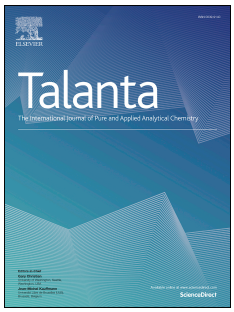
Journal Pre-proof



An integrated metabolomic and proteomic approach for the identification of covalent inhibitors of the main protease (M^{pro}) of SARS-COV-2 from crude natural extracts

Giovanna Baron, Sofia Borella, Larissa della Vedova, Serena Vittorio, Giulio Vistoli, Marina Carini, Giancarlo Aldini, Alessandra Altomare

PII: S0039-9140(22)00620-8

DOI: https://doi.org/10.1016/j.talanta.2022.123824

Reference: TAL 123824

To appear in: *Talanta*

Received Date: 13 July 2022

Revised Date: 3 August 2022

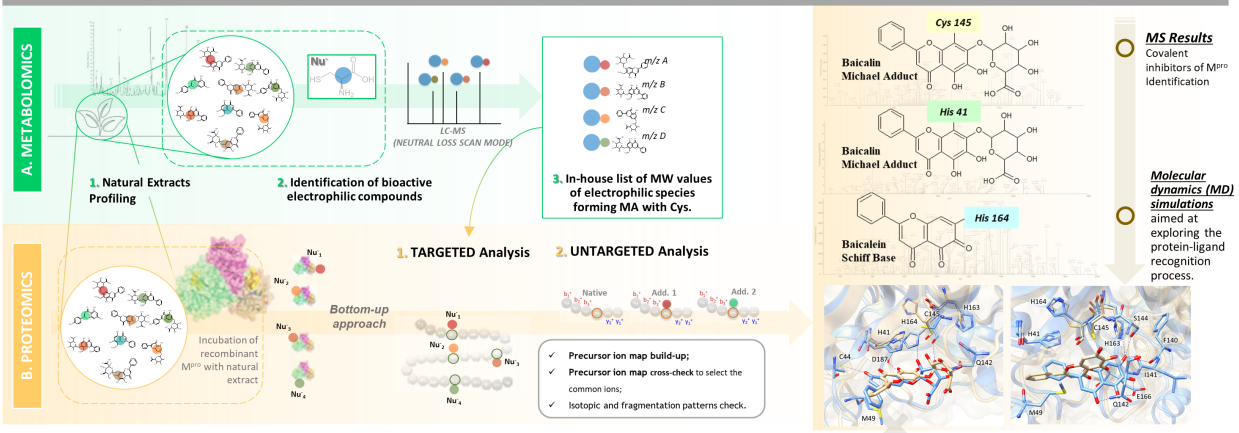
Accepted Date: 5 August 2022

Please cite this article as: G. Baron, S. Borella, L. della Vedova, S. Vittorio, G. Vistoli, M. Carini, G. Aldini, A. Altomare, An integrated metabolomic and proteomic approach for the identification of covalent inhibitors of the main protease (M^{pro}) of SARS-COV-2 from crude natural extracts, *Talanta* (2022), doi: https://doi.org/10.1016/j.talanta.2022.123824.

CRediT author statement

Conceptualization, AA, GB and GA; Formal analysis, AA, GB, SB, LDV, and SV; Funding acquisition, MC and GA; Investigation, AA, GB, SV, GA and GV; Methodology, AA, SV, and GB; Project administration, AA, GV and GA; Resources, MC and GA; Supervision, AA, GB, GV and GA; Validation, AA, SV and GB; Writing – original draft, AA, SV and GA; Writing – review & editing, AA, GB, GV and GA.

An integrated metabolomic and proteomic approach for the identification of covalent inhibitors of the main protease (M^{pro}) of SARS-COV-2 from crude natural extracts



An integrated metabolomic and proteomic approach for the identification of covalent inhibitors of the main protease (M^{pro}) of SARS-COV-2 from crude natural extracts

Giovanna Baron[1], Sofia Borella[1], Larissa della Vedova[1], Serena Vittorio[1], Giulio Vistoli[1], Marina Carini[1], Giancarlo Aldini[1], Alessandra Altomare[1*]

[1] Department of Pharmaceutical Sciences (DISFARM), Università degli Studi di Milano, Via Mangiagalli 25, 20133, Milan, Italy. *giovanna.baron@unimi.it, GB; sofia.borella@studenti.unimi.it, SB; larissa.dellavedova@unimi.it, LDV; serena.vittorio@unimi.it, SV; giulio.vistoli@unimi.it, GV; marina.carini@unimi.it, MC; giancarlo.aldini@unimi.it, GA; alessandra.altomare@unimi.it, AA.*

***** Correspondence: *alessandra.altomare@unimi.it*; (Department of Pharmaceutical Sciences (DISFARM), Università degli Studi di Milano, Via Mangiagalli 25, 20133, Milan, Italy) (A.A.)

ABSTRACT

M^{pro} represents one of the most promising drug targets for SARS-Cov-2, as it plays a crucial role in the maturation of viral polyproteins into functional proteins. HTS methods are currently used to screen M^{pro} inhibitors, and rely on searching chemical databases and compound libraries, meaning that they only consider previously structurally clarified and isolated molecules. A great advancement in the hit identification strategy would be to set-up an approach aimed at exploring un-deconvoluted mixtures of compounds such as plant extracts. Hence, the aim of the present study is to set-up an analytical platform able to fish-out bioactive molecules from complex natural matrices even where there is no knowledge on the constituents. The proposed approach begins with a metabolomic step aimed at annotating the MW of the matrix constituents. A further metabolomic step is based on identifying those natural electrophilic compounds able to form a Michael adduct with thiols, a peculiar chemical feature of many M^{pro} inhibitors that covalently bind the catalytic Cys145 in the active site, thus stabilizing the complex. A final step consists of incubating recombinant M^{pro} with natural extracts and identifying compounds adducted to the residues within the M^{pro} active site by bottom-up proteomic analysis (nano-LC-HRMS). Data analysis is based on two complementary strategies: (i) a *targeted* search applied by setting the adducted moieties identified as Michael acceptors of Cys as variable modifications; (ii) an untargeted approach aimed at identifying the whole range of adducted peptides containing Cys145 on the basis of the characteristic *b* and *y* fragment ions independent of the adduct. The method was set-up and then successfully tested to fish-out bioactive compounds from the crude extract of *Scutellaria baicalensis*, a Chinese plant containing the catechol-like flavonoid baicalin and its corresponding aglycone baicalein which are well-established inhibitors of M^{pro}. Molecular dynamics (MD) simulations were carried out in order to explore the binding mode of baicalin and baicalein, within the SARS-CoV-2 M^{pro} active site, allowing a better understanding of the role of the nucleophilic residues (i.e. His41, Cys145, His163 and His164) in the protein-ligand recognition process.

Keywords: M^{pro}, SARS-CoV-2, Proteomics, Metabolomics, Covalent Binder, Mass Spectrometry

1. Introduction

Coronavirus SARS-CoV-2 is widespread as a global pandemic, and has become the main global health crisis of our time, and although it has caused more than 4 million deaths since its first outbreak in December 2019, there are as yet limited validated antiviral drug candidates against coronavirus infections. This weakness has prompted massive scientific research aimed at, on the one hand, repurposing existing antiviral drugs, despite most of them having been previously rejected for practical use, and, on the other, assessing the efficacy of new molecules [1]. Diverse potential therapeutics, that mainly target the key "players" in the SARS-CoV-2 life cycle, have been reported as effective in inhibiting viral replication, by (i) interfering with the attack on the host cell, (ii) impeding the translation of its genome, (iii) viral replication and (iv) release of new virions [2-10]. In this context a keen interest has been shown in the papain-like protease and the viral 3CL protease, also called the **main protease (M^{pro})**, whose main purpose is to ensure the formation of new copies of the virus by processing the polyproteins pp1a and pp1ab into 16 vital mature non-structural proteins (nsps), including helicase, RNA-dependent RNA polymerase (RdRp), and methyltransferase, which are involved in the viral RNA replication and translation processes, and together ensure the progeny are able to propagate its species. In addition, M^{pro} has no human homolog, is highly conserved among all CoVs, and owing to its high cleavage specificity, all compounds structurally similar to its cleavage sites can selectively inhibit the viral protease with little or no damaging impact on host cell proteases [5, 11-13]. Basically, these considerations, strengthened by the latest literature which reports much evidence collected through biochemical and cell-based assays flanked by computational studies, have led to scientists betting on M^{pro} as a potential target of antiviral drugs, further laying a solid background for the development of broad-spectrum M^{pro} inhibitors as SARS-CoV-2 antivirals that interfere with nsp maturation [14-18].

After the COVID-19 outbreak, the crystal structure of SARS-CoV-2 M^{pro} was rapidly determined, which significantly facilitated its mechanistic study and the development of inhibitors [19]. M^{pro} is a 33.8-kDa cysteine protease characterized by three distinct domains, namely domains I and II connected to domain III through a long loop region that plays a role in protein dimerization. The active site, nested in a chymotrypsin-like fold in the cleft between domains I and II, is composed of very flexible loops intertwined with the catalytic dyad residues His41 and Cys145; with the aid of His41, which acts as a proton acceptor, the Cys145 exerts a nucleophilic attack on the carbonyl carbon of the substrate during the first step of hydrolysis [16]. Around this dyad, M^{pro} forms a conserved binding pocket which comprises four subsites (S1′, S1, S2, and S4) well accommodating the substrate; substrate insertion into the sub-pockets is maintained by a complex network of interactions, mediating polar contacts (hydrogen bonds) and hydrophobic interactions, with the side chains of the conserved aminoacidic residues in the substrate-binding cleft, i.e. Arg40, Cys44, His163, His164, Asp187 [20, 21].

