most cases, the pairs of interacting residues at the interface show a trend (increasing or decreasing) of the frequency of the contacts for the different systems. For example, the percentage of contacts during the simulations of the pair composed of the S494 residue of the spike protein and the H34 residue of the ACE2 receptor (S494-H34), progressively decreases for systems with low binding affinity. Therefore, this pair is more stable in high binding affinity systems. With an opposite trend, the pair of residues G502-Q325 shows an increasing trend, meaning that this is more present for systems with low binding affinity. This analysis provides information on which pairs of intermolecular interactions are more stable in high-affinity binding systems and which pairs, on the other hand, are most present in low-affinity binding systems.

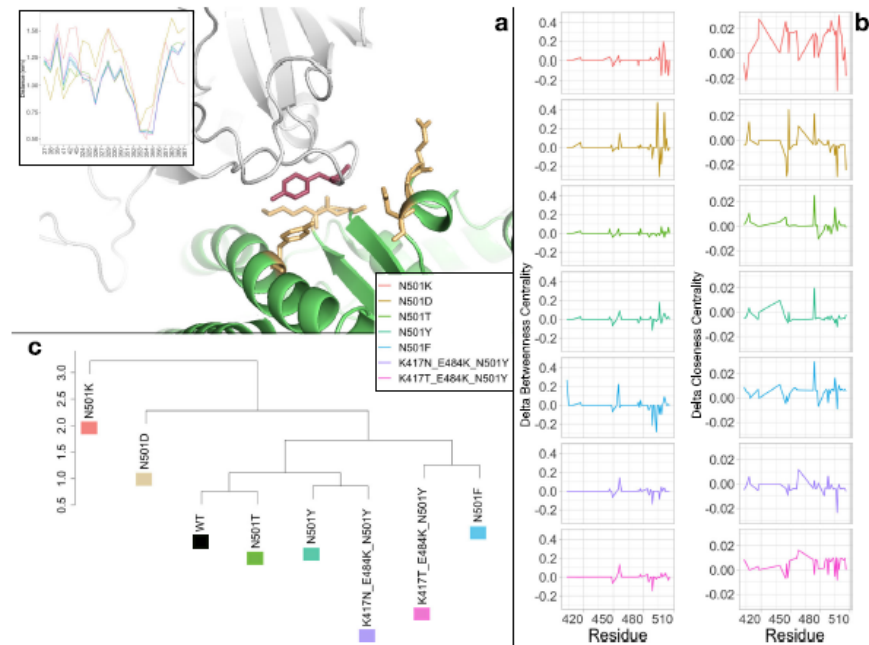


Fig. 5 Graph analysis of the interacting residues' network. **a** Cartoon representation of the complex (PDB id:6M0J) between SARS-CoV-2 spike protein (gray) and the human ACE2 receptor (green). Residue 501 of the spike protein is represented in red sticks, while a set of interacting residues of the ACE2 protein are identified in yellow sticks on the basis of a spatial distance threshold of 9 Å. The inset shows the average minimum distance of the ACE2 interacting residues with residue 501 of the spike protein for each variant. **b** Difference in Betweenness (left column) and Closeness (right column) centrality of each node of the SARS-CoV-2 spike found in interaction with ACE2 receptor between the seven studied spike variants and the WT one (see Table 1). **c** Hierarchical clustering of the seven spike variants together with the WT reference one (see Table 1).

Graph theory-based analysis of binding site residues. A higher level of complexity of the binding properties was addressed by analyzing the organization of interactions between the interface residues. To this end, we model the interaction between the two proteins as a bipartite network, schematizing each residue as a node of the network and each intermolecular interaction as a network edge. Only intermolecular interactions were considered for this analysis, so interactions between two residues belonging to the same protein are not included in the graph definition. In particular, we define a weighted graph for each system, weighing each link connecting two residues with the corresponding contact frequency calculated from the molecular dynamics configurations. In this case, we defined a contact between two residues if their distance is less than 8.5 Å in agreement with Chakrabarty et al.³⁷, where similar network analyses were performed on protein structures to assess the best threshold for evaluating the centroid-centroid contacts.

As an example for a better interpretation of this modeling for the molecular complex, we report the distances that every residue belonging to the ACE2 binding site has with the residue 501 of the spike protein since this is of primary importance in the variants considered in this work (see Fig. 5a). The mean distance between residue 501 of the spike protein and any other (within 12 Å) of the ACE2 receptor was calculated. This analysis shows

that the systems with lower experimental binding affinity are typically characterized by a greater distance (N501K and N501D single mutation) between the residues belonging to the binding site of ACE2 and the residue 501 of the spike protein. Interestingly, the system with the N501D mutation has a shorter distance than any other system between residues 501 and residues 37, 38, 39, 41, 42, and 45 of the receptor. Similarly, for the system having the N501K mutation, it shows the lowest distance between residue 501 and residue 383, 386, and 387 belonging to the ACE2 receptor. Extending the analysis to any other residue of the spike protein interface, we investigate the local organization of the intermolecular contacts through centrality measures. To this end, both the *Betweenness centrality* and *Closeness centrality* parameters (see Methods) were considered as a local descriptor of the interaction of each residue, since these are certainly two of the most widely used descriptors for the centrality analysis of a node³⁸ (See Fig. 5b).

