

University of Naples Federico II

Polytechnic and Basic Sciences School



PhD Thesis in Chemical Sciences
XXXVIII COURSE

Studies of the interactions between target
proteins and drugs by *in silico*, *in vitro* and *ex*
vivo new generation approaches

Tutors: Prof. Maria Monti
Dr. Daniela Iaconis

Author: Irene Cipollone

Coordinator: Prof. Alessandra Napolitano

Table of contents

1. Study of interactions between small molecules and proteins -----	1
1.1 Introduction-----	1
1.2 Computational approaches-----	3
1.3 X-Ray crystallography-----	4
1.4 Spectroscopic methods -----	5
1.4.1. Nuclear magnetic resonance spectroscopy -----	5
1.4.2. Circular dichroism spectroscopy -----	5
1.4.3. Fluorescence resonance energy transfer -----	6
1.4.4. Surface Plasmon Resonance -----	6
1.4.5. Bio-layer interferometry (BLI) -----	7
1.5 Thermodynamic methods-----	7
1.5.1. Isothermal Titration Calorimetry (ITC)-----	7
1.5.2. Microscale thermophoresis -----	8
1.6 Affinity-based techniques-----	9
1.6.1. Photoaffinity labelling -----	9
1.6.2. Affinity-Based Protein Profiling (ABPP)-----	10
1.7 Label-free approaches -----	11
1.7.1. Thermal Proteome Profiling (TPP)-----	11
1.7.2. Cellular thermal shift assay (CETSA) -----	12
1.7.3. Limited Proteolysis coupled with Mass Spectrometry (LiP-MS) -	13
1.7.4. Stability of Proteins from Rates of Oxidation -----	14
1.8 Differential proteomics-----	15
1.9 Tasks of research project -----	16
1.10 Bibliography-----	18
2. Study of the interactions between small molecules and proteins in the context of SARS-CoV-2 -----	22
2.1 Introduction-----	22
2.1.1. SARS-CoV-2 genome and focus on 3CL ^{pro} -----	22
2.1.2. Inhibition of 3CL ^{pro} with covalent inhibitors and the focus on halicin	25
2.2 Results -----	29
2.2.1. Crystal structure of 3CL ^{pro} /halicin complex -----	29

2.2.2.	ESI-MS and mass mapping results -----	30
2.2.3.	Mutant and enzymatic assay results-----	35
2.2.4.	Molecular dynamics-----	37
2.2.5.	Native gel electrophoresis and DSF results-----	38
2.2.6.	Discussion-----	40
2.3	Conclusions-----	43
2.4	Materials and methods -----	44
2.4.1.	3CL ^{pro} expression, purification, and characterization -----	44
2.4.2.	Identification and relative quantification of 3CL ^{pro} modifications induced by halicin -----	44
2.4.3.	Crystallization, data collection, data reduction, structure determination, refinement and final model analysis-----	45
2.4.4.	In vitro enzymatic activity assays -----	46
2.4.5.	Differential scanning fluorimetry assay -----	46
2.4.6.	Native gel electrophoresis-----	46
2.5	Bibliography -----	47

SUMMARY

The study of small molecule–protein interactions is of central importance, as it not only provides insights into the binding mode but can also shed light on the true mechanism of action of pharmacological agents. Over the past decades, a variety of techniques, including proteomics, have been developed to expand our understanding of these processes. Within this framework, my PhD project, co-funded by Dompé Farmaceutici S.p.A. (D.M. 352), has been carried out. During the first year, I focused on the covalent inhibition of the SARS-CoV-2 main protease (3CL^{pro}) by halicin, within a collaboration network including Elettra Sincrotrone (Trieste), University of Cagliari and the EXSCALATE platform (Dompé). By applying biochemical and mass mapping strategies, in combination with X-Ray crystallography, Thermal Shift Assay and Molecular Dynamics approaches, halicin was found covalently targeting not only the catalytic Cys145 but also Cys44, a residue crucial for the stabilization of the dimeric active form of 3CL^{pro}. Through this finding, an inhibition process by halicin, relying also on its destabilizing effect upon Cys44 modification was proposed. In the last two years, the interest was turned to pulmonary fibrosis, one of the major long-term consequences of this viral infection, with a particular focus on c-Kit, a tyrosine kinase implicated in fibrotic progression. Using a Limited Proteolysis–Mass Spectrometry strategy, in combination with Surface Plasmon Resonance, X-Ray crystallography and molecular dynamics (in collaboration with EXSCALATE, Dompé), the binding mode of four different c-Kit inhibitors was investigated. Since c-Kit exists in both hyperphosphorylated (active) and hypophosphorylated (inactive) states, phosphorylation-induced conformational changes were first examined, identifying regions that become more or less exposed depending on the kinase activation state. Structural alterations upon drug treatment were then assessed, demonstrating that all inhibitors targeted the active site but with distinct preferences: one compound, Cpd2, bound selectively to the active form, two, Cpd3 and Cpd5, to

the inactive form, while a fourth, Cpd7, was able to interact with both conformations with a comparable affinity. The research on fibrosis progressed also through functional studies aimed at identifying the target pathways in fibrotic fibroblasts exposed preventively to three selected compounds. These experiments were carried out at the Plateforme Protéomique Necker in Paris in collaboration with Dr. Chiara Guerrera, employing both proteomic and phosphoproteomic analyses to characterize the compounds' mechanisms of action. The results highlighted diverse outcomes: one compound, Cpd79, completely prevented fibrotic progression, another, Cpd18, showed anti-fibrotic effects but was associated with side effects, while the third compound, Cpd108, failed to exert protective activity. In conclusion, this work provides an integrated perspective on small molecule–protein interactions, spanning from structural binding modes to functional cellular consequences. The findings emphasize the value of combining biochemical, biophysical, computational and proteomic approaches to elucidate mechanisms of action, ultimately contributing to the rational development of more effective and selective therapeutics.

1. Study of interactions between small molecules and proteins

1.1 INTRODUCTION

The study of the interactions between small molecules and proteins is fundamental for human health and for advancing our understanding of biological processes. Proteins are the workhorses of the cell, executing various functions ranging from catalysing biochemical reactions to transmitting signals within the body. Small molecules, encompassing drugs and metabolites, have the capacity to modulate protein activity, thereby exerting an influence on cellular processes and disease pathways¹. The fundamental principle of drug therapy for diseases is to bind to the relevant functional proteins, thereby improving, reversing and terminating the symptoms of the disease in a holistic manner by modulating the function of the proteins. Consequently, probing the interaction, affinity size, and binding mode of small molecule compounds with target proteins is an important means and way to reveal the mechanism of drug action².

The analysis of protein–small molecule interactions is of fundamental importance, given that the binding event constitutes the atomic unit of pharmacology. The binding process governs the duration of a drug's activity *in vivo* (through binding kinetics and residence time), thereby shaping efficacy, dosing, and duration of action^{3,4}. The mapping of interactions across the proteome facilitates a better understanding of the distinction between selectivity and polypharmacology. The majority of approved drugs engage multiple targets, a property that has the capacity to engender synergy for efficacy, to reveal opportunities for repurposing, or to provide a rationale for adverse effects^{5,6}.

Beyond therapy, well-validated chemical binders have been shown to function as probes, enabling the precise assignment of function to proteins and pathways. However, it is imperative that these binders meet stringent criteria for potency, selectivity, and on-target action to ensure their efficacy⁷.

Considerable technological advances have been made in various scientific domains, allowing the characterization of the binding mode between small molecules and proteins. Each of these techniques has its own set of advantages and disadvantages. However, the combination of several of them has the potential to elucidate the precise mechanism of interaction. A graphical representation of some techniques employed for the characterization of the binding mode between a small molecule and a protein is reported in Figure 1.1.

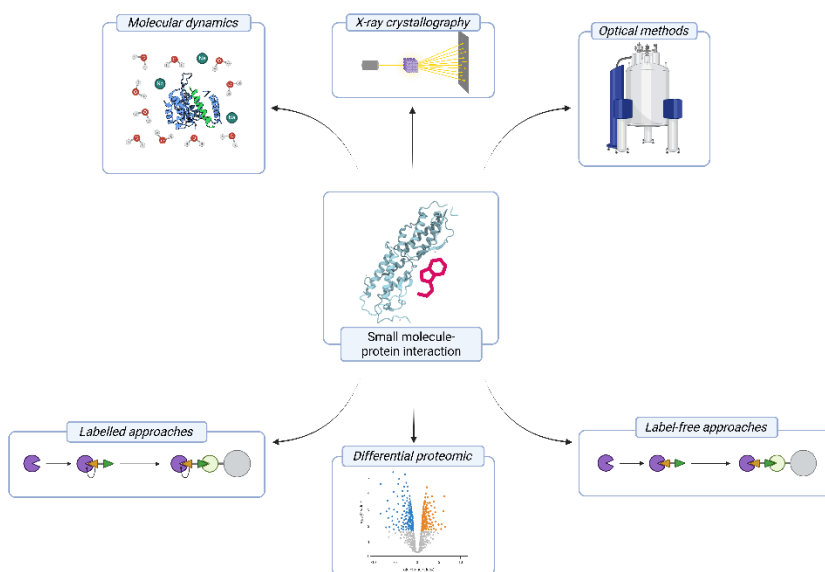


Figure 1.1. Graphical abstract of the main methodologies for the analysis of the interaction between small molecules and proteins

1.2 COMPUTATIONAL APPROACHES

Some of the technique for the study of the interaction between small molecule and proteins rely on computational predictions^{2,8}. This is primarily due to the advantages offered by high throughput, low cost, and short research cycle². Furthermore, computational methodologies are extensively employed in the domain of "drug repurposing", a process that facilitates the identification of novel applications for existing pharmaceutical agents, thereby reducing costs and accelerating the development timeline^{9,10}. In particular, at the inception of the drug discovery process, once the target protein has been selected, a screening procedure should be implemented for the purpose of identifying potential candidates. In this regard, structure-based Virtual Screening (VS) has emerged as a pivotal technique. The underlying principle of VS involves the docking of a library of small compounds into the binding pocket of a protein, followed by a subsequent theoretical calculation of the binding affinity⁸. This approach facilitates the identification of compounds that can interact more effectively with the target, paving the way for further validation through experimental approaches. However, it should be noted that computational approaches can also be used to confirm or better characterise a known binding between a small molecule and a protein. In this instance, the focus is on Molecular Simulations (MS), a method that utilises the Newtonian laws of motion to elucidate molecular systems, while preserving the physio-chemical properties of macrosystems¹¹. Two main classes of MS are identified: the Monte Carlo Simulation sampling method, in which an ensemble of conformations under specific thermodynamics conditions is generated through a stochastic approach¹², and the Molecular Dynamics simulation, in which a time-dependent Newtonian equation of motion for hard sphere particles in a system is iterated, providing different thermodynamic properties¹³.

CHAPTER 1

1.3 X-RAY CRYSTALLOGRAPHY

X-ray crystallography is a widely utilised technique for the characterisation of the binding site of a small molecule with a protein^{14,15}. X-Ray crystallography allows to obtain comprehensive structural information about the interaction of a ligand with a protein. Technically speaking, this methodology provides an electron density map, a time- and space- average of the electron density of molecules present in the crystallized protein¹⁴, allowing a complete characterization of the structure of the protein itself. Proteins can be crystallised by a variety of methods: for example, in the sitting-drop vapour-diffusion method, a drop of the protein is mixed with a reservoir solution containing salts or polymeric precipitant, which is located on an elevated platform, in a cell with 2-10 μL of liquid; by water equilibration through the vapour phase, the drop eventually approaches the reservoir in osmolarity, increasing both the concentration of precipitant in the drop and the protein concentration there¹⁶. Crystals of complexes can be prepared with two different protocols: soaking experiment, in which a single crystal of protein is exposed to a solution of the ligand for a specific period, and co-crystallization experiment, in which a solution mixture of protein and ligand is prepared and then the protein-ligand complex is crystallized¹⁵. After obtaining the protein–ligand complex crystals, they are mounted in loops and flash-cooled to approximately 100 K in order to mitigate radiation damage during X-ray exposure¹⁷. Diffraction data are collected usually from a synchrotron, with a rotation strategy that is meticulously chosen to maximise completeness and multiplicity. This process is closely monitored to ensure that the dose is kept to a minimum in order to limit any site-specific damage¹⁸. Subsequently, the crystallisation image is subjected to analysis using a range of software applications, with the objective of identifying the binding pocket^{19,20}.

1.4 SPECTROSCOPIC METHODS

In general, spectroscopic methods analyse how the matter can interact with a specific wavelength due to its intrinsic nature. In this context, the binding of a protein with a small molecule can be studied determining the changes in optical signals upon the binding of the molecule to the protein.

1.4.1. *Nuclear magnetic resonance spectroscopy*

Nuclear magnetic resonance is a technique based on the interaction of an externally applied radiofrequency radiation with atomic nuclei. The fundamental principle underlying NMR is that certain nuclei adopt specific nuclear spin states when exposed to an external magnetic field. NMR detects transitions between these spin states, which are unique to the specific nuclei under consideration and are influenced by their respective chemical environments¹⁹. In this particular field, it is possible to analyse binding by considering a phenomenon known as chemical shift perturbation, whereby resonance frequency differences are observed because of variations in the electronic environment surrounding the nucleus. When a protein interacts with a molecule, a change is evident in the electron cloud environment, resulting in a modification in the chemical shift²⁰. The observation of these changes thus facilitates the identification of the binding site of the small molecule to the protein. In this experiment, the binding is formed in a physiological buffer, the sample is placed in the probe of the NMR instrument, and the analysis is carried out. This process enables the observation of differences in chemical shifts, which in turn provides information on the occurred interactions^{21–23}.