Recently considerable research, based on *v*irtual screening campaigns and HTS tests aimed at identifying from approved drugs and known natural compounds those able to bind the SARS-CoV-2 protease as a target for potential anti-viral activity, has been carried out [22]; the most intriguing aspect is the high number of identified natural compounds acting as M^{pro} inhibitors, indicating that plants are a valuable source

of bioactive compounds, mostly found to be tightly bound to the very crucial key residue of Cys145, thus inhibiting SARS-CoV-2 replication and proliferation in the host [22, 23]. While this approach can be advantageous in terms of reducing costs and study time, the downside of this approach is its limited explorative potential, being as it is confined to the use of what is known: these investigations rely on searching chemical databases and libraries, meaning that they only consider previously structurally-clarified and isolated derivatives; a great advance in the hit identification strategy would be to set-up an analytical strategy aimed at exploring not isolated or even unknown compounds and hence based on un-deconvoluted mixtures of compounds.

Hence, we believe that a method able to fish-out bioactive compounds from crude extracts, containing hundreds of components in a mixture, meaning not deconvoluted in chemical libraries, would greatly extend our current knowledge. However, the search for bioactive compounds from a complex crude mixture is quite challenging and requires sophisticated OMICS based-untargeted methods. In light of these premises, the present study was aimed at developing an innovative High Resolution Mass Spectrometry (HRMS)-based analytical platform integrating metabolomics and proteomics approaches, designed to screen candidate inhibitors targeting the M^{pro} in early-stage drug discovery, in which OMICS approaches could generate unique insights.

The method was set-up and then tested to fish-out bioactive compounds from the crude extract of *Scutellaria baicalensis*, a Chinese plant containing the catechol-like flavonoid baicalin and its corresponding aglycone baicalein (non-covalent, non-peptidomimetic compounds), which are well-established inhibitors of M^{pro} [24].

2. Materials and methods

2.1. Reagents

The natural extract *Scutellaria baicalensis*, containing baicalin and baicalein, was provided by Plantex Research Srl, Milan, Italy. Ultrapure water was prepared by a Milli-Q purification system (Millipore, Bedford, MA, USA). Cysteine (Cys), iodoacetamide (IAA), iron(III) chloride hexahydrate ($FeCl_3 \cdot 6H_2O$), hydrogen peroxide solution 30% w/w (H_2O_2), and tris(2-carboxyethyl)phosphine (TCEP), tetraethylammonium bromide (TEAB), were provided by Sigma-Aldrich (Milan, Italy). Also formic acid (FA), trifluoroacetic acid (TFA), acetonitrile (ACN), and all ultrapure (99.5%) grade solvents used in LC-MS analysis were purchased from Sigma-Aldrich (Milan, Italy). S-TRAP™ columns were provided by Protifi (Huntington, NY).

2.2. Metabolomics analysis

2.2.1. Analytical profiling of the natural extract (*S. baicalensis)*

2.2.1.1. Sample preparation

The *S. baicalensis* extract was weighed and resuspended up to a concentration of 2 mg $\cdot$ mL^{-1} in methanol. Thereafter, a 1:4 dilution in the starting mobile phase of the LC gradient (0.1% formic acid in water) was performed to achieve the final concentration of 0.5 mg $\cdot$ mL^{-1}.

2.2.1.2. LC-HRMS analysis (LTQ Orbitrap XL *Mass Spectrometer*)

The analytical platform used comprises the Ultimate 3000 HPLC (Dionex), coupled to an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, USA), set to work as described by *Baron et al.* [25] to acquire both full MS and MS/MS spectra to achieve the qualitative extract profiling of polyphenols.

The database for the targeted data analysis was built searching in the literature for those compounds known to be present in *S. baicalensis* extract (n = 96) [26-37]. The identification was carried out on the basis of the exact mass (mass tolerance of 5 ppm), the isotopic and the fragmentation patterns using the Xcalibur *QualBrowser* tool (2.0.7, ThermoFisher Scientific Inc., Milan, Italy). The semi-quantitative composition was obtained by calculating the relative percentage of each component as described by Equation *(1)*:

$$\frac{AUC_{compound}}{\sum AUC} \times 100 \qquad (1)$$

The relative percentage calculated for each compound allowed the estimation of the content in the extract, although this without considering the different ionization efficiency of each molecule.

2.2.2. Electrophilic compound identification and reaction kinetics study

2.2.2.1. Sample preparation

S. baicalensis extract : Cysteine

Cysteine was dissolved in PBS 100 mM pH 7.4 up to a final concentration of 2 mM in the incubation mixture, while the *S. baicalensis* extract solution prepared at 5 mg · mL^{-1}in NaOH 0.02 M was adjusted with HCl 0.5 M to reach pH 7.4 (final concentration: 4.56 mg · mL^{-1}). The latter solution was diluted 1:4 in PBS 100 mM pH 7.4 in the incubation mixture containing cysteine (2 mM). An aliquot of the incubation mixture was withdrawn and diluted 1:10 with H_2O/CH_3CN/HCOOH (70:30:0.1, % v/v) after 0, 2, 4, 6 and 24 hours to stop the reaction. The stoichiometry of reaction (~5:1, *S. baicalensis* extract:cys) was chosen considering the high reactivity of cysteine.

Baicalin : Cysteine (oxidative and non-oxidative conditions)

Fenton reaction mixture was prepared by incubating 5 mM baicalin with 50 µM $FeCl_3 \cdot 6H_2O$ in PBS 10 mM pH 7.2, spiked with 500 µM H_2O_2 (FeII : H_2O_2, 1:10), for 10 minutes at room temperature with gentle shaking. Considering the relative content estimated for the baicalin compound in the natural extract and the stoichiometric ratio *S. baicalensis extract:cys* tested, the stoichiometry of reaction with pure baicalin was adjusted in favor of cysteine, 1:2, *baicalin:cys*. Cysteine was dissolved in PBS 100 mM pH 7.4 to reach a concentration of 4 mM, while baicalin in both oxidative and non-oxidative conditions was dissolved at 5 mM in PBS 10 mM pH 7.2. The incubation mixture was prepared as follows: 500 µL of 4 mM cysteine, 200 µL of 5 mM of baicalin (oxidative and non-oxidative conditions), 300 µL of PBS 100 mM pH 7.4. As previously described, an aliquot of 50 µL of the mixture was withdrawn and diluted 1:10 with H_2O/CH_3CN/HCOOH (70:30:0.1, % v/v) after 0, 15 min, 30 min, 1 h, 2 h, 4 h, 6 h and 24 h to stop the reaction.

2.2.2.2. Reaction kinetics study by LC-HRMS (*S. baicalensis* /baicalin : cysteine)

The LC-HRMS method described in paragraph *2.3.1.2* was modified only in the chromatographic conditions setting to reduce the time of analysis to 32 minutes. Briefly, the multistep gradient was set as follows: 0-15 min from 1 % of B to 20 % of B, 15-25 min from 20 % of B to 70 % of B, 25-28 min isocratic of 70 % B, 28.1-32 min isocratic of 1 % B.

2.3. Proteomics analysis

2.3.1. Characterization and localization of protein adducts

2.3.1.1. Sample preparation

M^{pro} incubation with S. baicalensis extract

Lyophilized recombinant M^{pro} was resuspended at a concentration of 1 µg/µL in PBS 100 mM pH 7.4, while the extract, based on its solubility, was dissolved in 100% DMSO at two different concentrations (330 µg/µL and 165 µg/µL). The extract solutions were then diluted in PBS 100 mM pH 7.4, so as to reduce the relative content of DMSO to 0.1%; the pH of the prepared extract solutions was adjusted with NaOH 0.1 M to mimic physiological conditions (pH = 7.4).

The incubation mixtures *M^{pro}:S. baicalensis* were prepared at two different stoichiometric ratios: 1:3.3, 1:1.65 (weight : weight); in order to initially speculate the reaction kinetics, which we expected to be slower than those obtained by incubating the extract with free cysteine, the *M^{pro}:S. baicalenis* mixtures were incubated in the Thermomixer at 37 °C, at a speed of 450 rpm, for 2, 4, and 12 hours.

M^{pro} incubation with pure baicalin (oxidative / non-oxidative conditions)

Lyophilized recombinant M^{pro} was resuspended at a concentration of 1 µg/µL in PBS 100 mM pH 7.4, while pure baicalin aliquots for both the oxidative (Fenton reaction) and non-oxidative experiments were prepared as reported in paragraph *2.3.2.1* at a final concentration of 1.33 µg/µL in PBS 100 mM (pH = 7.4), prior the co-incubation. In order to study the reaction in both conditions the *M^{pro} : baicalin* mixtures were incubated in the Thermomixer at 37 °C, at a speed of 450 rpm, for 2, 4, and 12 hours.