In order to compare the different interaction organizations of each molecular system, we performed a clustering analysis (see Fig. 5c) considering a single vector for each system, which is composed of the combination of the two descriptors considered. The profile is relative only to the residues belonging to the spike interface (see Fig. 4a).

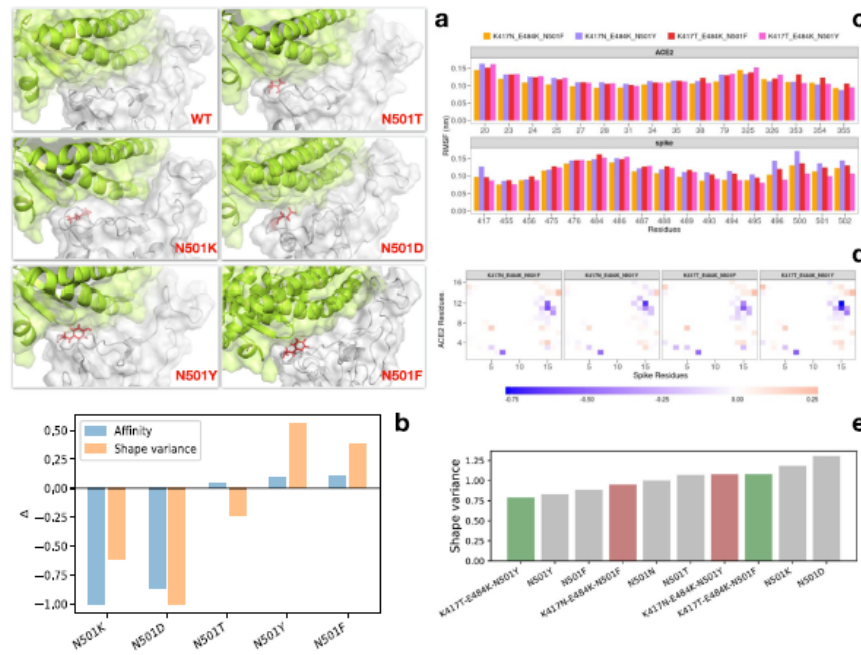


Fig. 6 Analysis of the shape complementary of the binding regions. **a** Cartoon representations of the binding region of the complex formed by SARS-CoV-2 spike protein (gray) and human ACE2 receptor (green) for the wild-type complex and the five single-mutation variants. The spike residue 501 is represented with red sticks. The molecular surfaces for both spike and ACE2 are shown. **b** Experimental affinity values (taken from ref. 27) for the spike-ACE2 complexes (blue bars) and shape variance as measured by the Zernike descriptors on molecular dynamics configurations (orange bars). Both quantities are obtained by difference with respect to the reference complex data (see Table 1). **c** Root mean square fluctuations (RMSF) of ACE2 (top) and SARS-CoV-2 spike (bottom) protein residues found in interaction during the dynamics for the South African and Amazonian variants together with the two versions carrying phenylalanine at position 501. **d** Difference between the contact probability matrices of the interacting residues between each three-mutation spike variant and the Wuhan (WT) one (see Table 1). Residues are considered as interacting if the distance between their α -carbons is lower than 9 Å. **e** Shape variance as measured by the Zernike descriptors on molecular dynamics configurations for all the variants reported in Table 1.

Also in this analysis, we find a clear separation of mutations with lower binding affinity from the others, showing the ability of our molecular dynamics-based approach to distinguish between high and low-affinity systems. Interestingly, the WT form of the complex occupies an intermediate position between low-affinity and high-affinity mutations. In particular, the WT system is characterized by a profile very similar to the N501T mutation, which has the same amino acid in position 501 of the SARS-CoV system.

On the other side, the Amazonian variant and the South African variant appear to have a very similar organization of contacts with the positive control and the English variant, respectively. Interesting to note that the Amazonian variant, which is causing an important concern in the world, is very similar to the system characterized by the single mutation N501F, the one with the best experimental binding affinity. Similarly, the South African and English variants are in close proximity in terms of their Betweenness and Closeness properties, as they exhibit similar behavior in terms of infection transmission.