1.4.2. *Circular dichroism spectroscopy*

Circular dichroism (CD) is defined as the difference in absorption of left and right circularly polarised light by a solution containing the molecules of interest²⁴. This effect arises when chiral molecules are analysed. Proteins are known to possess intrinsic chirality due to their secondary structural elements, and numerous ligands within biological systems are also chiral. This renders the CD

technique a valuable tool for investigating the interactions between small molecules and proteins²⁵. In the instance of a binding event, a structural chiral perturbation is initiated; the alterations in the CD signals of the isolated compound and protein, in comparison with those of the complexes, facilitate interaction analysis²⁶.

1.4.3. *Fluorescence resonance energy transfer*

Fluorescence resonance energy transfer (FRET) is a spectroscopic phenomenon in which a donor (the excited fluorophore) can transfer the energy of its de-excitation to an acceptor (a fluorophore) via long-range dipole-dipole coupling. This phenomenon occurs when the distance between the two fluorophores is minimal²⁷. The efficiency of the energy transfer, also known as FRET efficiency, is characterised by the Förster distance, denoted as R_0 . This is the distance between the fluorescent groups involved in the transfer of energy. When R_0 is between 2 and 8 nm, it is indicative of energy transfer and, consequently, the existence of interaction²⁸. In such cases, the selection of fluorescent dyes for labelling the protein and the molecule is often necessary. The most common labelling methods include chemical reactions, the use of antibodies or the production of the recombinant protein with a tag²⁹. The selection of the appropriate emission and excitation wavelength is also of significance, and this must be made according to the properties of the fluorescent molecule. The analysis of the fluorescent intensity, the FRET efficiency and the R_0 can then be calculated after the registration of the emission spectra of the targets³⁰.

1.4.4. *Surface Plasmon Resonance*

Surface Plasmon Resonance (SPR) is a technique that exploits the effect of the change in the refractive index of the surface of a metallic medium when there is a change in the mass on the surface³¹. In this experimental set-up, the protein is immobilized on a metal sensor chip, and the ligand solution is injected. When the ligand binds to the protein, the mass on the chip undergoes a change, resulting in

a concomitant alteration in the refractive index³². The signal intensity is directly proportional to the concentration of the ligand molecule. The resonance curves are examined firstly without the addition of the ligand, followed by the gradient addition of increasing concentrations of the molecule solution³³.

1.4.5. *Bio-layer interferometry (BLI)*

Bio-Layer Interferometry (BLI) is a technique based on the principle of real-time measurement of changes in optical interference signals by two surfaces, a proprietary reference surface and a sample surface. This technique can monitor the interaction between the ligand, which is immobilized onto the biosensor surface to form a biofilm, and the analyte which is kept in solution³⁴. The binding of the ligand molecule results in a change in the thickness of the biofilm and, consequently, a change in the spectrum of interference. The parameters of this interaction can be obtained by detecting the amount of interferometric spectral shifts before and after the binding³⁵ and the interferometric signal of the sensor chip without the addition of the small molecule is measured as baseline.

1.5 THERMODYNAMIC METHODS

In general, the application of thermodynamic methods involves the analysis of discrepancies in a range of calorimetric parameters. These parameters are indicative of the binding of the protein to the molecule in comparison to the isolated protein.

1.5.1. *Isothermal Titration Calorimetry (ITC)*

Isothermal Titration Calorimetry (ITC) is a technique based on the measurement of the heat generated or absorbed, calculated as the variation in enthalpy, during a specific reaction. In more detail, the binder is added to the solution of the protein over time using one or more individual injections, and the measured heat can be either as a change in temperature or as a change in power required to maintain temperature between the sample and a reference cell³⁶. During the analytical

process, the protein sample is introduced into the cuvette, and the baseline is subsequently measured. Thereafter, the small molecule sample is loaded, after which the heat change is recorded³⁷.

1.5.2. *Microscale thermophoresis*

Microscale thermophoresis is a thermophoretic method whereby the movement of molecules in solution is driven by the application of a temperature gradient. In the experiment, a fluorescent label is attached to the analysed protein and then both the protein and the small molecule are injected into a temperature gradient buffer generated using an infrared laser. The molecules will then undergo thermophoretic movement in accordance with the temperature gradient³⁸. Upon interaction of the protein with the molecule, the hydration shell of the molecule is known to be affected. This, in turn, results in a slowing down of the movement of the labelled molecule and alterations to the fluorescence distribution³⁹. In instances where the protein of interest is non-fluorescent, the implementation of the label process is imperative. This process entails the preparation of solutions at varying concentrations of the molecule under investigation. These solutions are subsequently injected into the thermophoresis instrument. The temperature gradient is then set, and the thermophoresis data are subjected to analysis^{40,41}.

Proteomic analyses also play a crucial role by enabling the systematic investigation of protein–small molecule interactions using diverse experimental workflows. Collectively, these strategies form the field of chemical proteomics, which primarily aims to identify the protein targets of bioactive compounds and increasingly focuses on pinpointing the specific binding regions⁴². Chemical proteomics techniques are broadly divided into labelled and label-free approaches according to the possibility to employ chemical tags to add to ligands (probes) in the protocols for protein targets identification. Affinity-based approaches

1.6 AFFINITY-BASED TECHNIQUES

In the affinity-based technique, also known as drug-pulldown⁴³, the target molecule (pharmacophore) is chemically modified with functional groups that enable its immobilization on solid supports, such as sepharose or magnetic beads. Subsequent to anchoring the labelled drug to the solid-phase matrix, the drug is incubated with cell lysates with a view to selectively capturing interacting target proteins. The bound proteins are then eluted, enriched, and subsequently identified using mass spectrometry (MS). This approach has been demonstrated to offer an effective and cost-efficient strategy for broad target analysis and rapid enrichment of protein interactors⁴⁴.

1.6.1. Photoaffinity labelling

In photoaffinity labelling, a chemical probe covalently binds to its target upon light activation. The design of such probes necessitates three essential components: a photoreactive group, a linker that connects this group to the small molecule of interest, and an affinity tag for detection or enrichment⁴⁵. Following exposure to irradiation, the photoreactive moiety is activated, thereby enabling the formation of a stable covalent bond with the interacting target. Upon activation, each of these groups produces a distinct reactive intermediate, which have a specific residue preference^{46,47}. A plethora of photo-reactive groups have been employed in this field, including diazo compounds⁴⁸, aryl azides⁴⁹, diazirines⁵⁰, aryl ketones⁴⁵, and tetrazoles⁵¹. The principal features of each group are set out in Table 1.1.

CHAPTER 1

Table 1.1. Major classes of phoreactive groups. The general structure, the reactive intermediate, the residue preferences and the activation wavelength are reported.

Photogroup	General structure	Reactive Intermediate	Residue Preferences	Activation wavelength
Diazo compounds ⁴⁸		Carbene	Leu, Val, Ile, Met	~350 nm
Aryl azides ⁴⁹		Nitrene (singlet/triplet)	Trp, Tyr, Phe + Lys, His, Arg	~254–300 nm
Diazirines ⁵⁰		Carbene (singlet/triplet equilibrium)	Asp/Glu + Leu, Val, Ile, Met	~330–370 nm
Aryl ketones ⁴⁵		Triplet carbonyl diradical	Met, Leu, Ile, Val, Ala; Trp, Tyr, Phe	~350–365 nm
Tetrazoles ⁵¹		Nitrile imine	Lys, His, Tyr	~300–365 nm

1.6.2. Affinity-Based Protein Profiling (ABPP)

Affinity-Based Protein Profiling (ABPP) is a powerful technique for identifying protein targets within complex biological samples (Figure 1.2). The activity-based probes under consideration are typically designed with three main components. Firstly, they contain a reactive group. Secondly, they contain a reporter group, which is often an alkyne. Thirdly, they contain a linker⁵² connecting the reactive group to the reporter. The reactive group is typically derived from known covalent enzyme inhibitors, thereby facilitating the formation of covalent adducts with potential protein targets. Reporter groups commonly include fluorescent dyes (e.g., cyanines) or biotin, which facilitate both the visualization of labelled proteins and their enrichment for downstream analysis. The composition of the linker may be a hydrophilic chain, a lipophilic chain, or a peptide. ABPs are incubated with proteins from cell lysates or up taken by living cells, allowing specific and

stable target labelling⁵³. To improve this approach, recent strategies have increasingly incorporated bioorthogonal chemistry, which relies on small, inert functional groups. The most commonly applied reactions include copper(I)-catalyzed azide–alkyne cycloaddition (CuAAC)⁵⁴, strain-promoted azide–alkyne cycloaddition (SPAAC)⁵⁵, and the inverse electron-demand Diels–Alder (iEDDA) reaction⁵⁶.

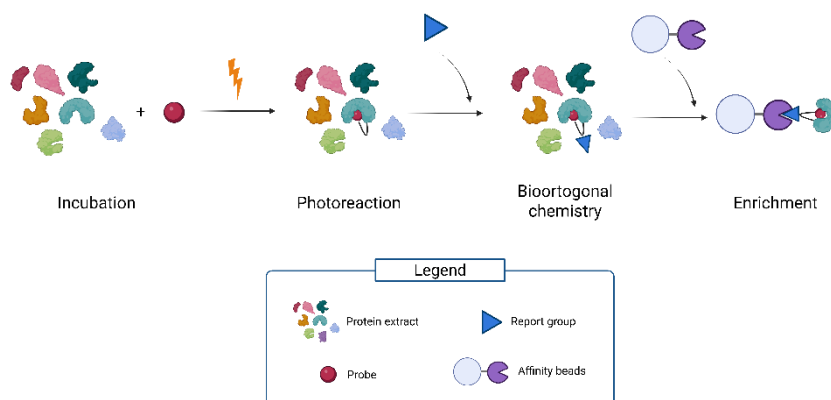


Figure 1.2. Schematic representation of the ABPP protocol.

1.7 LABEL-FREE APPROACHES

Label-free approaches employ small molecules in their native form, avoiding chemical modifications for immobilization on inert matrices and thereby preserving their functional properties.

1.7.1. Thermal Proteome Profiling (TPP)

Thermal Proteome Profiling (TPP) is a technique used to assess drug–target interactions by measuring changes in protein thermal stability and solubility following pharmacological treatment⁵⁷. In this method, cells are lysed under native conditions, and the resulting protein extracts are incubated with either the drug or vehicle control. For each condition, aliquots are subjected to a temperature gradient, after which soluble protein fractions are separated from precipitated proteins

by ultracentrifugation or multi-well filter plates. Proteins occurring in soluble fraction are then digested with trypsin and analysed by LC-MS/MS for protein identification and quantification, by using proteomics labelled⁵⁸ or label free approaches. The resulting data are processed using bioinformatic tools to generate sigmoidal melting curves and calculate melting temperatures for each protein in both treated and control samples. As pro-proteins exhibit alterations in their thermal stability upon binding small molecules, these shifts can be utilised for the identification and characterisation of the molecular targets of a given compound⁵⁹. Some more specific protocols have been developed starting from the classic idea of TPP, such as Temperature range TPP (TPP-TR)^{60,61}, TPP Compound Concentration Range (TPP-CCR)⁶⁰ and a combination of both within two-dimensional TPP (2D-TPP)⁶².

1.7.2. Cellular thermal shift assay (CETSA)

The Cellular Thermal Shift Assay (CETSA) relies on the principle of ligand-induced thermodynamic stabilization of protein targets. Unlike TPP, CETSA can be carried out directly in living cells, providing a major advantage. In a typical experiment, cells are treated with either the small molecule of interest or a vehicle control and then exposed to a temperature gradient. Following heating, the cells are lysed, and the stability of target proteins is assessed. This is typically performed using Western blotting, a technique that enables the determination of whether the target proteins undergo temperature-dependent denaturation. Furthermore, it can be used to ascertain whether the melting curves of the target proteins shift in the presence of the ligand. The use of comparative analysis of the thermal stability of proteins under treated and control conditions is enabled by CETSA, thus facilitating the identification of small molecule–protein interactions and estimation of binding affinity⁶³.

1.7.3. Limited Proteolysis coupled with Mass Spectrometry (LiP-MS)

Limited Proteolysis-Mass Spectrometry (LiP-MS) is a label-free proteomic strategy that combines controlled proteolysis with mass spectrometry to monitor conformational changes in proteins upon ligand binding, unfolding, aggregation, or allosteric modulation (Figure 1.3)^{64,65}. The method has been demonstrated to be applicable in two different ways. Firstly, it can be used to identify protein–ligand complexes in isolation within an *in vitro* setting. Secondly, it can be used to identify drug targets by comparing peptide profiles in samples that have been treated versus untreated. This enables the identification of drug targets by comparing peptide profiles in treated versus untreated samples. In a typical workflow, proteins are incubated with the compound or vehicle and then subjected to a brief digestion with a non-specific protease, usually proteinase K, under controlled conditions. The step of limited hydrolysis is succeeded by extensive digestion with trypsin, in accordance with the conventional bottom-up proteomics pipeline. Thereafter, peptides are identified and quantified by LC-MS/MS in a label-free manner⁶⁶. Increases in fully tryptic peptides or decreases in half-tryptic peptides have been demonstrated to be associated with conformational changes that are linked to ligand binding. In recent times, enhancements such as Data-Independent Acquisition (DIA) and statistical frameworks like MSstatsLiP have been employed to improve the throughput and robustness of LiP-MS analyses⁶⁴.