Protein digestion with S-TRAP™

Samples collected from all the prepared incubation mixtures were then processed according to the canonical bottom-up proteomics procedure. Given the need to miniaturize the experimental model due to the high cost of recombinant M^{pro}, the proteolytic digestion step was the most critical issue to be solved during the optimization of the analytical platform, starting from the selection of the most suitable proteolytic enzyme(s): the choice was guided by the need to obtain a satisfying mixture of peptides including the two residues of the catalytic dyad (Cys145 / His41). *In silico* protein digestion simulation was carried out by means of the PeptideMass software tool, (https://www.expasy.ch/tools/peptide-mass.html) with different enzymes, or combinations of them (e.g. trypsin and chymotrypsin).

To maximize the digestion yield from the negligible amount of recombinant protein incubated with the extract, the great potential of this new technology, using convenient spin columns, S-TRAP™, was exploited. Sample preparation begins by dissolving the samples with 5% SDS followed by further denaturation

by acidification and subsequent exposure to a high concentration of methanol. The collected incubation mixtures were then dissolved 1:1 in lysis buffer (10% SDS, 100 mM TEAB). The reduction of disulfide bridges was performed by adding 5 µL of the reducing solution of tris(2-carboxyethyl)phosphine (5 mM TCEP in 50 mM AMBIC) and incubating the mixtures under gentle shaking in the Thermomixer for 10 min at 95 °C. Next, a volume of 5 µL of iodoacetamide solution (20 mM IAA in 50 mM AMBIC) was added, with the aim of alkylating the free thiol residues; incubation in this case was carried out for 45 minutes, at room temperature in the dark. Proteins were further denatured by acidification to pH < 1 by adding a 12% phosphoric acid solution in water (1:10 relative to sample volume). The next step consisted of the sample loading: 165 µL of the binding buffer (90% methanol, 10% TEAB 1 M) and 25 µL of sample were simultaneously added onto the spin columns, then centrifuged at a speed of 4000 xg, for 1 minute at 15 °C; this step was repeated until the protein sample was fully loaded into the columns.

This was followed by three washing steps with the loading of 150 µL of binding buffer onto the S-TRAP columns, which were then centrifuged (1 min, 4000 xg, 15°C) to remove all the excess unbound sample. At this point the proteolytic digestion step was begun by adding 25 µL of a mixture consisting of: 1 µg of trypsin (sequencing-grade trypsin, Roche) diluted 0.2 µg/µL in 50 mM AMBIC, and 1 µg of chymotrypsin (sequencing-grade chymotrypsin, Roche) diluted 0.2 µg/µL in 50 mM AMBIC, which requires the cofactor calcium chloride (5 mM $CaCl_2$ in 50 mM AMBIC). Upon addition of proteases, the physical confinement within the submicron pores of the trap forces the substrate and protease interaction to yield rapid digestion; consequently, protein digestion requires much shorter incubation times, *i.e.* 1.5 hours, at 47 °C, under slow stirring (400 rpm). The peptide mixture was recovered by loading two different solutions onto the columns (elution and 1' centrifugation, 4000 xg, 15 °C): 40 µL of elution buffer 1 (10% H_2O, 90% ACN, 0.2% FA) and 35 µL of elution buffer 1 (60% H_2O, 40% ACN, 0.2% FA). The collected peptide mixtures were dried in the SpeedVac (Martin Christ.) at 37 °C and stored at -80°C until analysis.

2.3.1.2. nLC-HRMS Orbitrap Elite™ *Mass Spectrometer* analysis

Tryptic peptides, resuspended in an appropriate volume (30 µL, sufficient for three technical replicates) of 0.1% TFA mobile phase, were analyzed using a Dionex Ultimate 3000 nano-LC system (Sunnyvale CA, USA) connected to the Orbitrap Elite™ *Mass Spectrometer* (Thermo Scientific, Brema, Germania) equipped with an ionization source, the *Nanospray Ion Source* (Thermo Scientific Inc., Milano, Italia).

For each sample, 5 µL of solubilized peptides were injected in triplicate onto the Acclaim PepMap™ C18 column (75 µm x 25 cm, 100 Å pores, Thermo Scientific, Waltham, Massachusetts, USA), "protected" by a pre-column, the Acclaim PepMap™ (100 µm x 2 cm, 100 Å pores, Thermo Scientific, Waltham, Massachusetts, USA), thermostatically controlled at 40 °C. The chromatographic method used the binary pump system (LC / NC pumps) starting with sample loading onto the pre-column (3 minutes) using the loading pump with a flowrate of 5 µL/min of mobile phase consisting of 99% of buffer A_LC, 0.1% TFA / 1% of buffer B_LC, 0.1% FA in ACN. After the loading valve switched, peptide separation was performed by the Nano Column Pump (NC_pump) with a 117 minute linear gradient (0.3 µL/min) of buffer B_NC_pump (0.1%

FA in ACN) from 1% to 40%, and a further 8 minutes of linear gradient from 40% to 95% (Buffer B_NC_pump); 5 minutes at 95% of buffer B_NC_pump to rinse the column followed the separative gradient, and finally 7 minutes served to re-equilibrate the column to initial conditions. The total run time is 144 minutes. A washout injection with pure acetonitrile (5 µL) was performed between sample injections.

The nanospray ionization source was set as follow: positive ion mode, spray voltage at 1.7 kV; capillary temperature at 220 °C, capillary voltage at 35 V; tube lens offset at 120 V. The orbitrap mass spectrometer operated in data-dependent acquisition (DDA) mode set to acquire full MS spectra in "profile" mode over a scan range of 250-1500 *m/z*, with the AGC target at $5x10^5$, and resolution power at 120,000 (FWHM at 400 *m/z*); tandem mass spectra were instead acquired by the linear ion trap (LTQ), set to automatically fragment in CID mode the ten most intense ions for each full MS spectra (over $1x10^4$ counts) under the following conditions: centroid mode, normal mode, isolation width of the precursor ion of 2.5 *m/z*, AGC target $1x10^4$ and normalized collision energy of 35 eV. Dynamic exclusion was enabled (exclusion dynamics for 45 seconds for those ions observed 2 times in 10 seconds). Charge state screening and monoisotopic precursor selection were enabled, singly and unassigned charged ions were not fragmented. Xcalibur software (version 3.0.63, Thermo Scientific Inc., Milan, Italy) was used to control the mass spectrometer.

2.3.1.3. Data elaboration

Characterization and localization of protein adducts (targeted approach)

Raw data acquired by HR-MS were processed by means of Proteome Discoverer software (version 2.2.0.338, Thermo Fisher Scientific, USA), designed to computationally process the full and fragmentation mass spectra to obtain protein lists. Matching the experimental mass spectra with theoretical ones, obtained by the *in silico* digestion of the M^{pro} sequence (Uniprot ID: P0DTD1, AA 3264-3569), is accomplished by the SEQUEST algorithm, developed to automatically cross-validate the PSMs (peptide spectral matches) generated.

For the targeted analysis, aimed at characterizing the protein adducts of baicalin and/or baicalein with M^{pro}, specific experimental parameters concerning the instrument setting for HRMS acquisition were listed in the processing workflow: mass range between 350.0 Da - 5000.0 Da, activation type mode: Any, Total intensity Treshold 1, S/N Treshold 3, 10 ppm as Precursor Mass Tolerance, and 0.5 Da as Fragment Mass Tolerance; in addition, to allow *in silico* digestion of protein species, the proteolytic enzyme/s used (trypsin/chymotrypsin) and the maximum number of missed cleavages allowed (3), were set. Besides, cysteine carbamidomethylation was set as fixed modification (+57.02147), while methionine oxidation was allowed as a variable modification (+15.995 Da) along with potential aspecific carbamidomethylation of Lys and His. Furthermore, all the mass shifts considered plausible according to the hypothesized reaction mechanisms (Micheal addition) were also included as variable modifications targeting the nucleophilic moyeties of Cys, Lys and His, of which some structure formulas are reported as example (Figure 1).

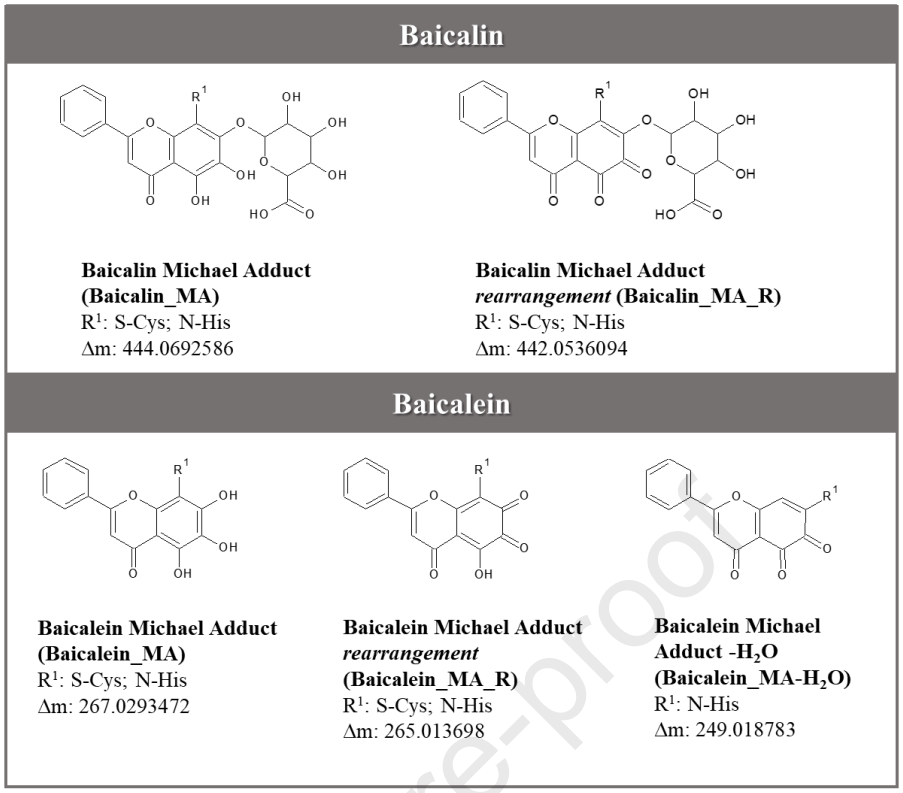


Figure 1. Structure formulas of some of the modifications to be investigated that target the nucleophilic amino acid residues of the protein searched for: (A) Baicalin_MA, (B) Baicalin_MA_R, (C) Baicalein_MA, (D) Baicalein_MA_R, (E) Baicalein_MA-H_2O.