Shape variation of binding site region. Finally, we focused on the whole binding region of both SARS-CoV-2 spike protein and ACE2 receptor and assessed the shape complementarity of the binding site. To do so, we first expanded the molecular surfaces of the two molecular partners (see Fig. 6a) on the basis of the 3D Zernike moments (see Methods) and then computed the Euclidean distance between the two sets of invariant descriptors. We repeated the procedure for 250 configurations sampled from the equilibrium of each of the molecular dynamics simulations we performed. We thus obtained a distribution of Zernike distances, i.e., of complementarity scores, for each investigated variant. As one could expect, all distributions are centered around similar values of complementarity since the overall shape of the binding site does not undergo substantial conformational changes upon a few point mutations. However, looking at the variance of the distributions for the five single-mutation variants, we found a similar trend between the experimentally measured complex affinities and the variance of the shape complementarity³⁹. This can be seen from Fig. 6b, where we compared the difference in

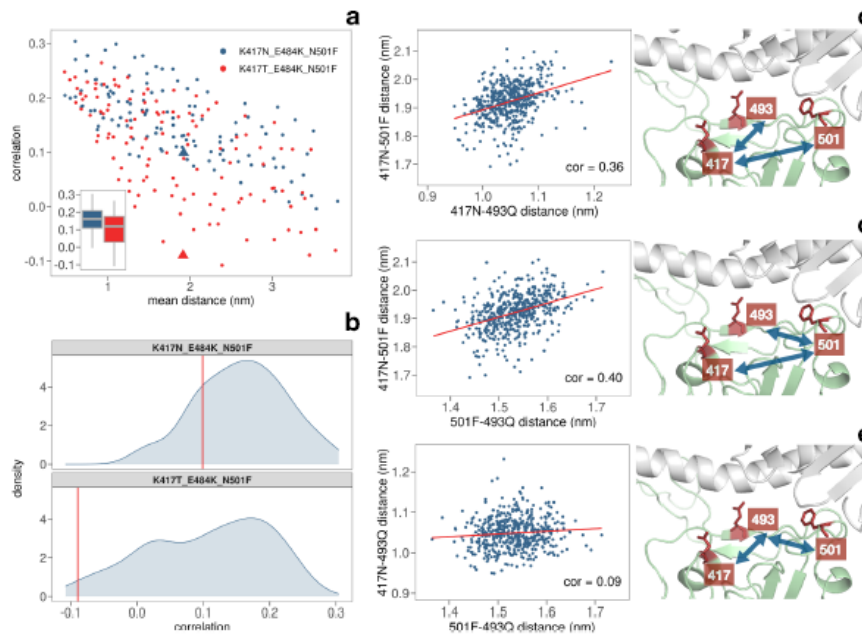


Fig. 7 Analysis on the residues' motion in variants carrying the 501F mutation. **a** Spatial correlation of the motion of all possible couples of residues of the spike protein as a function of their average spatial distance (of the residue centroids) during the simulation. Blue dots correspond to the residues of the South African variant with the putative N501F additional mutation with red dots represent all the couples of residues of the Amazonian variant with the N501F mutation. Box plots of the correlations are shown in the inset. **b** Distribution of the correlations in the motion of all the couples of residues having a sequence distance higher than three residues. Red vertical lines mark the correlation of residues 417–501. **c** Distance between the centroids of residues 417–501 vs residues 417–493 during the molecular dynamics simulation of the South African variant with the putative N501F additional mutation. Cartoon representation of the binding region between spike (green) and ACE2 receptor (gray), highlighting in red the key residues 417, 493, and 501. **d** Same as in (c) but for couples 417–501 vs 493–501. **e** Same as in (c) but for couples 417–493 vs 493–501.

affinity and shape variance of the five investigated variants with respect to the wild type. It is interesting to note that the N501Y variant displays a lower shape variance with respect to the wild type than the N501F variant. Unfortunately, as far as our knowledge goes, no binding affinity data are available for the three-mutation variants (i.e. the South African and Amazonian variants). We thus proceeded to compute the shape variance from the molecular dynamics configurations and compare them with single-mutation ones. Interestingly, we found that both variants show a lower variance with respect to the negative controls (N501D and N501K), with the South African variant being more motile than the English one while the Amazonian variant displaying the smallest variance (see Fig. 6c–d). This result provides an important example of the not-trivial effect of cooperativity. In fact, the same mutation (Y501F) gives an opposite outcome in the two variants, which now differ in terms of the interactions between a couple of residues 417T–501F against 417N–501F. Finally, we run two additional molecular dynamics simulations substituting amino acid F to Y in the two triple-mutated variants in order to check whether such mutation could bring an enhancement in the binding propensity. Looking again at the shape variance (see Fig. 6e), we observed that the modified South African and Amazonian variants behave oppositely: while a Y to

F mutation on the 501 residue increases the complementarity variability for the Amazonian variant thus worsening its binding capacity, the same mutation on the South African one brings a stabilization effect, with the complex maintaining a more stable shape complementarity during the dynamics.