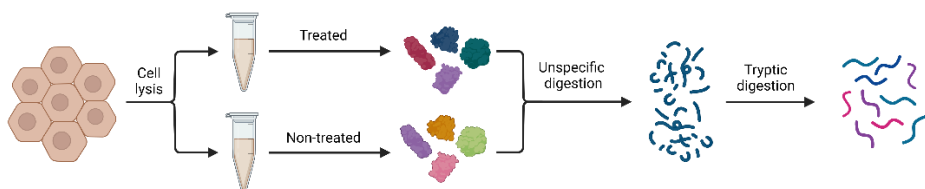


Figure 1.3. Schematic representation of LiP-MS approach

1.7.4. *Stability of Proteins from Rates of Oxidation*

The Stability of Proteins from Rates of Oxidation (SPROX) assay is based on the principle that protein stability increases upon small-molecule binding, thereby reducing the susceptibility of the protein to oxidation compared with its denatured state. In this approach, the compound of interest is incubated with a cell lysate, which is then subjected to chemical denaturation⁶⁷. After denaturation, all samples are exposed to an equal concentration of hydrogen peroxide, thereby inducing the oxidation of methionine residues that are accessible within the solvent⁶⁸. Since ligand binding alters the denaturation propensity of proteins, different concentrations of denaturant are required to unfold and expose methionine side chains. The resulting oxidized and non-oxidized peptides can be quantified using various analytical strategies, including chemical labelling approaches⁶⁹.

The most robust definition of the binding mode of a small molecule in native systems is achieved by integrating chemoproteomics with orthogonal readouts. For instance, the combination of photoaffinity chemoproteomics with CETSA has enabled the identification of the target and the subsequent in-cell stabilization of a nuclear receptor ligand. This progression signifies a shift from the initial engagement phase to the on-target mechanism⁷⁰. Conversely, an investigation employing affinity-based chemoproteomics in conjunction with isothermal cellular dose–response, docking, and functional assays identified ferrochelatase as an unanticipated pocket target of approved kinase inhibitors, thereby establishing a correlation between binding mode and a mechanistic toxicity phenotype⁷¹. The findings serve to underscore the significance of utilising complementary techniques for the better understanding of the binding mode between a small molecule and a target protein.

Following the characterisation of the binding mode, it is beneficial to comprehend the altered pathways within a living system (i.e. cells or *in vivo* models) subsequent to the administration of the molecule of interest. For this purpose, the

differential proteomics is a valid approach since it allows the identification of the dysregulated pathways in different conditions.

1.8 DIFFERENTIAL PROTEOMICS

Differential proteomics investigates proteome-wide changes under distinct conditions, such as physiological versus pathological states or drug-treated versus untreated samples. Its primary aim is to identify and quantify statistically significant variations in protein expression, thereby highlighting key proteins and pathways that may serve as potential therapeutic targets⁷². Differential strategies are broadly categorized into gel-based and shotgun approaches, with the latter being most widely adopted due to its reproducibility and streamlined sample preparation. Shotgun proteomics can be performed using label-based or label-free quantification. Label-based methods, such as SILAC and TMT, introduce isotopic labels into one or more experimental conditions, enabling the simultaneous analysis of all samples within a single LC-MS/MS run and improving the accuracy and reproducibility of quantitative comparisons⁷³. By contrast, label-free methods are cost-effective and broadly applicable since no labelling step is required, although they demand longer acquisition times and higher-performance mass spectrometers⁷⁴. In both strategies, quantification is based on the measurement and normalization of peptide ion currents, supported by software platforms that may be commercial (e.g., Proteome Discoverer, Spectronaut) or open source (e.g., MaxQuant, Proline, Mascot)⁷⁵. Relative protein abundances are then subjected to statistical testing (e.g., t-test, ANOVA) to identify significantly regulated proteins. Finally, bioinformatic analyses are performed to place these proteins within biological context, using tools such as Cytoscape⁷⁶, STRING⁷⁷, PANTHER⁷⁸, and Reactome⁷⁹, which enable the exploration of functional networks and pathway associations.

1.9 TASKS OF RESEARCH PROJECT

The present PhD project, co-funded by Dompé Farmaceutici S.p.A., concentrated on the characterisation of the binding mode of different molecules in two main contexts: the SARS-CoV-2 infection and pulmonary fibrosis. In the initial year of the study, I focused on the characterisation of the covalent binding of halicin, an inhibitor of the main protease of SARS-CoV-2, 3CL^{pro}. This task was carried out through mass spectrometry and biochemical based approaches for the identification of modified cysteine residues and complemented by X-ray crystallography for confirmation. Through native gel electrophoresis, thermal shift assay and molecular dynamics analysis of wt protein in comparison with the Cys44 mutant, an additional inhibition effect relying on dimer instability upon halicin treatment was proposed. A following task concerned the investigation of the binding modes of potential fibrotic inhibitors on a particular kinase of interest, c-Kit, which plays a pivotal role in the progression of pulmonary fibrosis. Multidisciplinary approaches including LiP-MS, x-ray crystallography, SPR and molecular dynamics data were employed to characterise the different binding sites of four compounds on active and inactive c-Kit forms. Finally, in the third task, proteomics and phosphoproteomics approaches were used to identify the dysregulated pathway following the induction of fibrosis, and to ascertain whether these pathways were restored upon treatment with specific drugs. The veracity of these findings was further substantiated by immunohistological analysis of fibroblasts. A graphical abstract of the approaches performed in the three years is reported in Figure 1.4.

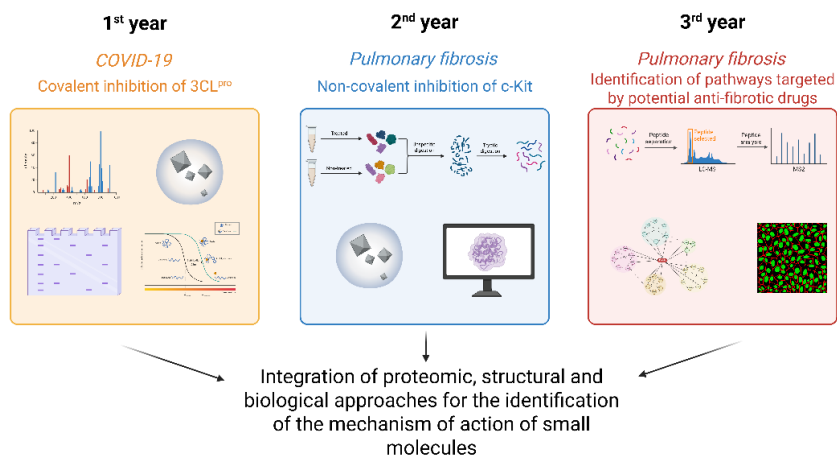


Figure 1.4. Graphical abstract of the main approaches employed during the PhD project. Created with BioRender.

1.10 BIBLIOGRAPHY

1. Lackner, A. *et al.* Measuring Interactions Between Proteins and Small Molecules or Nucleic Acids. *Curr Protoc* **4**, e1105 (2024).
2. Xu, H. *et al.* Detection of small molecule compounds bound to proteins. *Chinese Journal of Analytical Chemistry* **52**, 100398 (2024).
3. Knockenhauer, K. E. & Copeland, R. A. The importance of binding kinetics and drug-target residence time in pharmacology. *Br J Pharmacol* **181**, 4103–4116 (2024).
4. Copeland, R. A. The drug-target residence time model: a 10-year retrospective. *Nat Rev Drug Discov* **15**, 87–95 (2016).
5. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* **16**, 19–34 (2017).
6. Lounkine, E. *et al.* Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **486**, 361–367 (2012).
7. Arrowsmith, C. H. *et al.* The promise and peril of chemical probes. *Nat Chem Biol* **11**, 536–541 (2015).
8. Kontoyianni, M. Docking and Virtual Screening in Drug Discovery. *Methods Mol Biol* **1647**, 255–266 (2017).
9. Dotolo, S., Marabotti, A., Facchiano, A. & Tagliaferri, R. A review on drug repurposing applicable to COVID-19. *Brief Bioinform* **22**, 726–741 (2021).
10. Cavalcante, B. R. R. *et al.* In silico approaches for drug repurposing in oncology: a scoping review. *Front Pharmacol* **15**, 1400029 (2024).
11. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **9**, 646–652 (2002).
12. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**, 1087–1092 (1953).
13. Alder, B. J. & Wainwright, T. E. Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics* **27**, 1208–1209 (1957).
14. Maveyraud, L. & Mourey, L. Protein X-ray Crystallography and Drug Discovery. *Molecules* **25**, 1030 (2020).
15. Wienen-Schmidt, B., Oebbeke, M., Ngo, K., Heine, A. & Klebe, G. Two Methods, One Goal: Structural Differences between Cocrystallization and Crystal Soaking to Discover Ligand Binding Poses. *ChemMedChem* **16**, 292–300 (2021).
16. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun* **70**, 2–20 (2014).
17. Garman, E. F. & Owen, R. L. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **62**, 32–47 (2006).
18. Garman, E. F. & Weik, M. Radiation Damage in Macromolecular Crystallography. *Methods Mol Biol* **1607**, 467–489 (2017).
19. Marion, D. An introduction to biological NMR spectroscopy. *Mol Cell Proteomics* **12**, 3006–3025 (2013).
20. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **73**, 1–16 (2013).
21. Gee, C. T., Arntson, K. E., Koleski, E. J., Staebell, R. L. & Pomerantz, W. C. K. Dual Labeling of the CBP/p300 KIX Domain for 19 F NMR Leads to Identification of a New Small-Molecule Binding Site. *Chembiochem* **19**, 963–969 (2018).
22. Laveglia, V. *et al.* Automated Determination of Nuclear Magnetic Resonance Chemical Shift Perturbations in Ligand Screening Experiments: The PICASSO Web Server. *J Chem Inf Model* **61**, 5726–5733 (2021).

23. Arntson, K. E. & Pomerantz, W. C. K. Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery. *J Med Chem* **59**, 5158–5171 (2016).
24. Kelly, S. M. & Price, N. C. The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci* **1**, 349–384 (2000).
25. Rodger, A., Marrington, R., Roper, D. & Windsor, S. Circular Dichroism Spectroscopy for the Study of Protein-Ligand Interactions. in *Protein-Ligand Interactions* (ed. Ulrich Nienhaus, G.) vol. 305 343–363 (Humana Press, Totowa, NJ, 2005).
26. Daviter, T., Chmel, N. & Rodger, A. Circular and linear dichroism spectroscopy for the study of protein-ligand interactions. *Methods Mol Biol* **1008**, 211–241 (2013).
27. Selvin, P. R. [13] Fluorescence resonance energy transfer. in *Methods in Enzymology* vol. 246 300–334 (Elsevier, 1995).
28. Lu, Y. *et al.* Interaction of human serum albumin with N-(4-ethoxyphenyl)-N'-(4-antipyrinyl) thiourea using spectroscopies and molecular modeling method. *Journal of Luminescence* **129**, 734–740 (2009).
29. Shrestha, D., Jenei, A., Nagy, P., Vereb, G. & Szöllösi, J. Understanding FRET as a research tool for cellular studies. *Int J Mol Sci* **16**, 6718–6756 (2015).
30. Herman, B., Krishnan, R. V. & Centonze, V. E. Microscopic Analysis of Fluorescence Resonance Energy Transfer (FRET). in *Protein-Protein Interactions* vol. 261 351–370 (Humana Press, New Jersey, 2004).
31. Vo, T., Paul, A., Kumar, A., Boykin, D. W. & Wilson, W. D. Biosensor-surface plasmon resonance: A strategy to help establish a new generation RNA-specific small molecules. *Methods* **167**, 15–27 (2019).
32. Nguyen, H. H., Park, J., Kang, S. & Kim, M. Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel)* **15**, 10481–10510 (2015).
33. Huber, W. & Mueller, F. Biomolecular interaction analysis in drug discovery using surface plasmon resonance technology. *Curr Pharm Des* **12**, 3999–4021 (2006).
34. Petersen, R. L. Strategies Using Bio-Layer Interferometry Biosensor Technology for Vaccine Research and Development. *Biosensors (Basel)* **7**, 49 (2017).
35. Kumaraswamy, S. & Tobias, R. Label-free kinetic analysis of an antibody-antigen interaction using biolayer interferometry. *Methods Mol Biol* **1278**, 165–182 (2015).
36. Salim, N. N. & Feig, A. L. Isothermal titration calorimetry of RNA. *Methods* **47**, 198–205 (2009).
37. Paketurytė, V., Zubrienė, A., Ladbury, J. E. & Matulis, D. Intrinsic Thermodynamics of Protein-Ligand Binding by Isothermal Titration Calorimetry as Aid to Drug Design. in *Microcalorimetry of Biological Molecules* (ed. Ennifar, E.) vol. 1964 61–74 (Springer New York, New York, NY, 2019).
38. Seidel, S. A. I. *et al.* Label-free microscale thermophoresis discriminates sites and affinity of protein-ligand binding. *Angew Chem Int Ed Engl* **51**, 10656–10659 (2012).
39. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure* **1077**, 101–113 (2014).
40. Hellinen, L. *et al.* Microscale Thermophoresis as a Screening Tool to Predict Melanin Binding of Drugs. *Pharmaceutics* **12**, 554 (2020).
41. Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat Commun* **1**, 100 (2010).
42. Wright, M. H. & Sieber, S. A. Chemical proteomics approaches for identifying the cellular targets of natural products. *Nat Prod Rep* **33**, 681–708 (2016).
43. Ziegler, S., Pries, V., Hedberg, C. & Waldmann, H. Target Identification for Small Bioactive Molecules: Finding the Needle in the Haystack. *Angew Chem Int Ed* **52**, 2744–2792 (2013).
44. Rodriguez, E. L. *et al.* Affinity chromatography: A review of trends and developments over the past 50 years. *J Chromatogr B Analyt Technol Biomed Life Sci* **1157**, 122332 (2020).