Untargeted characterization of Cys 145 covalent adducts

In order to confirm the adducts identified through the targeted approach, and to expand the investigation to include reactive molecules not considered so far, an untargeted data processing approach was applied; the main focus of this part of the work was the residue Cys 145 (catalytic dyad). Starting from the sequence of the native peptide containing Cys 145 the most intense fragments of the *b*- and *y*-series not including the target residue were selected. The *m/z* values of the selected fragment ions were used to generate so-called ion maps. Using the Xacalibur Qual Browser tool, for each raw data acquired according to the proteomic approach described above, 7 extrapolated precursor ion maps were built with a tolerance value of 0.5 Da (Figure S1, panels A-G, *Supplementary Material*), then compared with each other to select for common precursor ions only. Fragmentation patterns of the selected precursor ions were then manually checked to confirm the presence of all characteristic input fragments that generated the ion maps.

2.5. Molecular modeling

2.5.1. Molecular docking

The binding mode of baicalin within the SARS-CoV-2 M^{pro} active site was investigated by molecular docking using the crystal structure of M^{pro} in complex with baicalein (PDB ID 6M2N) as 3D coordinates [38]. The water molecules were removed and missing atoms were added to the protein by Vega ZZ suite [39]. H++ webserver was employed to add hydrogens and define both the tautomeric state of the histidines and the

arrangement of asparagine and glutamine residues [40]. The structure underwent 10,000 steps of energy minimization keeping the backbone fixed in order to preserve the resolved folding. The ligand was removed from the optimized structure which was then employed to perform docking studies. The structure of baicalin was retrieved from PubChem [41] and optimized by the PM7 semi-empirical method [42]. The so prepared ligand was docked into the M^{pro} active site by employing the software Gold 5.8.1 [43]. Docking simulation was performed as described elsewhere [44] with minor modifications. Briefly, the binding site was defined in order to include the residues within 10 Å from the native ligand. Twenty genetic algorithm runs were performed applying the default settings keeping the protein rigid. The "allow early termination" option was deactivated and the docking solutions whose RMSD value was less than 0.75 Å were clustered together. The docking protocol was validated by re-docking the co-crystallized ligand into the M^{pro} active site allowing the successful reproduction of the experimental binding conformation with a RMSD value of 0.87 Å. The best scoreing docking pose was selected for the following computational studies.

2.5.2. Molecular dynamics

Molecular dynamics (MD) simulations were carried out by using the software Amber v18 [45]. The crystal structure of SARS-CoV-2 M^{pro} in complex with baicalein (PDB ID 6M2N) and the complex baicalin-M^{pro} obtained from the docking simulation were used as starting coordinates. General Amber force field (GAFF) parameters [46] were assigned to the ligands, while partial charges were computed by the AM1-BCC method as implemented in Antechamber [47]. The ff14SB force field [48] was employed for the parametrization of the protein. The systems were solvated in a box of TIP3P water molecules and neutralized adding an appropriate number of Na+ and Cl- ions to reproduce the physiological salt concentration of 0.15 M. The so prepared complexes were subjected to three steps of energy minimization involving first the hydrogen atoms, then the water molecules and finally the protein side chains. Subsequently, 20 ps of heating phase was performed gradually increasing the temperature to 300 K employing the Langevin thermostat and applying positional restraints (5 Kcal/mol) to the Cα atoms. Two equilibration steps were performed first using the NVT ensemble for 50 ps, maintaining the Cα restrained, and then the NPT ensemble keeping the pressure around 1 atm by means of the Berendsen barostat and gradually reducing the weight of the restraints. Finally, a production run of 500 ns was performed at constant pressure without any restraint. All the bonds involving hydrogen atoms were restrained by the SHAKE algorithm with a timestep of 2 fs. Electrostatic interactions were computed by particle-mesh Ewald (PME) method and periodic boundary conditions were applied. RMSD analysis was performed by the AmberTools 18 cpptraj module [49] while the obtained frames were clustered by means of the TTClust program [50].

3. Results

3.1. Method overview

Figure 2 shows an overview of the approach we here propose to identify hit compounds effective as M^{pro} inhibitors from un-deconvoluted compound mixtures such as natural extracts. The principle of the method is based on the fact that most M^{pro} inhibitors act by covalently binding the nucleophilic residues constituting the active site, forming a stable complex and inducing enzyme inactivation. Based on this mechanism, the proposed approach consists of selecting from the mixture, by an integrated metabolomic and proteomic approach, those molecules which covalently bind the nucleophilic sites of the M^{pro} catalytic domain, *i.e.* His41 and Cys145. In the literature other molecules, *e.g.* baicalin and baicalein, have been described as non-covalent binders; despite the presence of a pyrogallol moiety, which is a potentially electrophilic warhead, as found for analogues, *i.e.* myricetin, baicalin and baicalein have been found to inhibit M^{pro} through a non-covalent engagement. To detect this class of compound, the experimental conditions were forced in terms of incubation time and oxidation milieu in order to catalyze the protein adduction and also to detect compounds with a limited electrophilicity.

The OMICS-based analytical platform which was optimized to identify and characterize M^{pro} covalent binders as potential protease' inhibitors, is overall based on two different approaches, a targeted and an un-targeted approach.

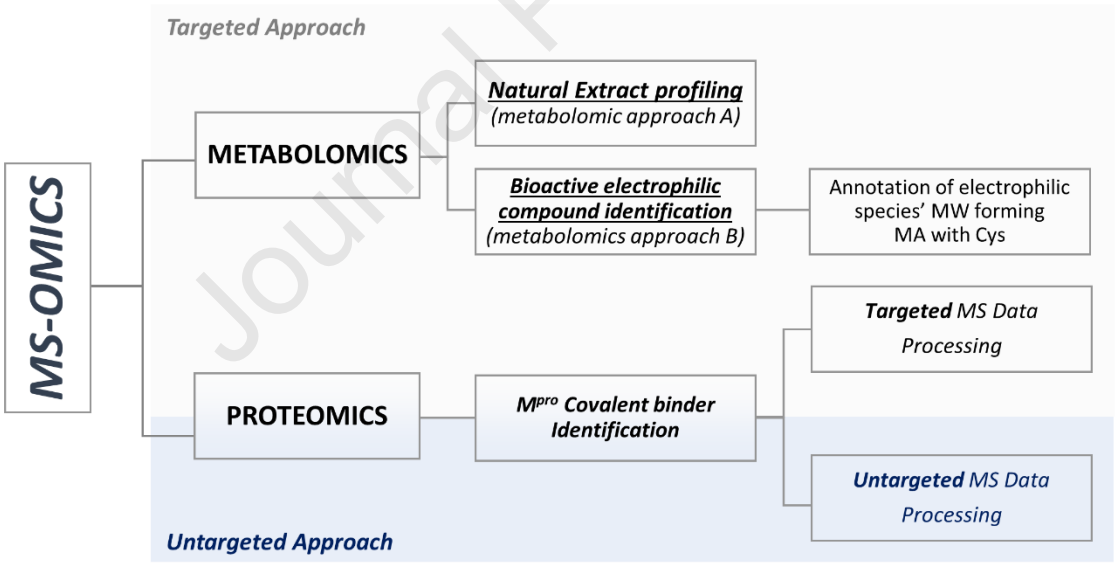


Figure 2. Experimental design of the HRMS analytical platform used to identify covalent binders of M^{pro} from complex natural extracts.

The targeted approach consists, as a first step, of annotating the MW of the components of the mixture (metabolomic approach A); the mixture is then incubated with cysteine as a model of soft nucleophilic substrate and the electrophilic compounds which form a Michael adduct (MA) are identified by calculating the mass difference between the detected MA and Cys, corresponding to the MW of the electrophilic compound. The proteomics approach follows the metabolomic analysis. RP-nanoLC-ESI-HRMS/MS analysis of M^{pro} incubated in the presence of the extract and then digested with trypsin and chymotrypsin is then carried out.