Analysis of motion between the key residues for the putative variants carrying the 501F mutation. To better elucidate the non-trivial interplay between residue 501 and 417, we performed a correlation analysis between the motions of the binding site residues, both considering the case in which the mutation increases the stability (South African variant with additional 501F mutation) and the one where it decreases (Amazonian variant carrying the additional 501F mutation). Results are displayed in Fig. 7a.

Indeed, the average correlation of the residues belonging to the South African variant and the Amazon variant is 0.16 and 0.10 respectively. Notably, this substantial difference in motion properties of the whole spike binding region is caused by just one different residue in position 417.

We then focused on the 501–417 residue pair and their cooperative action. In Fig. 7b, we show the correlations of the

501–417 residue pair in the two systems with respect to the distribution of the correlations between all pairs of spike residues forming the binding region. In particular, only residues that are more than two residues apart in sequence have been selected to remove trivial correlation due to close proximity. Interestingly, the 501F–417N pair in the modified South African variant (predicted to be more stable) has a positive correlation of 0.10, while the 501F–417T residue pair in the modified Amazonian variant (predicted to give a less stable binding) has a negative correlation of -0.09 .

Finally, we consider the correlated motion that the two residues have with the other residues of the binding region. Considering the distance between the centroids of the residue side chains, we compared the trend of the distance of each residue pair involving residue 417 and/or 501 as a function of time. Next, we looked for couples having a high correlation in the modified South African variant and a low correlation in the modified Amazonian system (see Supplementary Table 1). Interestingly, only two pairs were selected and they both involve the same residue, i.e., residue Q493.

In particular, if one looks at the modified South African variant, the pair of residues 417–493 has a high Pearson correlation value (0.36) with the pair 417–5017, thus indicating a possible functional role (see Fig. 7c). Similarly, the 417–501 and 501–493 residue pairs display a Pearson correlation value of 0.40 as one can see from Fig. 7d. On the other hand, the pairs formed by residues 501 and 417 with residue 493 do not move in a correlated way (Fig. 7e), showing that residues 417 and 501 shares a direct correlation. Overall, our analyses add further residue-level evidence of the cooperative effect of residues 417–501 and suggest that residue 493 could be another important residue involved in the binding stability of the complex⁴⁰.

Conclusions

The major goals of a virus are replication and spread, and SARS-CoV-2 coronavirus is constituting no exception. The vast, worldwide, diffusion of the SARS-CoV-2 epidemics indeed is providing the virus the possibility of exploring the genomic landscape and accumulating mutations. The resulting viral variants are subject to natural selection; thus, mutations that increase the diffusion rapidly get fixed in the viral genomic pool. Besides the transmission of the virus itself, antiviral treatments also contribute to selection. Indeed with the introduction of vaccines, new mutations inducing escape and resistance to available treatments are expected to couple with more virulent strains^{41–43}. The new frontier for fighting the COVID-19 pandemic seems to become even more based on controlling and understanding the accumulation of variations in the SARS-CoV-2 genome.

To date, several different variants have already emerged. In particular, those with mutations on the spike protein are attracting a lot of attention as this protein is responsible for the binding to cellular receptor and attachment factors but also the primary target of the antibodies of the immune system.

Here, we deployed a set of molecular dynamics simulations to characterize the effect of different possible mutations of the spike residues 417, 484, and 501, which are the mutations found in the most relevant observed variants. In particular, mutation N501Y has been found in both the English, South African, and Amazonian variants. Analyzing the equilibrium configurations acquired by the spike-ACE2 complex in terms of residue fluctuations, networks of contacts, and conservation of shape complementary of the binding region, we found that indeed mutations on residue 501 strongly influence the dynamical stability of the complex. Most importantly, a phenylalanine substitution in position 501 increases the stability of the South

African variant via a cooperative action with residue 417N. In conclusion, our results suggest close surveillance for the emergence of such mutation to anticipate (and minimize) its effects in the viral spread and eventual early incorporation of this information into diagnostic and pharmacological procedures.

Methods

Structural data. The complex of SARS-CoV-2 spike protein bound to the human ACE2 receptor has been taken from the PDB bank (PDB id: 6M0J). In particular, only the receptor-binding domain (RBD) of the spike and the extracellular domain of ACE2 are considered. Each variant considered in the present study has been obtained manually mutating the experimental complex via the PyMol software³⁵. All information about amino acid substitutions is reported in Table 1.