45. Burton, N. R., Kim, P. & Backus, K. M. Photoaffinity labelling strategies for mapping the small molecule-protein interactome. *Org Biomol Chem* **19**, 7792–7809 (2021).
46. Chen, S. *et al.* Self-renewal of embryonic stem cells by a small molecule. *Proc Natl Acad Sci U S A* **103**, 17266–17271 (2006).
47. Zhang, Q. *et al.* Small-molecule synergist of the Wnt/beta-catenin signaling pathway. *Proc Natl Acad Sci U S A* **104**, 7444–7448 (2007).
48. Singh, A., Thornton, E. R. & Westheimer, F. H. The photolysis of diazoacetylchymotrypsin. *J Biol Chem* **237**, 3006–3008 (1962).
49. Murale, D. P., Hong, S. C., Haque, M. M. & Lee, J.-S. Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci* **15**, 14 (2016).
50. Smith, R. A. & Knowles, J. R. Letter: Aryldiazirines. Potential reagents for photolabeling of biological receptor sites. *J Am Chem Soc* **95**, 5072–5073 (1973).
51. Karaj, E. *et al.* Tunable Cysteine-Targeting Electrophilic Heteroaromatic Warheads Induce Ferroptosis. *J Med Chem* **65**, 11788–11817 (2022).
52. Pan, S., Zhang, H., Wang, C., Yao, S. C. L. & Yao, S. Q. Target identification of natural products and bioactive compounds using affinity-based probes. *Nat Prod Rep* **33**, 612–620 (2016).
53. Fang, H. *et al.* Recent advances in activity-based probes (ABPs) and affinity-based probes (AfBPs) for profiling of enzymes. *Chem Sci* **12**, 8288–8310 (2021).
54. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective ‘ligation’ of azides and terminal alkynes. *Angew Chem Int Ed Engl* **41**, 2596–2599 (2002).
55. Jewett, J. C., Sletten, E. M. & Bertozzi, C. R. Rapid Cu-Free Click Chemistry with Readily Synthesized Biarylazacyclooctynones. *J. Am. Chem. Soc.* **132**, 3688–3690 (2010).
56. Blackman, M. L., Royzen, M. & Fox, J. M. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J Am Chem Soc* **130**, 13518–13519 (2008).
57. Martinez Molina, D. *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84–87 (2013).
58. Iacobucci, I. *et al.* From the Discovery of Targets to Delivery Systems: How to Decipher and Improve the Metallodrugs’ Actions at a Molecular Level. *Pharmaceutics* **15**, 1997 (2023).
59. Mateus, A. *et al.* Thermal proteome profiling for interrogating protein interactions. *Mol Syst Biol* **16**, e9232 (2020).
60. Franken, H. *et al.* Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat Protoc* **10**, 1567–1593 (2015).
61. Zinn, N. *et al.* Improved Proteomics-Based Drug Mechanism-of-Action Studies Using 16-Plex Isobaric Mass Tags. *J Proteome Res* **20**, 1792–1801 (2021).
62. Tang, G. *et al.* Orthogonal Strategies for Profiling Potential Cellular Targets of Anandamide and Cannabidiol. *Chemistry* **29**, e202300682 (2023).
63. Morretta, E. *et al.* Discovering the Biological Target of 5-epi-Sinuleptolide Using a Combination of Proteomic Approaches. *Marine Drugs* **15**, 312 (2017).
64. Malinowska, L. *et al.* Proteome-wide structural changes measured with limited proteolysis-mass spectrometry: an advanced protocol for high-throughput applications. *Nat Protoc* **18**, 659–682 (2023).
65. Holfeld, A. *et al.* Limited Proteolysis-Mass Spectrometry to Identify Metabolite-Protein Interactions. *Methods Mol Biol* **2554**, 69–89 (2023).
66. Cappelletti, V. *et al.* Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **184**, 545–559.e22 (2021).
67. West, G. M., Tang, L. & Fitzgerald, M. C. Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass spectrometry-based strategy. *Anal Chem* **80**, 4175–4185 (2008).

68. West, G. M. *et al.* Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements. *Proc Natl Acad Sci U S A* **107**, 9078–9082 (2010).
69. Strickland, E. C. *et al.* Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation. *Nat Protoc* **8**, 148–161 (2013).
70. Seneviratne, U. *et al.* Photoaffinity Labeling and Quantitative Chemical Proteomics Identify LXR β as the Functional Target of Enhancers of Astrocytic apoE. *Cell Chem Biol* **28**, 148–157.e7 (2021).
71. Klaeger, S. *et al.* Chemical Proteomics Reveals Ferrochelatase as a Common Off-target of Kinase Inhibitors. *ACS Chem Biol* **11**, 1245–1254 (2016).
72. Rozanova, S. *et al.* Quantitative Mass Spectrometry-Based Proteomics: An Overview. *Methods Mol Biol* **2228**, 85–116 (2021).
73. Wang, Z., Liu, P.-K. & Li, L. A Tutorial Review of Labeling Methods in Mass Spectrometry-Based Quantitative Proteomics. *ACS Meas Sci Au* **4**, 315–337 (2024).
74. Meng, X., Liu, D. & Guan, Y. Advances in the application of label-free quantitative proteomics techniques in malignancy research. *Biomed Chromatogr* **37**, e5667 (2023).
75. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513–2526 (2014).
76. Cline, M. S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2**, 2366–2382 (2007).
77. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–452 (2015).
78. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* **31**, 8–22 (2022).
79. Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res* **52**, D672–D678 (2024).

2. Study of the interactions between small molecules and proteins in the context of SARS-CoV-2

2.1 INTRODUCTION

Human coronaviruses are a group of respiratory pathogens that range in severity from the common cold to more severe diseases such as pneumonia and bronchiolitis¹. The emergence of SARS-CoV-2 in December 2019 rapidly escalated into a global crisis, leading the World Health Organization to declare a pandemic on 11 March 2020. In the early stages, the basic reproduction number (R_0) of SARS-CoV-2 was estimated to be between approximately 2.4 and 3.4, indicating a substantially higher transmissibility than typical influenza strains, which generally have an R_0 between 1.3 and 1.7². Even from a clinical perspective, the severity of the SARS-CoV-2 virus was significantly more pronounced than that of influenza, with considerably protracted periods of hospitalisation. The median duration from admission to death was 25.5 days for those infected with the novel coronavirus, in comparison to 9.5 days for patients afflicted with seasonal influenza³. A further study indicated that individuals aged between 18 and 49 exhibited a 5.6-fold elevated probability of hospitalisation due to COVID-19 in comparison with influenza, thereby underscoring the heightened severity of the infection even among younger adults⁴. Considering the dramatic pandemic experienced worldwide, the study of the vital processes of SARS-CoV-2 and of possible therapeutic targets with a very broad spectrum and therefore applicable in cases of general infection by other coronaviruses has been and remains a primary objective of a large part of microbiological, molecular and biomedical research.

2.1.1. SARS-CoV-2 genome and focus on 3CL^{pro}

The genome of SARS-CoV-2, akin to other β -coronaviruses, consists of 14 functional open reading frames (ORFs) that encode for nonstructural proteins (NSPs), accessory proteins, and structural proteins (Figure 2.1)^{5,6}.

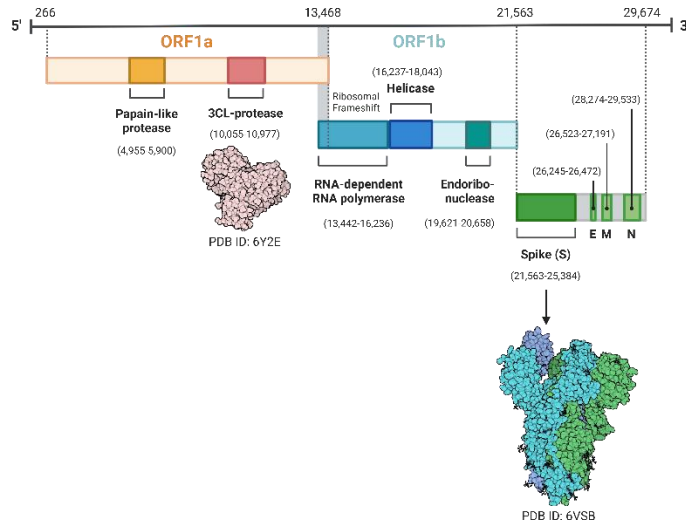


Figure 2.1. SARS-CoV-2 genome. (Created with BioRender)

ORF1a and ORF1b encode polyproteins pp1a and pp1ab, respectively. Together, these polyproteins account for approximately 16 nonstructural proteins (nsps) that assemble into the replication–transcription complex (RTC). The remaining 3' one-third of the genome encodes the four major structural proteins - Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N), as well as several accessory proteins that fine-tune virus–host interactions and contribute to immune evasion⁷. Proteolytic processing of pp1a and pp1ab is indispensable for viral replication and is mediated by two viral proteases embedded within the polyproteins themselves: the papain-like protease (PL^{pro}, nsp3) and the 3C-like cysteine protease (M^{pro} or 3CL^{pro}, nsp5)⁸. PL^{pro} cleaves the N-terminal region of pp1a/pp1ab at three sites, releasing nsp1, nsp2, and nsp3⁹. In addition to its role in viral protein maturation, PL^{pro} exhibits deubiquitinating and deISGylating activities, which suppress host antiviral pathways by interfering with ubiquitin- and ISG15-medi-

ated signalling cascades, particularly those regulating type I interferon responses¹⁰. This dual function makes PL^{pro} both a processing enzyme and an immune antagonist. 3CL^{pro}, a cysteine protease, hydrolyses the polyprotein at a minimum of 11 sites, preferring the Leu-Gln ↓(Ser, Ala, Gly) sequence (where ↓ marks the cleavage position). Through this activity, 3CL^{pro} releases most of the nsps, including the RNA-dependent RNA polymerase (RdRp, nsp12), helicase (nsp13), exonuclease (nsp14), and other enzymatic cofactors that together form the functional Replication-Transcription Complex (RTC)¹¹. Structurally, 3CL^{pro} is a homodimer. It is possible to distinguish three domains in each monomer. The N-terminal domain (residues 8-100) and the central domain (residues 101-183) are constituted of a series of six-stranded antiparallel β barrels, which envelop the substrate-binding site. The C-terminal domain (residues 198-303) is constituted of a cluster of five helices, the function of which is to regulate protein dimerization¹¹. The crystal structure of 3CL^{pro} is reported in Figure 2.2.

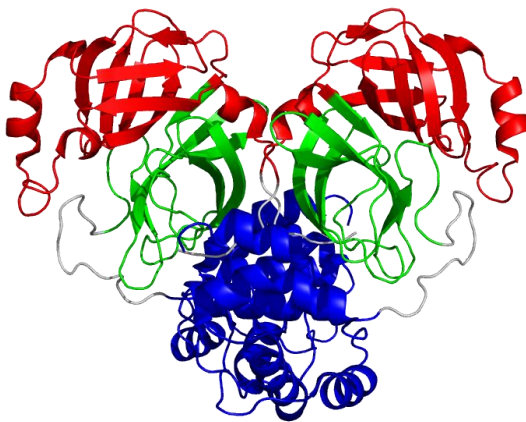


Figure 2.2. Crystal structure of 3CL^{pro} (PDB:7WOF). The N-terminus (in red), the central domain (in green) and the C-terminus (in blue) are highlighted.

Its catalytic machinery relies on a Cys145–His41 dyad, which performs a nucleophilic attack on peptide bonds during cleavage⁹. This protease plays a pivotal role in the viral life cycle, thus representing a primary target in the fight against the diffusion of SARS-CoV-2. Consequently, considerable efforts have been made to identify specific 3CL^{pro} inhibiting agents.