The MA between the electrophilic compounds identified in the previous metabolomic step and the nucleophilic sites of M^{pro} are then searched for by setting the electrophilic moiety as variable modification on Cys, His and Lys. Confirmation of the adduct is then obtained through a manual check of the MS/MS fragmentation spectra.

The untargeted approach is then applied with the intention of identifying the full range of target nucleophilic' containing peptides (M^{pro}) bearing a covalent adduct without any prior knowledge of the electrophilic compounds. Based on the sequence of the native peptides containing the target nucleophilic residues and their characteristic fragmentation spectra, the most intense fragments of the *b*-series and *y*-series, not including the target nucleophilic amino acid residue, were selected so that they would be independent of the modification. The *m/z* values of the selected fragment ions were then used to generate the so-called ion maps, namely the lists of the *m/z* values of those precursor ions (parent ions) which when fragmented release ions with the same *m/z* values as the product ions specified as input. The MW of the electrophilic species was then calculated on the basis of the identified precursor ion and adduction was confirmed by MS/MS experiments. In the case of M^{pro}, Cys145 represents one of the most recognized target nucleophilic residues and hence the following peptide $GSFLNGSC_{145}GSVGF$, arising from the enzymatic digestion of the protein and containing Cys145, was selected for the untargeted approach. The selected fragment ions arising from MS/MS fragmentation and not containing the Cys145 (Figure S2, *Supplementary Material*) were then selected to generate the ion maps as depicted in Table 1.

	m/z
$b3^+$	292.12918
$b4^+$	405.21325
$b5^+$	519.25617
$b7^+$	663.30967

	m/z
$y3^+$	322.17613
$y4^+$	409.20816
$y5^+$	466.22962

Table 1. *m/z* values of characteristic *b*- and *y*-series fragments not containing the Cys145 residue.

Targeted and untargeted results were then merged, and the MW of the electrophilic species contained in the mixture and able to react and covalently bind M^{pro} nucleophilic sites were determined. Finally, the nature of the selected electrophilic compound was defined on the basis of the accurate MW, isotopic pattern, elemental composition and MS/MS by searching the database of natural compounds provided in the literature or in the case that it was still unknown, characterized by isolation and structural analysis.

The proposed method was first applied, to test its suitability, to a natural extract containing well-known inhibitors of M^{pro}, namely the crude extract of *Scutellaria baicalensis* containing baicalin and baicalein [24]. In the next paragraphs the different steps of the approach are described and discussed.

3.2. Metabolomics analysis

3.2.1. Analytical profiling of the natural extract S. baicalensis

The TIC of the *S. baicalensis* extract obtained by LC-HRMS analysis in negative ion mode is shown in Figure S3 (*Supplementary Material*). The chromatogram identifies 12 well-resolved peaks eluting within 60 minutes. On the basis of the analysis, a table reporting the MW of all the extract constituents is generated. In line of principle, this step should be limited to the identification of the MW of the components while compound characterization is not required since it can be taken into account only at the last stage and is limited to the electrophilic compound/s which bind/s the target. However, for a better description of the method and also in consideration of the reduced number of the extract components, a full profiling of the extract was carried out at this stage.

Compounds were identified by comparing the experimental information obtained for each chromatographic peak, *i.e.* MW, elemental composition, isotopic pattern and MS/MS fragmentation pattern, with those contained in an in-house database compiled retrieving information of the plant constituents from the literature. The list of the identified constituents is reported in Table 2.

Peak	***Compound Name***	***Chemical Formulae***	***[M-H]⁻***	***MS/MS***	***AUC***	***Relative Abundance (%)***
1	*Scutellarin*	$C_{21}H_{18}O_{12}$	461.0715	285	44240099	4.694
2	*Hispidulin-7-O-glucuronide*	$C_{22}H_{20}O_{12}$	475.087	299	14411499	1.529
3	***Baicalin***	**$C_{21}H_{18}O_{11}$**	**445.0765**	**269**	**341455963**	**36.234**
4	*Dihydrobaicalin*	$C_{21}H_{20}O_{11}$	447.0918	271	54385342	5.771
5	*Apigenin-7-O-glucuronide*	$C_{21}H_{18}O_{11}$	445.076	269	35292559	3.745
6	*Scutevulin-7-O-glucuronide*	$C_{22}H_{20}O_{12}$	475.0869	299	8454936	0.897
7	*Chrysin-7-O-glucuronide*	$C_{21}H_{18}O_{10}$	429.0814	253	237540302	25.207
8	*Oroxylin A-7-O-glucuronide*	$C_{22}H_{20}O_{11}$	459.0922	283	85079780	9.028
9	*Wogonoside*	$C_{22}H_{20}O_{11}$	459.0921	283	53133150	5.638
10	*Apigenin*	$C_{15}H_{10}O_5$	269.0448	197-225	16780909	1.780
11	***Baicalein***	**$C_{15}H_{10}O_5$**	**269.044**	**241-251**	**27937120**	**2.964**
12	*Wogonin*	$C_{16}H_{12}O_5$	283.0605	268	23635134	2.508

Table 2. Summary of information on *S. baicalensis* compounds obtained by LC-HRMS analysis and compared with chemical libraries, listed by peak number in the MS profiling (Figure S3).

The two active compounds as M^{pro} inhibitors (positive control) baicalin and baicalein were identified with baicalin representing, at least on the basis of peak area, the most abundant component of the extract.

3.2.2. Identification of *S. baicalensis* electrophilic compounds and reaction kinetics study

Once the MWs of the extract components were defined, we moved to the second step, aimed at identifying the electrophilic compounds able to react with the thiolate, a reaction which is the basic mechanism for M^{pro} inhibitors acting as covalent binders. Cysteine was used as a thiol model and its incubation with the extract was carried out at different time points (0, 2, 4, 6, and 24 hours). Each incubated mixture was then analyzed by LC-HRMS (data-dependent scan mode) as reported in the method section. The $[M-H]^-$ value of the Michael adduct between Cys and each extract component was calculated, and the values summarized in Table 3. The ion current trace for each calculated $[M-H]^-$ value was then extracted with a 5 ppm tolerance and the presence of peaks in the SIC chromatograms searched for.

Adduct	***Formula***	***[M-H]⁻***
Cys-*Scutellarin*	$C_{24}H_{23}NO_{14}S$	580.07610
Cys-*Hispidulin*	$C_{25}H_{25}NO_{14}S$	594.09175
Cys-***Baicalin***	**$C_{24}H_{23}NO_{13}S$**	**564.08118**
Cys-*Dihydrobaicalin*	$C_{24}H_{24}NO_{13}S$	566.09683
Cys-*Apigenin*	$C_{24}H_{23}NO_{13}S$	564.08118
Cys-*Scutevulin*	$C_{25}H_{25}NO_{14}S$	594.09175
Cys-*Chrysin*	$C_{24}H_{23}NO_{12}S$	548.08627
Cys-*Oroxylin*	$C_{25}H_{25}NO_{13}S$	578.09683
Cys-*Wogonoside*	$C_{25}H_{25}NO_{13}S$	578.09683
Cys-*Apigenin*	$C_{18}H_{15}NO_7S$	388.04910
Cys-***Baicalein***	**$C_{18}H_{15}NO_7S$**	**388.04910**
Cys-*Wogonin*	$C_{18}H_{15}NO_6S$	402.06474
Cys-*Myricetin*	$C_{18}H_{15}NO_{10}S$	436.03384

Table 3. Molecular formulae and calculated monoisotopic masses of the hypothesized Michael adducts with Cys.

Analysis of the chromatograms revealed the formation of only one adduct relative to the Michael adduct involving cysteine and baicalin at *m/z* 564.08118 (experimental *m/z* 564.08167, Δppm 0.869). Time

dependent analyses also revealed that the adduct peaked after 2 hours of incubation and then reduced time-dependently at the following time-points (Figure 3).

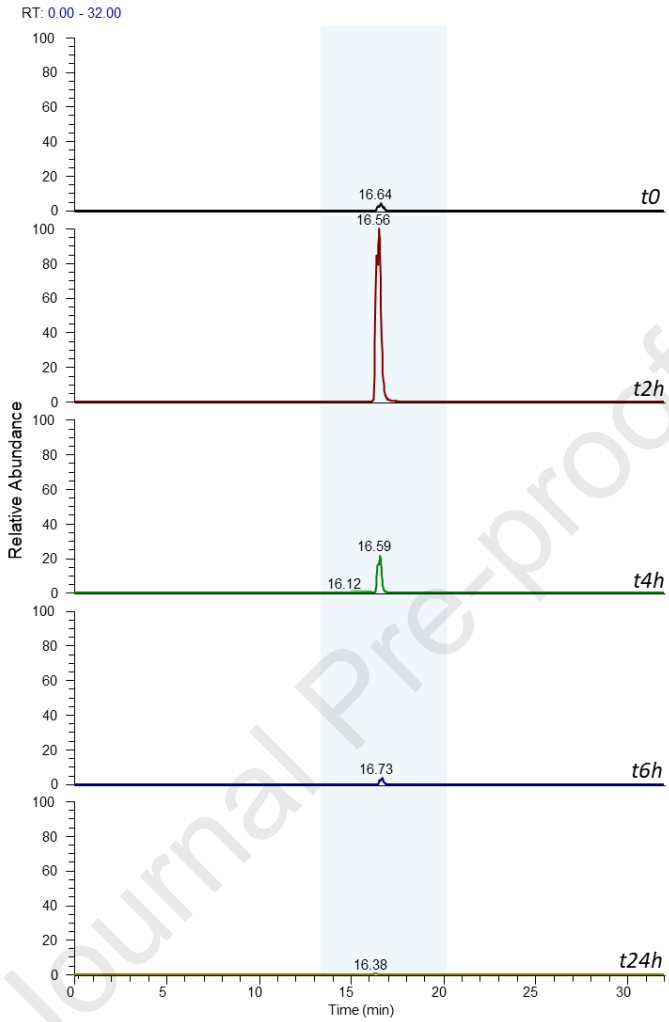


Figure 3. Ion current of the *m/z* 564.08118 ion relative to the Michael adduct between cysteine and baicalin; chromatograms are relative to the reaction mixture of Cys incubated with *S. baicalensis* incubated at the following time points: 0, 2, 4, 6 and 24 hours.