Four additional structures were used to validate our procedure, i.e.,

- the spike protein in trimeric APO form of SARS-CoV-2 South African variant (PDB code 7LYN);
- the spike protein in trimeric APO form of SARS-CoV-2 WT (PDB id: 7KJ5);
- the experimental molecular complexes of the SARS-CoV-2 RBD bound to human ACE2 for the English variant (PDB id: 7MJN);
- and the experimental molecular complexes of the SARS-CoV-2 RBD bound to human ACE2 for the Amazonian variant (PDB id: 7NXC).

Molecular dynamics simulations. All simulations were performed using Gromacs⁴⁴. Topologies of the system were built using the CHARMM-27 force field⁴⁵. The protein was placed in a dodecahedral simulation box, with periodic boundary conditions, filled with TIP3P water molecules⁴⁶. For all simulated systems, we checked that each atom of the proteins was at least at a distance of 1.1 nm from the box borders. Each system was then minimized with the steepest descent algorithm. Next, a relaxation of water molecules and thermalization of the system was run in NVT and NPT environments each for 0.1 ns at 2 fs time-step. The temperature was kept constant at 300 K with a v-rescale thermostat⁴⁷; the final pressure was fixed at 1 bar with the Parrinello–Rahman barostat⁴⁸.

LINCS algorithm⁴⁹ was used to constraint bonds involving hydrogen atoms. A cut-off of 12 Å was imposed for the evaluation of short-range non-bonded interactions and the Particle Mesh Ewald method⁵⁰ for the long-range electrostatic interactions. The described procedure was used for all the performed simulations.

Statistics and reproducibility. All molecular dynamics simulations were 500 ns long, which guarantees that every system has reached equilibrium conformations. All subsequent analysis were performed sampling 500 frames from the equilibrium range.

Molecular docking protocol. Molecular docking between the unbound conformation of spike RBD and ACE2 receptor has been performed using HDOCK software⁵¹. Apo form of RBD with K417N-E484K-N501Y mutations, is obtained by selecting the B chain from the trimeric form of the spike protein (PDB code: 7KJ5). Indeed, the B chain has the least number of missing residues with respect to the other two chains. The missing residues were modeled using SwissModel⁵² in order to obtain the complete structure of a single chain of the spike protein. Subsequently, we performed the K417N, E484K, and N501Y mutations using the PyMol software³⁵. Starting from this structure, we performed 500 ns of MD simulation for extrapolating the most representative structures through a clustering analysis on the first two principal components of the PCA. The clustering provided three representative structures, on which we performed the molecular docking procedure. On the other hand, the ACE2 receptor was selected from the spike-ACE2 complex (PDB code: 6M0J). Constraints on the binding residues were applied to molecular docking, selecting a large possible interaction region (116 residues and 194 residues for RBD and ACE2 interface respectively), using the known information on the binding regions.

Interface residue definition and contact probability calculation. For each analyzed complex, the interface was defined by taking ACE2 and SARS-CoV-2 spike residues whose α -carbons have a distance lower than 12 Å at time $t = 250$ ns, that is, after the equilibration phase. To end with a comparable set for all complexes, we selected the residues common to all interfaces obtaining 16 residuals for ACE2 and 17 for the spike protein. For each couple of interacting residues among the two proteins, we calculated the contact frequency, counting how many times each couple of residues had a distance lower than 9 Å between two α -carbons in each frame of the dynamics at equilibrium.

We got a 16×17 matrix for every complex reported in Table 1. Contact matrices shown in Fig. 3b are obtained by subtracting the wild-type one.

Principal component analysis and clustering. From the contact matrices of all complexes, we performed a principal component analysis (PCA) in which the starting matrix consisted of seven rows (the variants) and 272 columns (the contact frequencies of each pair of spike-ACE2 α -carbons). The clustering analysis was performed using the "hclust" function of R, preserving the default clustering algorithm (the "complete" method). We first computed the contact percentage matrix for each system, as obtained from the molecular dynamics frames. For graphics analysis, we define contact between two residues if the distance between their center is less than 8.5 Å, as proposed in ref. 37. Then, we defined a weighted graph for each matrix and calculated the betweenness centrality and closeness centrality parameters for each residue.

In particular, the betweenness centrality of a node i is given by:

$$b_i = \sum_{k \neq i \neq l} \frac{s_k(i)}{s_k} \quad (1)$$

where $s_k(i)$ is the number of weighted shortest paths (s) that go from node k to node l passing through node i while s_k is the total number of weighted shortest paths from node k to node l .

Similarly, the closeness centrality of a node i is defined by the inverse of the average length ($\langle c \rangle$) of the weighted shortest paths to/from all the other nodes in the network:

$$c_i = \frac{1}{\langle c_{ki} \rangle_{k \neq i}} \quad (2)$$

Both quantities were obtained via the corresponding functions of the "igraph" package of R³³.