2.1.2. *Inhibition of 3CL^{pro} with covalent inhibitors and the focus on halicin*

By following repurposing strategies, a plethora of small molecules and drug candidates have been investigated as inhibitors of the SARS-CoV-2 main protease, 3CL^{pro}. The priority of them was focused on the possibility to covalently modify the catalytic cysteine residue (Cys145) thus to block protease activity. A notable example is Nirmatrelvir (developed by Pfizer as part of Paxlovid), which irreversibly targets the active cysteine in the catalytic site through a nitrile warhead and has been approved for clinical use¹². Further examples of covalent inhibitors encompass peptidomimetic aldehydes, such as GC376, α -ketoamide inhibitors designed to form reversible thiohemiketal adducts with Cys145, and compounds bearing Michael acceptors or epoxyketones⁹. All these inhibitors exploit the high nucleophilicity of the catalytic cysteine. However, it is important to note that the catalytic activity and structural stability of 3CL^{pro} are not solely dictated by Cys145. The enzyme functions as a homodimer, and the correct orientation of its subdomains is essential for maintaining the active conformation. This finding prompted Günther et collaborators to undertake an X-ray crystallography screen of approximately 5,000 compounds, which led to the identification of not only classical active-site binders but also two distinct allosteric binding sites¹³. Of the 37 molecules identified, the majority interacted covalently with Cys145 (e.g., tolperisone and GC376) or non-covalently with the active site (e.g., MUT056399), while 13 compounds bound to alternative pockets. The first allosteric site is located within the C-terminal dimerization domain and is composed primarily of

hydrophobic residues (Ile213, Leu253, Gln256, Val297, and Cys300). This region exhibits a high degree of conservation, with approximately 96% sequence identity shared between SARS-CoV and SARS-CoV-2. It has been demonstrated that mutations in key residues, such as Glu290, Arg298 and Glu299, result in the destabilisation of the dimer and consequent impairment of the enzyme's activity¹⁴. The second allosteric site is located between the catalytic and dimerization domains. In this region, AT7519 establishes van der Waals interactions with Val202, Thr292, Ile249, and Phe294, and hydrogen bonds with Gln110 and Asp153. These interactions result in the reorientation of Asp153 and the shift of Tyr154, thereby facilitating the formation of a novel salt bridge with Arg298¹³. The pivotal role of Arg298 was substantiated by Shi et colleagues. Their findings indicated that the R298A mutation resulted in the destabilisation of the S1 pocket and the compromise of dimer formation¹⁵. Computational studies showed that allosteric inhibitors may exert an indirect effect on the catalytic pocket. Zhang et al. utilised molecular dynamics simulations to report that compounds such as pelitinib and AT519 could induce conformational changes that reduce substrate accessibility¹⁶. Taken together, these observations emphasise that while covalent inhibitors targeting Cys145 have proven highly effective, allosteric inhibitors offer an alternative strategy by exploiting less conserved and potentially more druggable pockets. The combination of these approaches may result in the development of inhibitors that exhibit enhanced potency, resistance profiles, and therapeutic potential. In a previously work, a screening campaign on 3CL^{pro} protein activity involving more than 8,700 compounds was performed¹⁷. In addition to the previously reported inhibitors of 3CL^{pro} protease, 62 other compounds with IC₅₀ values lower than 1 μ M have been identified. In particular, the potential of halicin (also known as SU3327) to act as an inhibitor of the SARS-CoV-2 main protease was explored.

Halicin contains a 1,3,4-thiadiazole nucleus (Figure 2.3), a heterocyclic scaffold that is widely recognised in the field of medicinal chemistry due to its

presence in a variety of natural products and bioactive compounds. This nucleus is regarded as a pharmacologically privileged structure, forming the foundation for agents that exhibit antimicrobial, anti-inflammatory, anticancer, and central nervous system activities¹⁸.

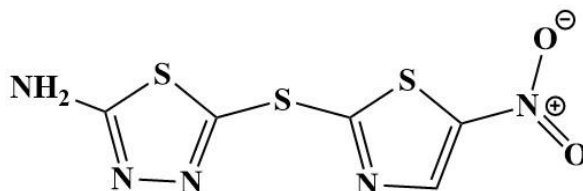


Figure 2.3. Structure of halicin

Initially, halicin was developed as an inhibitor of c-Jun N-terminal kinase (JNK), a member of the mitogen-activated protein kinase (MAPK) signalling pathway. The MAPK cascade is a key regulator of cellular responses to stress, and its dysregulation has been implicated in a wide range of pathologies, including type-2 diabetes¹⁹, obesity²⁰, tumourigenesis²¹, chronic inflammation²², and cardiovascular disease²³. In this context, halicin was demonstrated to suppress JNK activity and downstream inflammatory signalling, thus highlighting its potential as a therapeutic modulator of stress and immune responses. In recent years, halicin has attracted attention in a wholly distinct therapeutic domain, the antimicrobial resistance. A study by Stokes et collaborators utilised a deep-learning method to screen libraries of compounds for antimicrobial activity. This approach led to the unexpected identification of halicin as a potent broad-spectrum antibiotic²⁴. It exhibited bactericidal activity against multidrug-resistant strains, including *E. coli*²⁵, *Staphylococcus aureus*²⁶, *Acinetobacter baumannii*²⁷, *Enterococcus faecalis*²⁸, and *Enterococcus faecium*²⁹. Mechanistically, its antibacterial effect is linked not to traditional enzyme inhibition but to disruption of the bacterial transmembrane electrochemical gradient, rendering it less susceptible to pre-existing resistance mechanisms. In the context of SARS-CoV-2, the pharmacology of halicin is of

particular interest. The virus infection is frequently driven by a cytokine storm and inflammasome activation^{30–32} both of which involve JNK signalling, thus suggesting that JNK inhibitors, including halicin, could serve as host-directed therapies to mitigate hyperinflammation³³. Furthermore, Yang and colleagues proposed the repurposing of halicin as a covalent inhibitor of 3CL^{pro}³⁴. Utilising native mass spectrometry and X-ray crystallography, they were the firsts in publishing that halicin covalently modifies the catalytic cysteine of the enzyme. However, the incubation of 3CL^{pro} with halicin, even in limiting conditions, leads to the targeting of other cysteine residues, opening up the question about their potential role in protease activity and, consequently, their role in 3CL^{pro} inhibition when they are modified. Indeed, the modification of secondary cysteine residues may further influence protease stability, dimerization, or conformational dynamics, thus expanding the inhibitory mechanism beyond simple catalytic blockade.

Starting from these hypothesis and observations, in collaboration with Dompé Farmaceutici S.p.A., Elettra Sincrotrone Trieste and University of Cagliari we deeply characterised the effects of the other cysteines modified by halicin on 3CL^{pro} stability. The findings of this research have been published in a peer-reviewed journal as the following reference: Iacobucci I*, Cipollone I*, Cozzolino F, Iaconis D, Talarico C, Coppola G, Morasso S, Costanzi E, Malune P, Storici P, Tramontano E, Esposito F, Monti M. *Cys44 of SARS-CoV-2 3CL^{pro} affects its catalytic activity*. Int J Biol Macromol. 2025 Mar;295:139590. doi: 10.1016/j.ijbiomac.2025.139590. Epub 2025 Jan 7. PMID: 39788258. (*equal contribution).

2.2 RESULTS

2.2.1. *Crystal structure of 3CL^{pro}/halicin complex*

The co-crystals of 3CL^{pro} in conjunction with halicin was obtained by Dr. Storici and colleagues, and the X-ray structure was solved in space group P212121 at 1.93 Å resolution. The structure contained one dimer composed by two chains of 3CL^{pro}, chain A and chain B, in an asymmetric unit. The overall structure was preserved and exhibited a high degree of overlap with the 7TUU structure previously obtained by Yang et al. (RMSD 0.529 Å) through crystal soaking techniques³⁴. As illustrated in Figure 2.4, the 2Fo-FC map clearly identified multiple covalent derivatisations of solvent-accessible cysteines by halicin. In each site, the ligand has reacted, resulting in the release of the thiadiazol-amine and the consequent covalent binding of the nitrothiazole group to the sulphur. It is interesting to note that, in addition to the derivatisation of catalytic Cys145, they found halicin has also derivatised Cys44, Cys156 and Cys300. However, a divergent situation was observed: in chain A, all the residues of the active site were visible in the electronic densities. The cysteine with optimal electron density was Cys44(A). In contrast, residues 46-52 within chain B remained obscured by the electronic density, and histidine 41 underwent a rotational movement, thereby engaging in a stacking interaction with the ring of halicin bound to the catalytic C145(B). The present data corroborate the well-documented flexibility and adaptability of the 3CL^{pro} active site, as well as the observed asymmetrical behaviour of the subunits and, consequently, of the two active sites³⁵. Furthermore, the absence of electron density in the region 46-52 may indicate a destabilising effect resulting from halicin derivatisation.

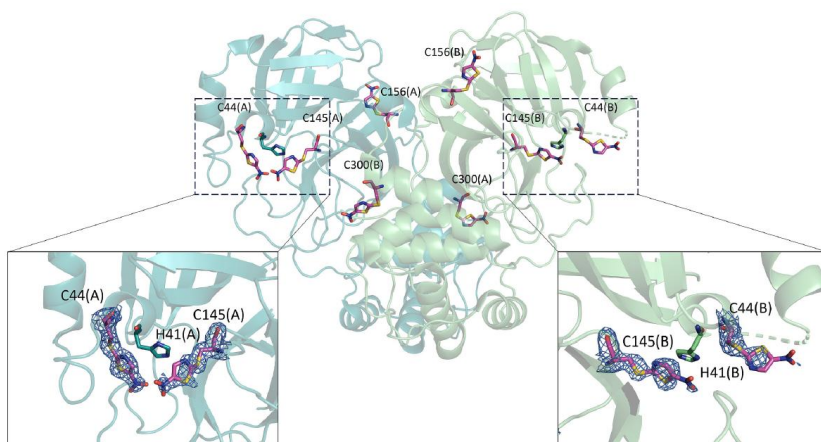


Figure 2.4. Cartoon model of the crystal structure of the dimer of 3CL^{pro} (PDB ID 7NBY, chain A = light blue, chain B = light green) with halicin bound to cysteines 44, 145, 156 and 300 of both chains shown as sticks and colored in magenta. Left insert: active site of chain A with derivatized Cys145 and Cys44 in magenta contoured by 2fo-fc map and catalytic His41 colored in teal. Right insert: active site of chain B with derivatized Cys145 and Cys44 in magenta contoured by 2fo-fc map and catalytic His41 is shown in sticks and colored in orange. 2Fo-Fc maps are shown in dashed blue lines contoured at 1 sigma.

2.2.2. ESI-MS and mass mapping results

In order to confirm whether other cysteine residues in addition to the catalytic Cys145 are targeted by halicin, thereby impairing the catalytic activity of 3CL^{pro}, the recombinant protein was treated with 10 μ M of halicin for 30 min. Subsequently, the modified protein mixture was analysed by LC-ESI-MS in order to confirm halicin binding and to characterise the resulting species. As shown in Figure 2.5 Panel A, the ESI-MS spectrum exhibited a species with a molecular weight of 33926.66 ± 0.93 Da, which is consistent with monomeric 3CL^{pro} modified by a single halicin molecule. This finding is in line with the results reported by Yang et al³⁴. An increase of +128 Da was observed on the expected average molecular weight of 3CL^{pro} (33796.6 Da), attributable to the covalent modification of a cysteine residue with one halicin molecule, through an aromatic nucleophilic substitution mechanism, as shown in Figure 2.5 Panel B.

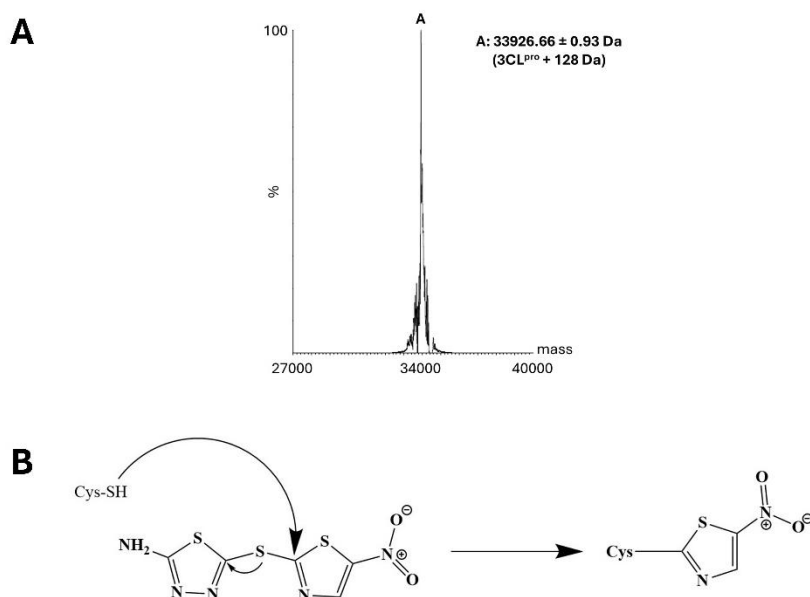


Figure 2.5. A) Transformed ESI spectrum of the reaction product of 3CL^{proM} with halicin in a 1:1 molar ratio. B) Aromatic nucleophilic substitution mechanism of a cysteine residue with one halicin molecule.