Formation of the Michael adduct was then confirmed by checking the MS/MS fragmentation pattern obtained in data dependent scan mode. Figure 4, panel A, shows the full MS spectrum of the peak, characterized by a main molecular ion at *m/z* 564.08167 (Δppm = 0.869) together with the putative chemical structure which was further confirmed by the fragmentation pattern (Figure 4, panel B).

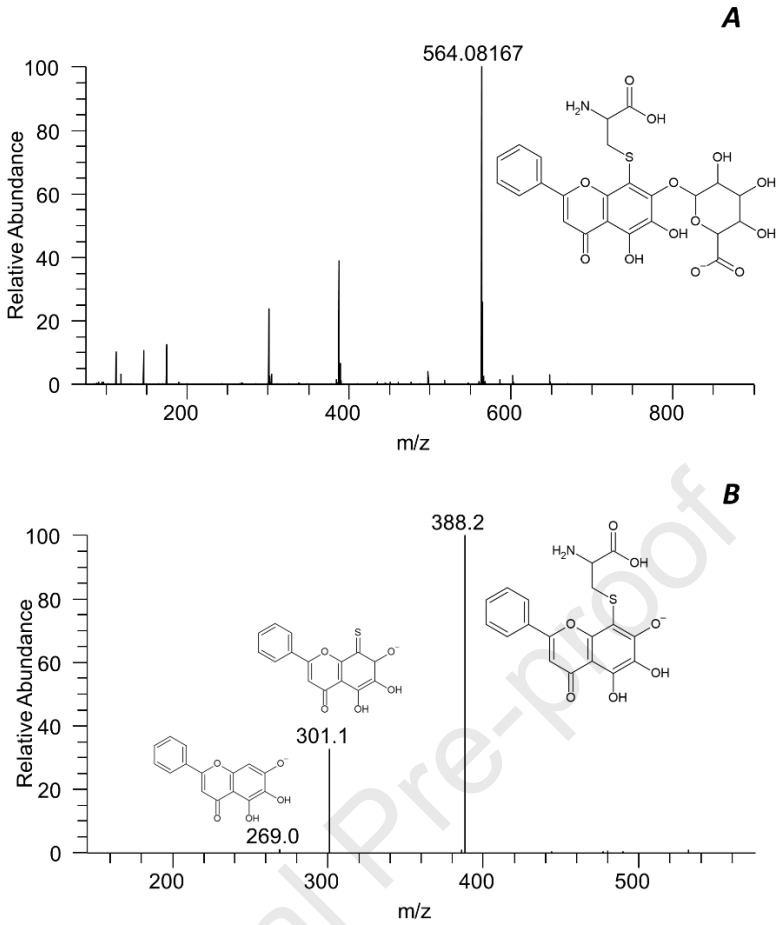
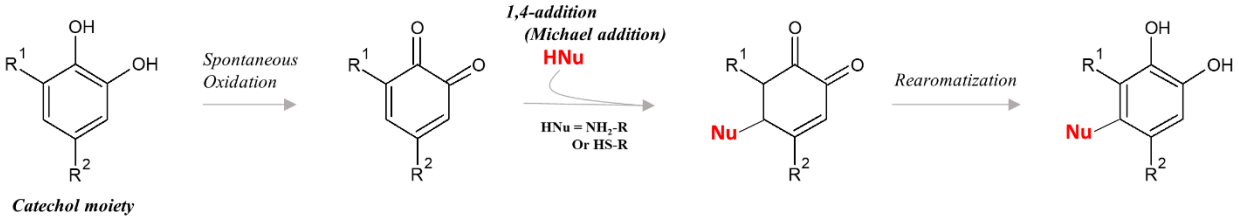


Figure 4. Full MS spectrum of the peak with RT 16.6 min characterized by the main ion at *m/z* 564.08167 relative to the Michael adduct between the thiol group of cysteine and baicalin (*panel A*). MS/MS spectrum of the ion at m/z 564.08167 is reported in *panel B* confirming the structure attribution.

3.2.3. Reaction of baicalin with cysteine in oxidative and non-oxidative reaction conditions

The reaction between baicalin and the thiolate of Cys most likely occurs through the formation of the quinone electrophilic intermediate which is formed by an oxidative activation of the catechol moiety. This is the basic mechanism through which many catechol-containing natural products react with M^{pro}, as recently demonstrated for myricetin. Figure 5 reports a general reaction mechanism which explains the reaction of catechol containing compounds with nucleophilic sites forming the corresponding Michael adducts.

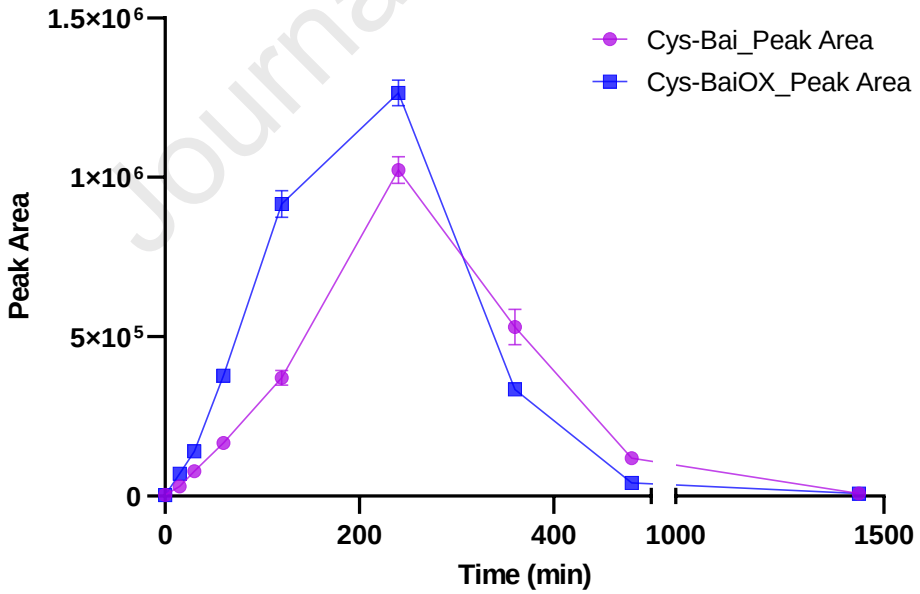


435

436 **Figure 5.** Spontaneous reaction between Catechol containing compounds and nucleophiles such as thiols and
437 amines to perform a 1,4-addition (Michael addition); scheme adapted from [51].

438

439 Hence, catalyzing the activation of the catechol moiety to the corresponding quinone could be an
440 interesting strategy to catalyze the formation of the adduct with Cys and to identify the less reactive
441 compounds. Moreover, this condition mimics the oxidizing milieu at the site of viral replication where an
442 inflammatory condition occurs, accompanied by oxidative stress. The std of baicalin was used to test the effect
443 of an oxidizing condition on the adduct formation and the formation of quinone was catalyzed by adding H_2O_2
444 and ferrous ion (Fenton reaction) to the reaction mixture (Figure S4, *Supplementary material*). Figure 6 shows
445 the time-dependent peak-areas relative to the Cys-baicalin adduct at different time points. It is quite evident
446 that the catalyzed formation of quinone by Fenton reaction not only accelerates the reaction kinetic but also
447 produces a greater amount of the Michal adduct.



448

449 **Figure 6.** Time-dependent peak-areas of Cys-baicalin adduct at different time points.

450

3.3. Proteomic analysis

3.3.1. Targeted approach

The selection of the proteolytic enzyme/s to be used for protein digestion represented an issue that needed to be scrupulously optimized to generate the peptides containing the nucleophilic target residues of the catalytic dyad (Cys145 / His41). *In silico* protein digestion simulation (PeptideMass software, https://www.expasy.ch/tools/peptide-mass.html, accessed on 14th June 2021) with different enzymes, or combinations of enzymes, was carried out identifying the combination of trypsin (cleavage sites: Arg and Lys) and chymostrypsin (cleavage sites: Tyr, Trp, Met, Phe and Leu) as the most suitable to generate LC-HRMS detectable peptides containing the nucleophilic targets. Moreover, the great potential of a new technology based on spin columns S-TRAP™ was tested to maximize the digestion yield from a negligible amount of recombinant protein incubated with the extract. Figure S5 (*Supplementary Material*) shows the ion current relative to the RP nanoLC-NSI-HRMS/MS analysis of M^{pro} digested with trypsin and chymotrypsin, characterized by at least 30 chromatographic peaks, corresponding to peptides whose sequence was determined by MS/MS analysis performed in data dependent scan mode. All the peptides containing the targeted residues which were simulated by *in silico* digestion were confirmed experimentally. The sequence coverage of recombinant M^{pro} was close to 99% (Figure S5, *Supplementary Material*), thus confirming the efficiency of the optimized experimental protocol, and in particular highlighting the advantages of using S-TRAP™ spin columns, which maximize the digestion yield and peptide recovery.