For each system, we combined the two vectors (which have values of betweenness and closeness for single residues) normalized to 1. We used the Euclidean distance to compare every pair of vectors. Analyses were performed using R standard libraries³⁴.

Molecular surface analysis via Zernike descriptors. Given an ACE2-RBD complex simulated in molecular dynamics, for each protein, we calculate separately the molecular surface using DMS software³⁵. Once extracted the portion of protein surface in interaction, with a voxelization procedure we represent the protein patch as a 3D function.

This 3D function can be described as a series expansion on the basis of the 3D Zernike Polynomials^{36,38,37}. Taking the norm of the expansion coefficients we deal with an ordered set of numerical descriptors that compactly summarize the shape of the examined molecular surface.

Indeed, a function $f(r, \theta, \phi)$ can be written as:

$$f(r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{l=0}^n \sum_{m=-l}^l C_{nlm} Z_{nlm}^*(r, \theta, \phi) \quad (3)$$

where Z_{nlm}^* are the 3D Zernike polynomials, while the coefficients C_{nlm} are called Zernike moments.

The precision of the description can be selected by modifying the order of the expansion N . In this work, we fix $N=20$, corresponding to 121 numerical descriptors representing such function.

The 3D Zernike Moments can be seen as:

$$C_{nlm} = \int_{|r| \leq 1} f(r) \overline{Z_{nlm}(r, \theta, \phi)} dr \quad (4)$$

where \overline{Z} represent the complex conjugate.

The norms of such moments, with respect to the index m , are invariant under translation and rotation. Indeed, the Zernike Descriptors are defined as:

$$D_{nl} = \|C_{nlm}\| = \sqrt{\sum_{m=-l}^l |C_{nlm}|^2} \quad (5)$$

The shape complementarity between two surfaces can be easily evaluated by applying a metric between the two vectors of numbers describing them^{39,38,39}. Indeed, we adopted the euclidean distance. When two surfaces have a low distance between them, they are characterized by a similar shape and therefore they are suitable for binding.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

All codes used to produce the findings of this study are available from the corresponding author upon request. The code for the Zernike algorithm is available at <https://github.com/matmi8/Zernike3D>.

Received: 27 April 2021; Accepted: 26 November 2021;

Published online: 06 January 2022

References

- Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *Lancet* **395**, 497–506 (2020).
- Zhu, N. et al. A novel coronavirus from patients with pneumonia in china, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
- Fontanet, A. et al. Sars-cov-2 variants and ending the covid-19 pandemic. *Lancet* **397**, 952–954 (2021).
- Duchene, S. et al. Temporal signal and the phylogenetic threshold of sars-cov-2. *Virus Evol.* **6**, veaa061 (2020).
- Portelli, S. et al. Exploring the structural distribution of genetic variation in sars-cov-2 with the covid-3d online resource. *Nat. Genet.* **52**, 999–1001 (2020).
- Plante, J. A. et al. Spike mutation d614g alters sars-cov-2 fitness. *Nature* **592**, 116–121 (2020).
- Zhang, J. et al. Structural impact on sars-cov-2 spike protein by d614g substitution. *Science* **372**, 525–530 (2021).
- Trucchi, E. et al. Population dynamics and structural effects at short and long range support the hypothesis of the selective advantage of the g614 sars-cov-2 spike variant. *Mol. Biol. Evol.* **38**, 1966–1979 (2021).
- Baric, R. S. Emergence of a highly fit SARS-CoV-2 variant. *N. Engl. J. Med.* **383**, 2684–2686 (2020).
- Tang, J. W., Tambyah, P. A. & Hui, D. S. Emergence of a new sars-cov-2 variant in the UK. *J. Infect.* **82**, e27–e28 (2020).
- Galloway, S. E. et al. Emergence of sars-cov-2 b. 1.1. 7 lineage—united states, december 29, 2020–january 12, 2021. *MMWR* **70**, 95 (2021).
- Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269, <https://doi.org/10.1038/s41586-021-03470-x> (2021).
- Challen, R. et al. Risk of mortality in patients infected with sars-cov-2 variant of concern 202012/1: matched cohort study. *BMJ* **372**, n579 (2021).
- Chand, M. et al. Investigation of novel sars-cov-2 variant: variant of concern 202012/01 (pdf). *Public Health England*, p. 1–18 (2020).
- Tegally, H. et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443, <https://doi.org/10.1038/s41586-021-03402-9> (2021).
- Naveca, F. et al. Sars-cov-2 reinfection by the new variant of concern (voc) p. 1 in Amazonas, Brazil. Preprint at <https://virological.org/t/sars-cov-2-reinfection-by-the-new-variant-of-concern-voc-p-1-in-amazonas-brazil/596>, 1–18 (2021).
- Hulawit, R. J. et al. Human coronaviruses oc43 and hku1 bind to 9-o-acetylated sialic acids via a conserved receptor-binding site in spike protein domain 1. *Proc. Natl Acad. Sci. USA* **116**, 2681–2690 (2019).
- Schwegmann-Wesfels, C. & Herberich, G. Sialic acids as receptor determinants for coronaviruses. *Glycoconj. J.* **23**, 51–58 (2006).
- Milanetti, E. et al. In-silico evidence for a two receptor based strategy of SARS-CoV-2. *Front. Mol. Biosci.* **8**, 690655 (2021).
- Bò, L., Miotta, M., Di Rienzo, L., Milanetti, E. & Ruocco, G. Exploring the association between sialic acid and sars-cov-2 spike protein through a molecular dynamics-based approach. *Front. Mol. Technol.* **2**, 24 (2020).
- Baker, A. N. et al. The sars-cov-2 spike protein binds sialic acid and enables rapid detection in a lateral flow point of care diagnostic device. *ACS Cent. Sci.* **6**, 2046–2052 (2020).
- Schoeman, D. & Hedding, B. C. Coronavirus envelope protein: current knowledge. *Virus* **16**, 1–22 (2019).
- Zhu, Z. et al. From sars and mers to covid-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir. Res.* **21**, 1–14 (2020).
- Ou, X. et al. Characterization of spike glycoprotein of sars-cov-2 on virus entry and its immune cross-reactivity with sars-cov. *Nat. Commun.* **11**, 1–12 (2020).
- Yan, R. et al. Structural basis for the recognition of the sars-cov-2 by full-length human ace2. *Science* **367**, 1444–1448 (2020).
- Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- Starr, T. N. et al. Deep mutational scanning of sars-cov-2 receptor binding domain reveals constraints on folding and ace2 binding. *Cold Spring Harbor Lab. Perspect. Biol.* **12**, 1295–1310 (2020).
- Di Rienzo, L., Milanetti, E., Lepore, R., Olimpieri, P. P. & Tramontano, A. Superposition-free comparison and clustering of antibody binding sites: implications for the prediction of the nature of their antigen. *Sci. Rep.* **7**, 1–10 (2017).