An aliquot of the reaction mixture was then subjected to a mass mapping procedure in order to localise the halicin-induced modification on the 3CL^{pro} sequence. The initial approach adopted was a classical digestion protocol, incorporating the carboamidomethylation of free cysteines, followed by trypsin digestion in 50 mM ammonium bicarbonate at pH 8. However, this approach yielded no peptide containing a modified cysteine, suggesting the intrinsic instability of the adduct in mild basic conditions. In order to avoid the dissociation of the covalent adduct with halicin and prevent scrambled disulfide bridge formation, it was necessary to employ a proteolytic protocol effective in acidic conditions. An aliquot of the modified protein was therefore digested with pepsin, a digestive proteolytic enzyme active at acidic pH (approximately pH 2). The resulting peptide mixture

was analysed by nanoLC-MS/MS. Pepsin's broad specificity renders it a particularly efficacious tool to achieve a high sequence coverage percentage of 3CL^{pro}. However, the MS/MS spectra of peptides resulting from pepsin digestion are frequently of substandard quality and challenging to interpret. As expected, Cys145 was identified as being modified within a variety of peptides that were generated during the process of pepsin digestion (Table 2.1). Surprisingly, Cys44, Cys85, and Cys300 were also found to be covalently bound to halicin. It was hypothesised that, within the limits of the experimental conditions explored (10 μ M halicin), a heterogeneous population of mono-modified 3CL^{pro}-halicin adducts would be generated, as indicated by the ESI-MS spectrum. In order to assess the contribution of each cysteine residue to the modified 3CL^{pro} fraction, the extent of modification for each cysteine-containing peptide was quantified. Quantification was performed using the ion extracted chromatogram (XIC) procedure. Specifically, the chromatographic peak areas of all peptide species containing a given modified cysteine were summed up and divided by the sum of the areas of all modified and unmodified peptides containing that residue. The amount of each modified cysteine was expressed as percentages and is shown in Table 2.1. Moreover, the MS/MS spectra of the Cys44-modified peptides 32-44 and 35-44 are also reported in Figure 2.6.

Table 2.1. This table consists of 4 sections, each reporting the list of peptides identified by LC-MS/MS and containing Cys44, Cys85, Cys145 and Cys300 modified or not. Each list contains information concerning starting and ending point of each peptide within its amino acid sequence and the modified cysteine, The calculated molecular weight (Da), the experimental m/z, the charge, the retention time (min) and the percentage of modification are also reported.

Start	Stop	Peptide sequence	Modified Cys	Calc. Mass (Da)	Exp. m/z	Area	% mod
32	47	LDDVVYCPRHVICTSE		1847.8546	924.9358	5.43E+06	
32	44	LDDVVYCPRHVIC		1530.7323	766.3719	1.34E+07	
32	44	LDDVVYCPRHVIC	SU3327 (C44)	1658.7004	830.3583	1.23E+06	
33	44	DDVVYCPRHVIC		1417.6483	709.8288	7.06E+06	
33	47	DDVVYCPRHVICTSE		1734.7706	868.3967	2.94E+06	
35	47	VVYCPRHVICTSE		1504.7167	753.3649	2.07E+08	
35	45	VVYCPRHVICT		1288.6420	645.3283	4.77E+07	
35	44	VVYCPRHVIC		1187.5944	594.8036	3.27E+08	
35	48	VVYCPRHVICTSED		1619.7436	810.8783	1.56E+07	
35	44	VVYCPRHVIC	SU3327 (C44)	1315.5624	439.5283	9.52E+07	
36	44	VYCPRHVIC		1088.5259	545.2704	2.17E+06	
75	87	LRVIGHSMQNCVL		1468.7643	735.3883	1.20E+07	7.02
75	89	LRVIGHSMQNCVLKL		1709.9433	570.9885	2.96E+06	
76	87	RVIGHSMQNCVL		1355.6802	678.847	1.07E+09	
76	89	RVIGHSMQNCVLKL		1596.8592	799.4386	7.30E+07	
76	103	RVIGH-SMQNCVLKLKVDTPKTPKY KF		3214.7471	804.6934	1.98E+06	
76	85	RVIGHSMQNC	SU3327 (C85)	1271.4958	636.7543	8.27E+06	
76	89	RVIGHSMQNCVLKL	SU3327 (C85)	1724.8273	575.9495	7.60E+07	
76	86	RVIGHSMQNCV		1242.5961	622.3037	9.26E+06	
76	87	RVIGHSMQNCVL	SU3327 (C85)	1483.6483	742.8301	3.74E+06	

CHAPTER 2

129	150	AMRPNFTIKGSFLNGSCGSVGF		2289.1035	764.0403	2.83E+06	46.45
135	150	TIKGSFLNGSCGSVGF	SU3327 (C145)	1700.7287	851.3719	3.43E+08	
135	151	TIKGSFLNGSCGSVGFN		1686.8036	844.4108	9.11E+06	
135	150	TIKGSFLNGSCGSVGF		1572.7607	787.3878	3.70E+08	
135	154	TIKGSFLNGSCGSVGFNIDY		2077.9779	1039.998	1.27E+06	
135	153	TIKGSFLNGSCGSVGFNID		1914.9146	958.4668	1.96E+06	
135	147	TIKGSFLNGSCGS		1269.6023	635.8085	6.16E+05	
135	146	TIKGSFLNGSCG	SU3327 (C145)	1310.5384	656.2756	1.97E+06	
135	149	TIKGSFLNGSCGSVG		1425.6922	713.8543	2.63E+06	
136	150	IKGSFLNGSCGSVGF		1471.7130	736.8612	2.50E+05	
139	150	SFLNGSCGSVGF		1173.5125	587.762	9.18E+05	
141	150	LNGSCGSVGF		939.4120	470.7135	7.94E+06	
289	306	DEFTPFDVVRQCSGVTFQ		2073.9466	1037.981	1.13E+06	4.85
289	302	DEFTPFDVVRQCSG		1598.7035	800.3607	8.69E+05	
290	306	EFTPFDVVRQCSGVTFQ		1958.9197	980.4676	1.24E+06	
290	302	EFTPFDVVRQCSG		1483.6766	742.847	1.31E+06	
291	306	FTPFDVVRQCSGVTFQ		1829.8771	915.9473	1.52E+07	
291	306	FTPFDVVRQCSGVTFQ	SU3327 (C300)	1957.8452	979.9348	3.84E+05	
292	306	TPFDVVRQCSGVTFQ		1682.8087	842.4123	6.72E+06	
295	306	DVVRQCSGVTFQ		1337.6398	669.8279	9.50E+06	
296	306	VVRQCSGVTFQ		1222.6129	612.3138	1.12E+09	
296	306	VVRQCSGVTFQ	SU3327 (C300)	1350.5809	676.2979	5.87E+07	
297	306	VRQCSGVTFQ		1123.5444	562.7784	2.50E+06	

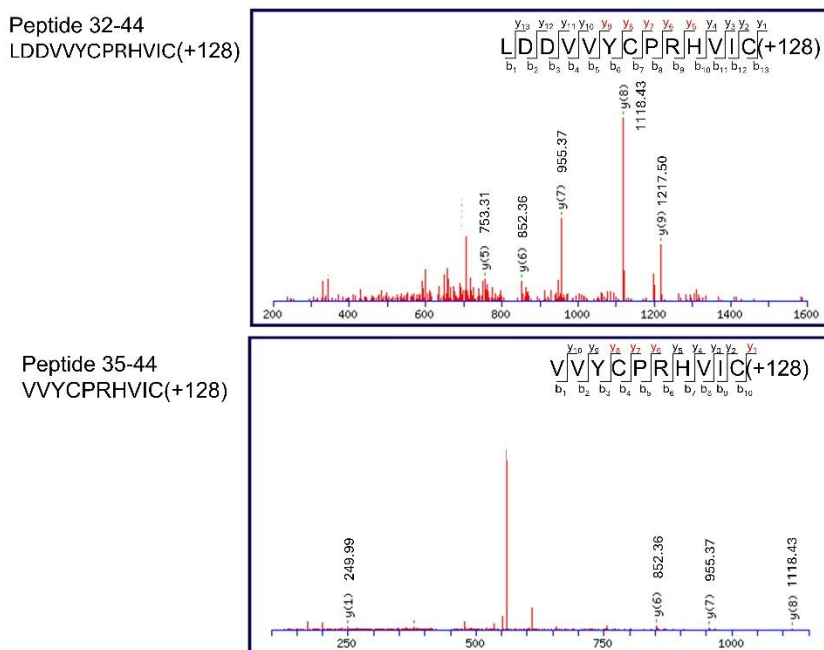


Figure 2.6. MS/MS spectra of the peptides modified at Cys44 32-44 and 35-44.

As expected, Cys145 was the most targeted by the drug, reaching approximately 47% of modification. The second cysteine residue with the highest percentage of modification was Cys44 (12.31%), while Cys85 and Cys300 appeared less modified (7.02% and 4.85%, respectively). These results highlight an unexpected reactivity of these other Cys residues in addition to the catalytic one, suggesting that a combination of mechanisms may be responsible for 3CL^{pro} inhibition, beyond the blockage of Cys145.

2.2.3. Mutant and enzymatic assay results

While the function of Cys145 is well established and this residue is essential for enzyme activity, the role of Cys44 has received comparatively less attention. In light of this and considering the high reactivity and exposure level of Cys44 suggested by the relevant targeting with halicin, the role of this residue on protein

stability and its enzymatic activity was more deeply analysed. To this aim, a mutant of the protein containing an alanine residue at position 44, rather than cysteine, was produced. The inhibitory activity of the halicin compound was evaluated in relation to its 3CL^{pro} WT and Cys44Ala mutant counterparts. The GC376 compound, a well-validated covalent inhibitor that forms a hemithioacetal adduct exclusively with Cys145³⁶, was utilised as the positive control in these experiments. As reported in Table 2.2, halicin exerted an effect on the 3CL^{pro} WT enzymatic activity, exhibiting an IC₅₀ value of 0.45 μ M. Furthermore, the inhibition was observed to occur in a dose-dependent manner (Figure 2.7, Panel A), with a profile that decreased in the presence of increasing concentrations of the compound, similar to the control GC376 (Figure 2.7, Panel B). Therefore, a biochemical assay was conducted to ascertain whether the mutation Cys44Ala exerted an effect on the enzymatic activity.

The IC₅₀ value recorded on the 3CL^{pro} Cys44Ala mutant was 6.52 μ M, about 14-fold lower if compared with 3CL^{pro} WT (IC₅₀=0.45 μ M). This data demonstrated that the presence of a cysteine residue in position 44 has an additional inhibiting effect on 3CL^{pro} catalytic activity, upon halicin targeting.

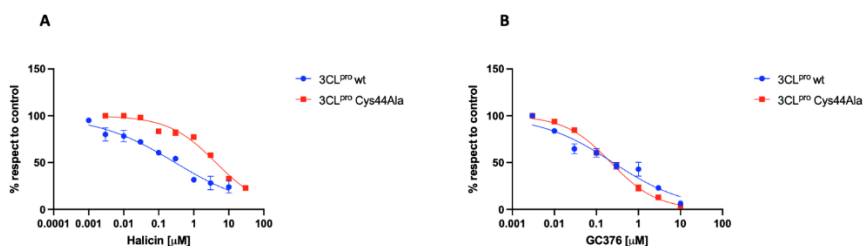


Figure 2.7. (A) Dose-dependent curve of halicin in the presence of 3CL^{pro} WT and 3CL^{pro} Cys44Ala mutant. (B) Dose-dependent curve of GC376 in the presence of 3CL^{pro} WT and 3CL^{pro} Cys44Ala mutant.

Table 2.2. Effect of halicin compound on 3CL^{pro} WT and Cys44Ala mutant.

Compound	SARS-CoV-2 3CL ^{pro} ^a IC ₅₀ (μM)	
	WT	Cys44Ala
Halicin	0.45 ± 0.04	6.52 ± 1.02
GC376	0.24 ± 0.03	0.24 ± 0.07

2.2.4. Molecular dynamics

The molecular dynamics (MD) approach was employed to elucidate the effects of the Cys44Ala mutation on the 3CL^{pro} monomer structure, with consideration given to the dynamics of the proteins. The values of Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) were evaluated by collaborating with EXSCALATE group of-Dompè S.p.A. collaborators, Dr. Iaconis Daniela and Dr. Talarico Carmine, in order to get information on differences occurring between 3CL^{pro} WT and the Cys44Ala mutant during the simulation. RMSD values (Figure 2.8, Panel A, green line) demonstrated that the WT monomer exhibited excellent global stability (RMSD range 1.5 to 3.5 Å), whilst the mutated shape (Figure 2.8, Panel A, blue line) underwent significant conformational alterations (RMSD range 1.0 to 5.5 Å), particularly in the final phase of the simulation (after TIME > 300 ns). Alterations in the range of 1-3 Å were expected for small, globular proteins. However, more significant changes, such as those observed in the case of mutant, suggest that the protein is experiencing a substantial conformational shift during the simulation.

As indicated by Root Mean Square Fluctuation (RMSF) data, residues that fall into the region around Cys44 (Figure 2.8, Panel B, green line) exhibited a high degree of mobility. This phenomenon may provide a rationale for the apparent ease of functionalisation experienced by Cys44. In the presence of the mutation Ala44, the same region remained exposed and appeared stable (Figure 2.8, Panel B, blue line). It is evident from the combined results of these simulations that the Cys44 residue is accessible to the ligand, and that the mutation at this position

exerts a substantial influence on the functional globular conformation of the monomer.