The targeted analysis consisted of including in the peptide search algorithm, as variable modifications, the mass shifts of adducts as determined by the metabolomic approach. Since baicalin was found as an adducted moiety, the mass shifts of 444.0692586 Da and 442.0536094 Da (Figure 1) referred to the MA adduct of the phenol and quinoid forms were considered as variable modifications of the following nucleophilic moieties: Cys, Lys and His. Although baicalein did not produce any adducts during the first stage (metabolomics), at this step we also considered potential Michael adducts with the aglycone (Δm: 267.0293472) on the same nucleophilic sites; two additional mass shifts were included as variable modifications, relative to a rearrangement of the MA produced and to the corresponding dehydrated form (Δm: 265.013698 Da and Δm: 249.018783 Da). The structures of the putative reaction products are shown in Figure 1.

Since PD software allows searches against a restricted set of variable modifications (6) for each SEQUEST node, a processing workflow consisting of sequential 2 search algorithms linked together in a hierarchical pattern was created; this enabled the search for all mass variations to be considered in one single PD analysis. To minimize the number of false positives, the mass data recorded were reprocessed by means of the Decoy database, in which the protein sequences are inverted and randomized. This operation allows the calculation of the False discovery rate (FDR) for each match, thus excluding all proteins out of the range of FDR Strict set to 0.01 and FDR Relaxed set to 0.05.

To improve the quality of matching, post-analysis filters were also applied: only those PSMs associated with an XCorr value greater than or equal to 2.5 were considered true identifications. The MS/MS spectra of the adducted peptides were further manually inspected: a modified peptide was confirmed only if the fragmentation spectra showed *b* and/or *y* fragments neighboring the modified amino acid residue both at the N- and C-termini.

This approach permitted the identification of adducts between baicalin and Cys145, His41, and His163/164. Cys145 and His41, constitute the catalytic dyad, and the conserved residues His163/164, exposed to the binding pocket, are involved in a complex network of interactions mediating polar contacts between a water molecule (named H_2O_{cat}) and the catalytic His41 [20]. Figure S6 (*Supplementary Material*) shows the MS/MS spectrum of the ion at *m/z* 838.3085 relative to the $[M+2H]^{2+}$ of the peptide GSFLNGSC*GSVGF where the Cys residue (Cys145) is adducted as Michael adduct by baicalin. The MS/MS spectrum is characterized by almost all the fragment ions of the *y*- and *b*-series with an XCorr value of 4.02; as further confirmation of the attribution, the immonium ion at *m/z* 520.09081, corresponding to the cysteine-Baicalin MA adduct, was detected.

Targeted analysis also identified the peptide H*VICTSEDMLNPNYEDLLIR with the His41 residue adducted with baicalin and baicalein as Michael adduct *Baicalin_MA/Baicalein_MA_R*; as an example in Figure S7 (*Supplementary Material*) is reported the MS/MS spectrum of the precursor ion $[M+3H]^{3+}$ at *m/z* 940.7389 corresponding to the *Baicalin_MA:* in this case the fragmentation pattern is characterized by the majority of the fragment ions of the *b* and *y* series confirming the presence of the adduct with baicalin on the catalytic His41, with an XCorr value of 4.22.

Baicalein adducts have also been found to involve His163 and/or His164: the electrophilic species forms an MA that undergoes de-hydration in the ion source. The MS/MS cannot indicate whether there is a mixture of peptides containing both His163 and Hs164 as adducted residues or if one aa is more targeted than the other. In Figure S8 (*Supplementary Material*) is reported as an example the fragmentation spectrum of the precursor ion $[M+2H]^{2+}$ at *m/z* 1332.0648, corresponding to the peptide with sequence MH*H*MELPTGVHAGTDLEGNFY, in which the marked histidine residue indicates His163 as modified with baicalein by forming the adduct *Baicalein_MA-H_2O*.

Overall, targeted data processing allowed the identification of a series of covalent adducts with baicalin/baicalein involving specific M^{pro} nucleophilic amino acid residues, namely Cys145, His41 and His163/164, to be considered as potential drug targets of molecules with inhibitory activity (covalent inhibitors). Table 4 gives an overview of the identified adducts, for each of which the aminoacidic residue involved and the number of peptide spectrum matches (PSMs) are listed on the basis of incubation time and concentration of the extract tested. Specifically, the number of PSMs indicates the total number of fragmentation spectra identifying specific peptide sequences (native and/or modified), and is proportional to the peptide content as peptides can be fragmented and acquired several times during an LC-HRMS analysis depending on their abundance.

Incubation: 2h, E.C.: 165 µg/µL		
Coverage: 97%; Peptides: 127; PSM: 9018		
Adduct	***AA Residues***	***#PSMs***
Baicalein_MA-H_2O	His163/164	175
Baicalein_MA_R	His163/164	22
Incubation: 2h, E.C.: 330 µg/µL		
Coverage: 99%; Peptides: 131; PSM: 7771		
Adduct	***AA Residues***	***#PSMs***
Baicalein_MA_R	His41	156
Baicalin_MA	His41	10
Baicalein_MA-H_2O	His163/164	113
Baicalein_MA_R	His163/164	26
Baicalein_MA	His163/164	17
Incubation: 4h, E.C.: 165 µg/µL		
Coverage: 99%; Peptides: 129; PSM: 7202		
Adduct	***AA Residues***	***#PSMs***
Baicalein_MA-H_2O	His163/164	91
Baicalein_MA_R	His163/164	60
Baicalein_MA	His163/164	25
Incubation: 4h, E.C.: 330 µg/µL		
Coverage: 97%; Peptides: 125; PSM: 12036		
Adduct	***AA Residues***	***#PSMs***
Baicalin_MA_R	Cys 145	2
Baicalein_MA_R	His41	31
Baicalin_MA_R	His41	4
Baicalin_MA	His41	15
Baicalein_MA-H_2O	His163/164	332
Baicalein_MA_R	His163/164	38
Baicalein_MA	His163/164	55
Incubation: 12h, E.C.: 165 µg/µL		
Coverage: 99%; Peptides: 137; PSM: 3449		
Adduct	***AA Residues***	***#PSMs***
Baicalein_MA_R	His163/164	31
Baicalein_MA-H_2O	His163/164	32
Baicalin_MA_R	Cys 145	23
Incubation: 12h, E.C.: 330 µg/µL		
Coverage: 97%; Peptides: 154; PSM: 7357		
Adduct	***AA Residues***	***#PSMs***
Baicalin_MA	Cys 145	23
Baicalin_MA_R	Cys 145	34
Baicalein_MA	His163/164	32
Baicalein_MA_R	His163/164	145
Baicalein_MA-H_2O	His163/164	91

Baicalin_MA_R	His41	130
Baicalin_MA	His41	96
Baicalein_MA_R	His41	18
Baicalein_MA-H_2O	His41	11

Tabella 4. List of M^{pro} adducts with baicalin/baicalein identified by RP-nanoLC-NSI-HRMS/MS analysis; for each adduct, the amino acid residue involved and the number of PSMs returned by Proteome Discoverer analysis are listed by incubation time and concentration of extract tested (E.C.).

M^{pro} incubation with pure baicalin (oxidative / non-oxidative conditions)

In the proteomic experiment relating to the incubation of M^{pro} with standard baicalin, unlike in the case of incubation with cysteine alone, the oxidation boosting by the Fenton reaction did not greatly improve/accelerate the reaction kinetics except for a few PSMs (*data not shown*).

3.3.2. U*ntargeted* approach

To confirm, and eventually broaden, the spectrum of covalent adducts identified by the targeted approach we then used a proteomic untargeted approach which was applied to the peptide containing Cys145 which is the most druggable nucleophilic site of M^{pro} so far reported.

For this analysis, nanoLC-HRMS raw data were acquired and reprocessed through an untargeted approach as reported in the Materials and Methods section. Ion maps were generated on the basis of the fragment ions of the peptide GSFLNGSC_{145}GSVGF selected since it does not contain Cys145 thus enabling the identification of variable modifications of Cys145. Adducts were identified on the basis of the mass shift between the precursor ion identified by the untargeted approach and the native peptide with the results summarized in Table 5.

Precursor Ion m/z	***z***	***m/z***	***M (Da)***	***$M_{unknown}$ - M_{native}***	***Species Description***
616.27625	2		1231.541173		*Native Peptide*
600.28	2	600.28644	1199.565055	-31.9761176	-O_2
632.27	2	632.26654	1263.525255	31.9840824	+2 O su Cys (*Sulfinic*)
640.27	2	640.26617	1279.524515	47.9833424	+3 O su Cys (*Sulfonic*)
644.78	2	644.78285	1288.557875	57.0167024	Cys *Carbamidomethylation*
645.27	2	645.27481	1289.541795	58.0006224	Cys *Carbamidomethylation* / Asn *Deamidation*
648.26	2	648.25555	1295.503275	63.9621024	Cys-SO_2-SH (T*hiosulfonic acid*)
656.26	2	656.26538	1311.522935	79.9817624	Ser *Phosphorylation*

837.30	**2**	**837.30035**	**1673.592875**	**442.0517024**	***Baicalin_MA_R on Cys145***

Table 5. Table summarizing the *m/z*, charge (z) and the molecular weight values M (Da) of those parent ions selected through the untargeted data processing. The mass shifts calculated from the mass of the native peptide are also reported, along with their description.