29. Di Rienzo, L., Milanetti, E., Alba, J. & D'Abramo, M. Quantitative characterization of binding pockets and binding complementarity by means of zernike descriptors. *J. Chem. Inf. Model.* **60**, 1390–1398 (2020).
30. Milanetti, E. et al. 2d Zernike polynomial expansion: finding the protein-protein binding regions. *Comput. Struct. Biotechnol. J.* **19**, 29–36 (2021).
31. Miozzo, M. et al. Molecular mechanisms behind anti SARS-CoV-2 action of lactoferrin. *Front. Mol. Biosci.* **8**, 607443 (2021).
32. Ali, A. & Vijayan, R. Dynamics of the ace2-sars-cov-2/sars-cov spike protein interface reveal unique mechanisms. *Sci. Rep.* **10**, 1–12 (2020).
33. Kerber, B. et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827.e19 (2020).
34. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
35. The PyMOL molecular graphics system, version 1.8 (LLC, Schrödinger LLC, 2015).
36. Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S.-Y. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **45**, W365–W373 (2017).
37. Chakrabarty, B. & Parekh, N. Naps: network analysis of protein structures. *Nucleic Acids Res.* **44**, W375–W382 (2016).
38. Miozzo, M. et al. Insights on protein thermal stability: a graph representation of molecular interactions. *Bioinformatics* **35**, 2569–2577 (2018).
39. Desantis, F., Miozzo, M., Rienzo, L. D., Milanetti, E. & Ruocco, G. Investigating the side-chain structural organization behind the stability of protein folding and binding. Preprint at [arXiv:2107.08099](https://arxiv.org/abs/2107.08099) (2021).
40. Fiorentini, S. et al. First detection of SARS-CoV-2 spike protein n501 mutation in Italy in august, 2020. *Lancet Infect. Dis.* **21**, e147 (2021).
41. Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **73**, 4433–4448 (2016).
42. Miozzo, M. & Monacelli, L. Genome heterogeneity drives the evolution of species. *Phys. Rev. Res.* **2**, 043026 (2020).
43. De Maréno, A., Gueudré, T. & Miozzo, M. Exploration-exploitation tradeoffs dictate the optimal distributions of phenotypes for populations subject to fitness fluctuations. *Phys. Rev. E* **99**, 012417 (2019).
44. Spoel, D. V. D. et al. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
45. Brooks, B. R. et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
46. Jørgensen, W. L., Chandross-dhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
47. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
48. Parrinello, M. & Rahman, A. Crystal structure and pair potentials: a molecular-dynamics study. *Phys. Rev. Lett.* **45**, 1196–1199 (1980).
49. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
50. Chentham, T. E. L., Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. Molecular dynamics simulations on solvated biomolecular systems: the particle mesh ewald method leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* **117**, 4193–4194 (1995).
51. Yan, Y., Tao, H., He, J. & Huang, S.-Y. The HDOCK server for integrated protein-protein docking. *Nat. Protoc.* **15**, 1829–1852 (2020).
52. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
53. Caardi, G. & Nepusz, T. et al. The igraph software package for complex network research. *Int. J. Complex Syst.* **1695**, 1–9 (2006).
54. R Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2020).
55. Richards, F. M. Areas, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng.* **6**, 151–176 (1977).
56. Venkataraman, V., Suel, L. & Kihara, D. Potential for protein surface shape analysis using spherical harmonics and 3d zernike descriptors. *Cell Biochem. Biophys.* **54**, 23–32 (2009).
57. Novotni, M. & Klein, R. Shape retrieval using 3d zernike descriptors. *Comput. Aided Des.* **36**, 1047–1062 (2004).
58. Venkataraman, V., Yang, Y. D., Suel, L. & Kihara, D. Protein-protein docking using region-based 3d zernike descriptors. *BMC Bioinform.* **10**, 407 (2009).
59. Daberdaku, S. & Ferrari, C. Exploring the potential of 3d zernike descriptors and svm for protein-protein interface prediction. *BMC Bioinform.* **19**, 35 (2018).