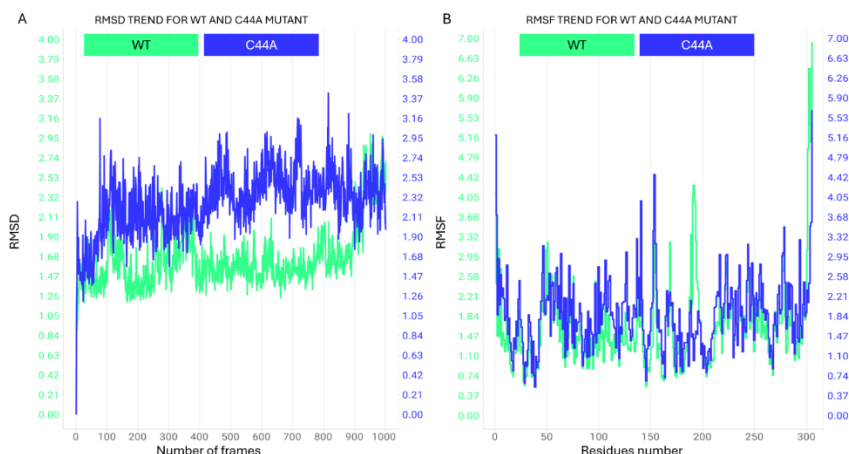


Figure 2.8. (A) Plot showing the RMSD trend in both WT (green) and Cys44Ala mutant monomer (blue) of 3CL^{pro}. (B) RMSF trend in both WT (green) and Cys44Ala mutant monomer (blue) of 3CL^{pro}.

2.2.5. Native gel electrophoresis and DSF results

In the light of previous findings concerning the possible effects of Cys44 on protein stability and, consequently, also on protease catalytic activity, the potential impact of modification of Cys44 on the monomer-dimer equilibrium was then investigated, by using native polyacrylamide gel electrophoresis. This method, employing sodium lauroyl sarcosinate (sarkosyl, SAR) at a concentration of 0.05% w/v, has been demonstrated to effectively resolve oligomeric states in the gel³⁷. As illustrated in Figure 2.9, the recombinant proteins 3CL^{pro} WT and the Cys44Ala mutant exhibited divergent distributions between monomeric and dimeric states. The WT protein displayed two bands with apparent molecular weights around 65kDa and 37kDa, corresponding to those expected for the dimeric and monomeric forms, respectively. Conversely, the Cys44Ala mutant exhibited an absence of the dimer band, suggesting that this mutation exerts a direct influence on the stability of the active dimeric form of 3CL^{pro}.

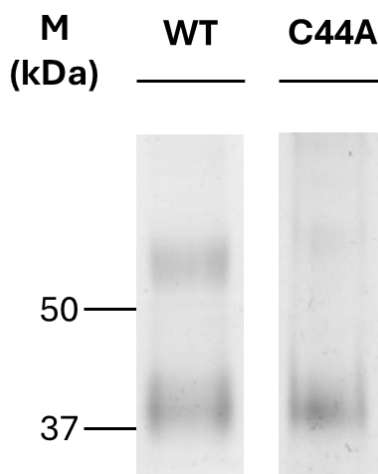


Figure 2.9. Sodium lauroyl sarcosinate (sarkosyl, SAR) 0.05% gel of 3CL^{pro} WT and Cys44Ala proteins. The image shows the conformers' resolution of WT and mutant enzyme (5 μ g each). Protein molecular weight markers are shown on the left side of the gel image.

To further validate the shifting between the dimeric and monomeric 3CL^{pro}, a Differential Scanning Fluorimetry (DSF) assay was carried out. A comparison of the profiles of wild-type (WT) and mutant proteins (Figure 2.10) revealed a significant decrease in the melting temperature of the mutant, estimated at -9°C in comparison to the WT. This finding indicated that, in the absence of drug intervention, the Cys44Ala mutant exhibits reduced stability in comparison to the WT 3CL^{pro}.

In the presence of halicin, the thermal profiles of both 3CL^{pro} WT and Cys44Ala mutant changed, with a different impact on the two proteins. As illustrated in Figure 2.10C, halicin induced a negative thermal shift of -6°C in 3CL^{pro} WT in comparison to the untreated protein. This behavior is distinct from that which was recorded in the presence of GC376 (Figure 2.10, Panel B), the other covalent inhibitor of the 3CL^{pro} WT which was used as positive control. This latter had no effects on the WT protein stability. Conversely, when the 3CL^{pro} Cys44Ala mutant was treated with halicin, the DSF profile demonstrated a positive shift

(Figure 2.10, Panel F) that is comparable to that recorded in the presence of GC376 (Figure 2.10, Panel E). These latter findings suggest that in the absence of Cys44, the apparent effect induced by halicin treatment aligns with that of GC376.

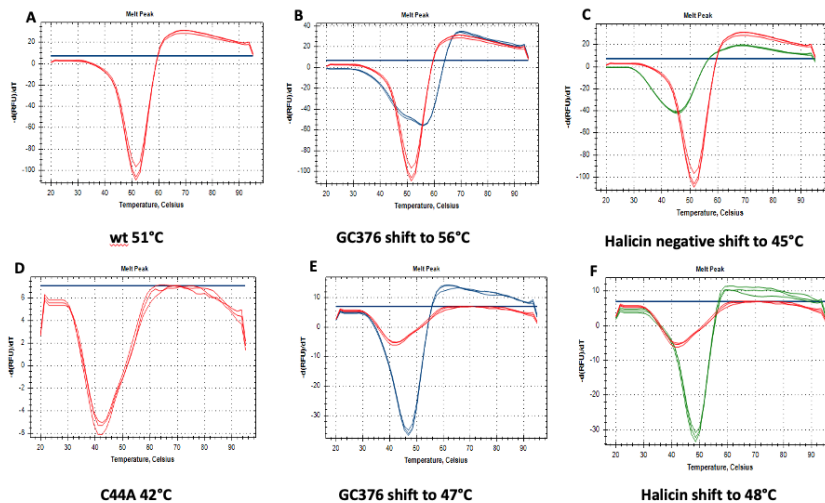


Figure 2.10. Differential scanning fluorimetry. A) 3CL^{pro} WT (in red). B) 3CL^{pro} WT (in red) in the presence of GC376 (in blue). C) 3CL^{pro} WT (in red) in the presence of halicin (in green). D) 3CL^{pro} Cys44Ala mutant (in red). E) 3CL^{pro} Cys44Ala mutant (in red) in the presence of GC376 (in blue). F) 3CL^{pro} Cys44Ala (in red) in the presence of halicin (in green).

2.2.6. Discussion

The main protease of SARS-CoV-2 (3CL^{pro}) has rapidly emerged as one of the most extensively investigated antiviral targets, owing to its essential role in viral replication. Its inhibition has been demonstrated to impede the processing of viral polyproteins into functional nonstructural proteins, thereby effectively arresting the viral life cycle^{9,38}. Conventionally, research has been concentrated on the catalytic active residues, with a particular emphasis on Cys145, which is deemed to be indispensable for enzymatic activity. However, the 3CL^{pro} structure contains 11 additional cysteine residues, many of which are highly conserved and

may influence folding, stability, or oligomerization⁹. The exploration of these residues presents a valuable opportunity to expand the spectrum of inhibitory mechanisms and to identify unconventional vulnerabilities in the protease.

In this study, the current role of 3CL^{pro} cysteine not catalytic residues was expanded by demonstrating that inhibition of 3CL^{pro} is not exclusively mediated by blockade of Cys145. Halicin, which was originally proposed as a Cys145-targeting inhibitor³⁴, has the capacity to modify other cysteine residues, most notably Cys44, alongside Cys85 and Cys300. The functional importance of Cys44 was validated through a combination of structural and biochemical means. The mutation of this residue to alanine resulted in a significant reduction in catalytic efficiency, compromised stability, and hindered dimer formation. This provides compelling evidence that Cys44 plays a previously underappreciated stabilizing role, directly supporting the structural integrity of the active protease. The high nucleophilicity of Cys44 at physiological pH renders it a particularly attractive site for covalent inhibition³⁹. The findings from this study provide compelling evidence that halicin interacts with Cys44, thereby impacting the conformational dynamics of 3CL^{pro}. The 14-fold reduction in potency observed with the Cys44Ala mutant lends further support to the hypothesis that effective inhibition of 3CL^{pro} can arise not only from catalytic blockade, but also from disruption of auxiliary stabilizing residues. It is important to note that molecular dynamics simulations revealed that the Cys44Ala mutation induces long-range structural perturbations, which were corroborated by differential scanning fluorimetry. These findings emphasize the broader principle that cysteine residues outside the catalytic site can modulate protease activity through effects on protein folding and oligomerization state⁹. This concept is further reinforced by the findings of previous studies. For instance, Cys300, located at the dimerization interface, has been shown to regulate quaternary structure, and its covalent modification has been demonstrated to destabilize

the functional dimer⁴⁰. Consequently, the present data on Cys44 extends a growing body of evidence that structural cysteines act as molecular switches, influencing dimerization, stability, and catalytic readiness.

The present study also highlights the context-dependent effects of covalent inhibition. While halicin has been shown to induce significant conformational destabilization in the wild-type enzyme, the Cys44Ala mutant incubated with the compound has been observed to display a melting-temperature (T_m) profile that is more closely similar with that of Cys44Ala mutant incubated with GC376, a Cys145-targeting covalent inhibitor. This finding suggests the possibility of a more classical, active-site-focused mechanism in the absence of Cys44⁴¹. In accordance with this interpretation, GC376 is a well-validated covalent inhibitor that forms a hemithioacetal adduct with Cys145^{36,42}, and it produces modest T_m increases relative to nirmatrelvir in thermal-shift assays⁴¹. These findings emphasize the potential of multi-cysteine engagement to enhance potency and robustness.

This work finally highlights the importance of integrating orthogonal approaches – namely, mass spectrometry, crystallography, biochemical assays, and computational modelling – for the mechanistic analysis of drug–protein interactions. The integration of such a workflow is becoming increasingly standardized in contemporary drug discovery pipelines, and the findings presented herein illustrate its capacity to unveil unexpected mechanisms of inhibition.

2.3 CONCLUSIONS

In this second chapter, different approaches for the characterization of the inhibition of halicin against 3CL^{pro} and for the understanding of the mechanism of action of Spike protein during the infection were reported. All these results create a solid base for our better understanding of the virus and how to inhibit it.

The work on 3CL^{pro} proves that Cys44 is an important cysteine in the SARS-CoV-2 main protease, and that changing it has a measurable effect on how well it works. By combining biochemical tests with targeted modifications, the study shows that Cys44 changes protease activity, suggesting that distant cysteines and the His41–Cys145 catalytic pair can communicate. These findings show that there are more places on 3CL^{pro} that can be targeted by drugs, beyond the active site and dimer interface, and suggest Cys44 as a useful control point that can be used to adjust catalysis. Halicin is a drug that can influence this residue, which suggests that it has the potential to be used as a treatment. Overall, understanding how Cys44 is regulated helps us to understand how 3CL^{pro} works better. This also creates a new way to design antivirals, making covalent or reversible binders that target Cys44 to reduce the effects of proteolysis and, in turn, viral replication.

2.4 MATERIALS AND METHODS

2.4.1. *3CL^{pro} expression, purification, and characterization*

The WT 3CL^{pro} of SARS-CoV-2 plasmid was kindly provided by Prof. Rolf Hilgenfeld's research group at the Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck (Germany) (ORF1ab poly-protein residues 3264-3569, GenBank code: MN908947.3). Protein production and purification followed the protocol reported by Zhang et al. 2020, with eluted fractions containing the target protein pooled and subjected to buffer exchange in 20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 7.8^{43,44}. The protein was flash frozen in LN2 at a concentration of 10-20 mg/ml and stored in aliquots at -80°C. The 3CL^{pro} WT and Cys44Ala mutant proteins for biochemical and differential scanning fluorimetry assays were expressed in the *E. coli* BL21(DE3) strain and purified using the same protocol described in Messori et al⁴⁵. Both recombinant proteins were characterized by mass mapping strategies involving *in situ* hydrolysis of the band fractionated by SDS-PAGE and LC-MS/MS analysis, according to Cozzolino et al⁴⁶.