Acknowledgements

The research leading to these results has been also supported by European Research Council Synergy grant ASTRA (n. 855923).

Author contributions

M.M. performed the in silico mutations, carried out the statistical analyses, and developed the numerical methods; M.M., L.D.R., L.B., and E.M. performed the molecular dynamics simulations and analyzed the data. G.P. and R.P. contributed additional ideas to the work and suggested biological tests with computational methods. A.B. and G.R. contributed with additional ideas and directed the computational choices on the basis of the biological and physical knowledge of the system; E.M. conceived the research. All authors wrote and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information


Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02946-w>.

Correspondence and requests for materials should be addressed to Edoardo Milanetti.

Peer Review Information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Karli Montague-Cardoso. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Appendix III

Other publications:

Spizzichino, S., Boi, D., Boumis, G., Lucchi, R., Liberati, F.R., Capelli, D., Montanari, R., Pochetti, G., **Piacentini, R.**, Parisi, G., Paone, A., Rinaldo, S., Contestabile, R., Tramonti, A., Paiardini, A., Giardina, G. and Cutruzzolà, F. (2022), *Cytosolic localization and in vitro assembly of human de novo thymidylate synthesis complex*. FEBS J, 289: 1625-1649. <https://doi.org/10.1111/febs.16248>

Affatigato, L., Sciortino, A., Sancataldo, G., Incocciati, A., **Piacentini, R.**, Bonamore, A., Cannas, M., Messina, F., Licciardi, M. and Militello, V. (2023), *Engineered Ferritin with Eu³⁺ as a Bright Nanovector: A Photoluminescence Study*. Photochem Photobiol, 99: 1218-1224. <https://doi.org/10.1111/php.13759>

Incocciati, A., Kubeš, J., **Piacentini, R.**, Cappelletti, C., Botta, S., Bertuccini, L., Šimůnek, T., Boffi, A., Macone, A., Bonamore, A. (2023), *Hydrophobicity-Enhanced Ferritin Nanoparticles for Efficient Encapsulation and Targeted Delivery of Hydrophobic Drugs to Tumor Cells*. Protein Science, e4819. <https://doi.org/10.1002/pro.4819>

Parisi, G.*, **Piacentini, R.***, Incocciati, A., Bonamore, A., Macone, A., Rupert, J., Zacco, E., Miotto, M., Milanetti, E., Tartaglia, G. G., Ruocco, G., Boffi, A., Di Rienzo, L., *Design of protein-binding peptides with controlled binding affinity: the case of SARS-CoV-2 receptor binding domain and ACE2-derived peptides* (SUBMITTED TO *Frontiers in Molecular Bioscience*, UNDER REVISION)

Kubiak, M., Krol, M., Gorczak, M., Krzemiński, Ł., Marszalek, I., Kurpiel, D., Bialasek, M., Gorka, E., Guzek, J., Skórzyński, M., Rygiel, T. P., Kutner, J., Kisiala, M., Wozniak, K., Parisi, G., **Piacentini, R.**, Boffi, A., & Kucharzewska, P., *Investigating the mechanisms of heavy-chain ferritin uptake by macrophages: Implications for drug delivery and immunotherapy* (SUBMITTED TO Nature Communications, UNDER REVISION)