2.4.2. *Identification and relative quantification of 3CL^{pro} modifications induced by halicin*

The reaction of 3CL^{pro} with halicin was performed by incubating the native 3CL^{pro} with halicin in a 1:1 molar ratio with the monomeric form of the protein for 30 minutes at room temperature (buffer: 20mM Tris, pH 7.8, 150 mM NaCl, 1 mM EDTA). An aliquot of the sample was analyzed by LC-MS a Q-ToF Premier (Waters, Milliford, MA, USA) mass spectrometer coupled with an 1100 Agilent HPLC (Palo Alto, California, US)⁴⁷ to verify the reaction and determine the number of molecules that had reacted with each 3CL^{pro}. The mass mapping procedure for identifying modified residues involved setting up a pepsin hydrolysis protocol. The optimal hydrolysis conditions consisted of a 1:50 w/w enzyme/substrate ratio

for 2 hours⁴⁸. The peptide mixtures were then analyzed by nano LC-MS/MS using an LTQ Orbitrap XL mass spectrometer according to previously reported protocols⁴⁶. Peptide identification was carried out with the Mascot software as previously described⁴⁹. The halicin molecular weight addition was inserted in the in-house software, and modified peptides were searched for variable modifications on cysteine residues. The relative quantification of halicin-modified peptides was performed using the extracted ion chromatogram (XIC) procedure. Specifically, the current of each peptide ion (modified and unmodified) was extracted from the whole chromatogram, and the area of the corresponding chromatographic peak was measured and employed as quantitative parameter. The percentage of modified species was calculated as reported in the formula below.

$$\% \text{ Cys}_x \text{ modification} = \frac{\Sigma (\text{Area of all peptide ions with Cys}_x \text{ modified by halicin})}{\Sigma (\text{Area of all peptide ions containing modified and not modified Cys}_x)}$$

2.4.3. *Crystallization, data collection, data reduction, structure determination, refinement and final model analysis*

Prior to crystallization 3CL^{pro} WT was incubated at a concentration of 5 mg/ml (148 μ M) with halicin at molar ratio of 1:2 for 1 h (5% DMSO final). Crystals were obtained at RT after 1-2 days with the vapour diffusion technique, in sitting drops, using seeding and 0.09 M NPS (Sodium nitrate, Sodium phosphate dibasic, Ammonium sulfate), 0.1M Hepes/Mops pH 7.5, 20% v/v PEG 500 MME, 10 % w/v PEG 20000 as a precipitant solution. Diffraction data were collected at 100 K, at the XDR2 beamline of the Elettra Sincrotrone Trieste using a 1.000 Å wavelength. The collected dataset was processed with XDS⁵⁰ and Aimless⁵¹ from the CCP4 suite⁵². The structure was solved by molecular replacement with Phaser⁵³ using as a search model 7BB2 (PDB ID). The initial model was

CHAPTER 2

refined alternating cycles of manual model building in COOT⁵⁴ and automatic refinement using Phenix⁵⁵.

2.4.4. *In vitro enzymatic activity assays*

The activity of 3CL^{pro} WT and Cys44Ala was measured using a FRET method⁵⁶. The assay was performed in a 384-well black plate with a reaction volume of 20 μ L containing 20 mM Tris (pH 7.3), 100 mM NaCl, 1 mM DTT, 1 mM EDTA, and various concentrations of tested compounds. After a 30-minute incubation at 37°C, 12 μ M of DABCYL-KTSAVLQ↓SGFRKM-EDANS substrate was added to each well, and the fluorescent signal was read after 15 minutes of incubation at room temperature at Ex 340 nm and Em 490 nm.

2.4.5. *Differential scanning fluorimetry assay*

The stability of 3CL^{pro} WT and Cys44Ala mutant proteins was evaluated using a differential scanning fluorimetry (DSF) assay⁵⁷. The assay was performed in a 50 μ L reaction containing 20 mM HEPES pH 7.8, 120 mM NaCl, and 2 μ M of 3CL^{pro} WT or Cys44Ala mutant proteins. The reactions were pre-incubated with 40 μ M of compounds for 30 minutes at room temperature. After adding 5 $\times$ SYPRO Orange dye, thermal denaturation from 20°C to 95°C was initiated at a scan rate of 1.5°C/min, and the thermal shift (ΔT_m) was calculated as described⁵⁸.

2.4.6. *Native gel electrophoresis*

The conformational analysis of 3CL^{pro} WT and mutant conformers was carried out by native gel electrophoresis as previously described³⁷. Briefly, 5 μ g of protein were loaded in a buffer containing Tris-HCl (pH 7.8), 150 mM NaCl, and 1 mM EDTA onto a 10% 0.5SAR-PAGE running gel at 4°C. Following electrophoresis, the gel was stained with colloidal Coomassie Blue for 1 hour, and the excess dye was removed by washing in deionized water for 12 hours.

2.5 BIBLIOGRAPHY

1. Mesel-Lemoine, M. *et al.* A human coronavirus responsible for the common cold massively kills dendritic cells but not monocytes. *J Virol* **86**, 7577–7587 (2012).
2. da Costa, V. G., Saivish, M. V., Santos, D. E. R., de Lima Silva, R. F. & Moreli, M. L. Comparative epidemiology between the 2009 H1N1 influenza and COVID-19 pandemics. *J Infect Public Health* **13**, 1797–1804 (2020).
3. Chu, R. B. H. *et al.* Comparison of COVID-19 with influenza A in the ICU: a territory-wide, retrospective, propensity matched cohort on mortality and length of stay. *BMJ Open* **13**, e067101 (2023).
4. Kopel, H. *et al.* Comparison of COVID-19 and Influenza-Related Outcomes in the United States during Fall-Winter 2022-2023: A Cross-Sectional Retrospective Study. *Diseases* **12**, 16 (2024).
5. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
6. Wu, F. *et al.* Author Correction: A new coronavirus associated with human respiratory disease in China. *Nature* **580**, E7 (2020).
7. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
8. Bai, C., Zhong, Q. & Gao, G. F. Overview of SARS-CoV-2 genome-encoded proteins. *Sci China Life Sci* **65**, 280–294 (2022).
9. Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).
10. Mussetti, A. *et al.* Handling the COVID-19 pandemic in the oncological setting. *Lancet Haematol* **7**, e365–e366 (2020).
11. Cannalire, R., Cerchia, C., Beccari, A. R., Di Leva, F. S. & Summa, V. Targeting SARS-CoV-2 Proteases and Polymerase for COVID-19 Treatment: State of the Art and Future Opportunities. *J Med Chem* **65**, 2716–2746 (2022).
12. Vandyck, K. & Deval, J. Considerations for the discovery and development of 3-chymotrypsin-like cysteine protease inhibitors targeting SARS-CoV-2 infection. *Curr Opin Virol* **49**, 36–40 (2021).
13. Günther, S. *et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **372**, 642–646 (2021).
14. Goyal, B. & Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb Sci* **22**, 297–305 (2020).
15. Shi, J., Sivaraman, J. & Song, J. Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *J Virol* **82**, 4620–4629 (2008).
16. Zhang, Y. *et al.* An allosteric mechanism for potent inhibition of SARS-CoV-2 main proteinase. *Int J Biol Macromol* **265**, 130644 (2024).
17. Kuzikov, M. *et al.* Identification of Inhibitors of SARS-CoV-2 3CL-Pro Enzymatic Activity Using a Small Molecule in Vitro Repurposing Screen. *ACS Pharmacol Transl Sci* **4**, 1096–1110 (2021).
18. Jain, A. K., Sharma, S., Vaidya, A., Ravichandran, V. & Agrawal, R. K. 1,3,4-thiadiazole and its derivatives: a review on recent progress in biological activities. *Chem Biol Drug Des* **81**, 557–576 (2013).
19. Li, G. *et al.* Recent Advances in c-Jun N-Terminal Kinase (JNK) Inhibitors. *Curr Med Chem* **28**, 607–627 (2021).
20. De, S. K. *et al.* Design, synthesis, and structure-activity relationship of substrate competitive, selective, and in vivo active triazole and thiadiazole inhibitors of the c-Jun N-terminal kinase. *J Med Chem* **52**, 1943–1952 (2009).

21. Serban, G. 2-Amino-1,3,4-thiadiazoles as prospective agents in trypanosomiasis and other parasitoses. *Acta Pharm* **70**, 259–290 (2020).
22. Jang, S. & Javadov, S. Inhibition of JNK aggravates the recovery of rat hearts after global ischemia: the role of mitochondrial JNK. *PLoS One* **9**, e113526 (2014).
23. Blanton, H. L., Pietrzak, A., McHann, M. C. & Guindon, J. Sex and dose-dependent antinociceptive effects of the JNK (c-Jun N-terminal kinase) inhibitor SU 3327 are mediated by CB2 receptors in female, and CB1/CB2 receptors in male mice in an inflammatory pain model. *Brain Res Bull* **177**, 39–52 (2021).
24. Stokes, J. M. *et al.* A Deep Learning Approach to Antibiotic Discovery. *Cell* **180**, 688–702.e13 (2020).
25. van Gent, M. E. *et al.* Synergism between the Synthetic Antibacterial and Antibiofilm Peptide (SAAP)-148 and Halicin. *Antibiotics (Basel)* **11**, 673 (2022).
26. Higashihira, S. *et al.* Halicin Is Effective Against *Staphylococcus aureus* Biofilms In Vitro. *Clin Orthop Relat Res* **480**, 1476–1487 (2022).
27. Booq, R. Y., Tawfik, E. A., Alfassam, H. A., Alfahad, A. J. & Alyamani, E. J. Assessment of the Antibacterial Efficacy of Halicin against Pathogenic Bacteria. *Antibiotics (Basel)* **10**, 1480 (2021).
28. Hussain, Z. *et al.* Study on antibacterial effect of halicin (SU3327) against *Enterococcus faecalis* and *Enterococcus faecium*. *Pathog Dis* **80**, ftac037 (2022).
29. Leeb-Zatorska, B. CORR Insights®: Halicin Is Effective Against *Staphylococcus aureus* Biofilms In Vitro. *Clin Orthop Relat Res* **480**, 1488–1490 (2022).
30. Rodrigues, T. S. *et al.* Inflammasomes are activated in response to SARS-CoV-2 infection and are associated with COVID-19 severity in patients. *J Exp Med* **218**, e20201707 (2021).
31. Mohamed Khosroshahi, L. & Rezaei, N. Dysregulation of the immune response in coronavirus disease 2019. *Cell Biol Int* **45**, 702–707 (2021).
32. Frohman, E. M. *et al.* Part II. high-dose methotrexate with leucovorin rescue for severe COVID-19: An immune stabilization strategy for SARS-CoV-2 induced ‘PANIC’ attack. *J Neurol Sci* **415**, 116935 (2020).
33. Seif, F. *et al.* JAK Inhibition as a New Treatment Strategy for Patients with COVID-19. *Int Arch Allergy Immunol* **181**, 467–475 (2020).
34. Yang, K. S. *et al.* Repurposing Halicin as a potent covalent inhibitor for the SARS-CoV-2 main protease. *Curr Res Chem Biol* **2**, 100025 (2022).
35. Albani, S. *et al.* Unexpected Single-Ligand Occupancy and Negative Cooperativity in the SARS-CoV-2 Main Protease. *J Chem Inf Model* **64**, 892–904 (2024).
36. Fu, L. *et al.* Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat Commun* **11**, 4417 (2020).
37. Huang, L. *et al.* 05SAR-PAGE: Separation of protein dimerization and modification using a gel with 0.05% sarkosyl. *Anal Chim Acta* **1101**, 193–198 (2020).
38. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
39. Verma, N., Henderson, J. A. & Shen, J. Proton-Coupled Conformational Activation of SARS Coronavirus Main Proteases and Opportunity for Designing Small-Molecule Broad-Spectrum Targeted Covalent Inhibitors. *J Am Chem Soc* **142**, 21883–21890 (2020).
40. Xiong, Y. *et al.* Flavonoids in *Ampelopsis grossedentata* as covalent inhibitors of SARS-CoV-2 3CLpro: Inhibition potentials, covalent binding sites and inhibitory mechanisms. *Int J Biol Macromol* **187**, 976–987 (2021).
41. Paciaroni, A. *et al.* Stabilization of the Dimeric State of SARS-CoV-2 Main Protease by GC376 and Nirmatrelvir. *Int J Mol Sci* **24**, 6062 (2023).
42. Vuong, W. *et al.* Author Correction: Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat Commun* **11**, 5409 (2020).

43. Dai, W. *et al.* Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **368**, 1331–1335 (2020).
44. Puxeddu, M. *et al.* 4-(3-Phenylsulfonylindol-2-yl)-1-(pyridin-2-yl)piperazinyl-methanones as Potent Inhibitors of both SARS-CoV-2 and HCoV-OC43 Viruses. *ACS Infect Dis* **10**, 3158–3175 (2024).
45. Messore, A. *et al.* New Thiazolidine-4-One Derivatives as SARS-CoV-2 Main Protease Inhibitors. *Pharmaceuticals (Basel)* **17**, 650 (2024).
46. Cozzolino, F., Iacobucci, I., Monaco, V., Angrisano, T. & Monti, M. Lysines Acetylome and Methylome Profiling of H3 and H4 Histones in Trichostatin A-Treated Stem Cells. *Int J Mol Sci* **22**, 2063 (2021).
47. Florio, D. *et al.* Platinum(II) O,S Complexes Inhibit the Aggregation of Amyloid Model Systems. *Int J Mol Sci* **20**, 829 (2019).
48. Bocedi, A. *et al.* Ultra-Rapid Glutathionylation of Ribonuclease: Is this the Real Incipit of its Oxidative Folding? *Int J Mol Sci* **20**, 5440 (2019).
49. Cozzolino, F. *et al.* ADAM10 hyperactivation acts on piccolo to deplete synaptic vesicle stores in Huntington's disease. *Hum Mol Genet* **30**, 1175–1187 (2021).
50. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132 (2010).
51. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–1214 (2013).
52. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242 (2011).
53. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).
55. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).
56. Costanzi, E. *et al.* Structural and Biochemical Analysis of the Dual Inhibition of MG-132 against SARS-CoV-2 Main Protease (Mpro/3CLpro) and Human Cathepsin-L. *Int J Mol Sci* **22**, 11779 (2021).
57. Stefanelli, I. *et al.* Broad-spectrum coronavirus 3C-like protease peptidomimetic inhibitors effectively block SARS-CoV-2 replication in cells: Design, synthesis, biological evaluation, and X-ray structure determination. *Eur J Med Chem* **253**, 115311 (2023).
58. Qiao, J. *et al.* SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science* **371**, 1374–1378 (2021).