Table 5 represents an overview of the post-translational modifications/adduct attributable to the Cys145 residue, without diversifying identifications depending on the experimental conditions tested (incubation time and concentration of extract). Overall, the untargeted analysis (i) confirmed the presence of the native peptide with cysteine residue properly carbamidomethylated, (ii) revealed the presence of multiple oxidation states of the cysteine-thiol group (+2 O on Cys, sulfinic acid; +3 O on Cys, sulfonic acid; Cys-SO_2-SH, thiosulfonic acid), and (iii) detected deamidation of the asparagine residue and (iv) phosphorylation on the serine residue, but did not significantly expand the spectrum of adducts with bioactive molecules in the extract; only the presence of a Michael adduct involving baicalin (*Baicalin_MA_R*) on the catalytic cysteine was confirmed.

Overall, having demonstrated that there is a good overlap of results between the targeted and untargeted approaches, it can certainly be stated that this investigation procedure may represent a breakthrough in the search for a priori unknown mass shifts that may expand our knowledge of the reactivity of potential covalent ligands acting as inhibitors for the M^{pro} viral protease.

3.4. Molecular modeling studies

On the basis of the of the number of PSMs as an index of adduct abundancy as reported in Table 4, some considerations can be drawn for the reaction mechanism of baicalein. Firstly, adduct formation was found to be dose-dependent and, for some adducts, also time-dependent. Then, by clustering the adducts on the basis of the aa site of modification, His163/164 aa modifications occur at earlier incubation times and with a great abundancy while Cys adduct formation occurs at longer incubation times and in a limited amount. The data suggested a greater reactivity of the His163/164 residues towards baicalin/baicalein which could be due to a better engagement of the molecules in the cavity where the His residues are exposed. The data also shed some light on the different reactivities of baicalein and baicalin. After 2 hours of incubation only the aglycone baicalein was found to form covalent adducts with the residues His163/164 and His41 suggesting that it is presumably more reactive than the corresponding glucuronide, given the reduced steric hindrance which allows easier access to the active site of the protease. To better explain the reactivity of baicalin and baicalein and the engagement of different nucleophilic residues, MD simulations of the SARS-CoV-2 M^{pro} in complex with these molecules were carried out.

The crystal structure of SARS-CoV-2 M^{pro} in complex with baicalein (PDB ID 6M2N) [38] was used as starting coordinates while the complex involving the glycoside was generated by molecular docking by

using the Gold suite [43]. As shown in Figure 7A the X-ray structure revealed that baicalein occupies the M^{pro} binding site by establishing a network of H-bonds involving i) the hydroxyl groups of the pyrogallol portion and the residues Leu141, Gly143 and Ser144, and ii) the carbonyl group of the pyranone moiety and the backbone of Glu166. The phenyl group is engaged in π- π stacking interaction with His41 and in hydrophobic contact with Met49, Cys44, His41, Gln189 and Arg188. Finally, the catalytic residue Cys145 elicits π-S interaction with the pyrogallol ring. Even though the resolved complex does not involve covalent adducts, the electrophilic site of baicalein faces towards the thiol group of Cys145 and the Nδ of His41 assuming distances potentially conducive to a nucleophilic attack in the appropriate conditions. In contrast, the side chain of His163 is located at ~7Å from the pyrogallol moiety, while the imidazole ring of His164 faces the opposite side in respect to the ligand position.

Concerning baicalin, the docking simulation suggested that the glycoside might bind the M^{pro} active site by forming H-bonds with the backbone of His164 and the side chains of Glu166 and Gln189, through the glucuronide moiety, while the benzopyranone portion could establish π-π stacking and π-S interaction with His41 and Cys44, respectively (Figure 7B). Hydrophobic contacts were observed between the phenyl ring and Met49. Taken together, the complexes depicted in Figure 7 can rationalize the expected reactivity of baicalein towards the catalytic dyad His41/Cys145, while baicalin appears to be more distant to the catalytic residues, due to the shielding effect exerted by the glycoside ring.

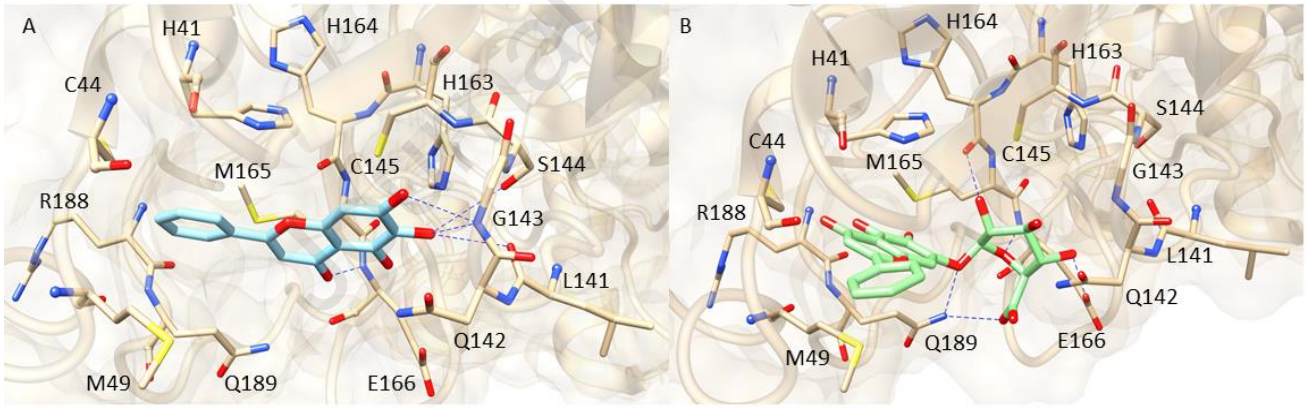


Figure 7. A) Experimentally resolved binding pose of baicalein (cyan sticks) within the catalytic site of SARS-CoV-2 M^{pro} (PDB ID 6M2N). B) Plausible binding mode of baicalin (green sticks) docked into M^{pro} binding pocket. The residues of the active site are represented as beige sticks while H-bonds are displayed as blue dashed lines.

By considering the well-known flexibility of the SARS-CoV-2 M^{pro} binding site, the two complexes underwent 500 ns of MD simulation by using the software Amber18 to explore their dynamic behavior and to obtain more insight into the role of the nucleophilic residues involved in the formation of the covalent adducts. The stability of the two systems during the simulations was checked by monitoring the RMSD of both protein and ligand. As shown in Figure S9 (*Supplementary Material*), both ligands exhibited a stable behavior during the simulation time. However, when in complex with baicalein (Figure S9A), the protein showed a greater flexibility compared to baicalin (Figure S9B).

Cluster analysis was carried out to obtain representative conformations of the systems. The representative structures of the most populated clusters of both systems are depicted in Figure 8.

The MD output revealed that baicalein was able to maintain the crucial interactions with the residues His41, Cys145, Ser144, Ile141, Glu166 and Met49 (Figure 8A). Moreover, additional H-bonds were observed between (i) the pyrogallol ring and the residues His163 and Phe140, and (ii) the carbonyl group of the benzopyranone moiety and Gln142. As shown in Figure 8A, , baicalein moved during the simulation from its original position to conveniently approach His163, while remaining close to the catalytic Cys145. The distances recorded during the trajectory between the Nε of His163 and the electrophilic carbon of baicalein were in a range between 3.11 Å and 8.35 Å. Similar values were registered between the sulfur atom of Cys145 and the electrophilic carbon of the ligand (3.13 Å-10.96 Å). These results suggested that baicalein is able to adopt suitable distances for a nucleophilic attack by both His163 and Cys145, thus supporting the formation of the detected covalent adducts in an oxidative environment.

Concerning baicalin, during the MD simulation the glycoside maintained its interactions with His164 and Met49 as shown in Figure 8B. New H-bonds were detected between i) Asp187 and the pyrogallol moiety, ii) Asn142 and the hydroxyl groups of the sugar portion, and iii) Cys44 and the carbonyl group of baicalin. Moreover, the side chain of His41 approached the pyrogallol ring, with the minimum distance registered between the Nε of His41 and the electrophilic carbon of baicalin being 3.19 Å, which is conducive to the Michael addiction in the appropriate conditions. Cys145 shows only limited movements towards the pyrogallol ring of baicalin (its distance ranges from 5.6 Å to 7.4 Å during the MD run); however, in an oxidative milieu structural rearrangements in the M^{pro} active site might further favor the exposure of these nucleophilic residues to the electrophilic warhead. Although the side chain of His164 moved towards the ligand during MD simulation, His163 and His164 remain too far from the pyrogallol ring due to the steric hindrance elicited by the glycoside and this is can explain why the corresponding covalent adducts are not detected with baicalin.

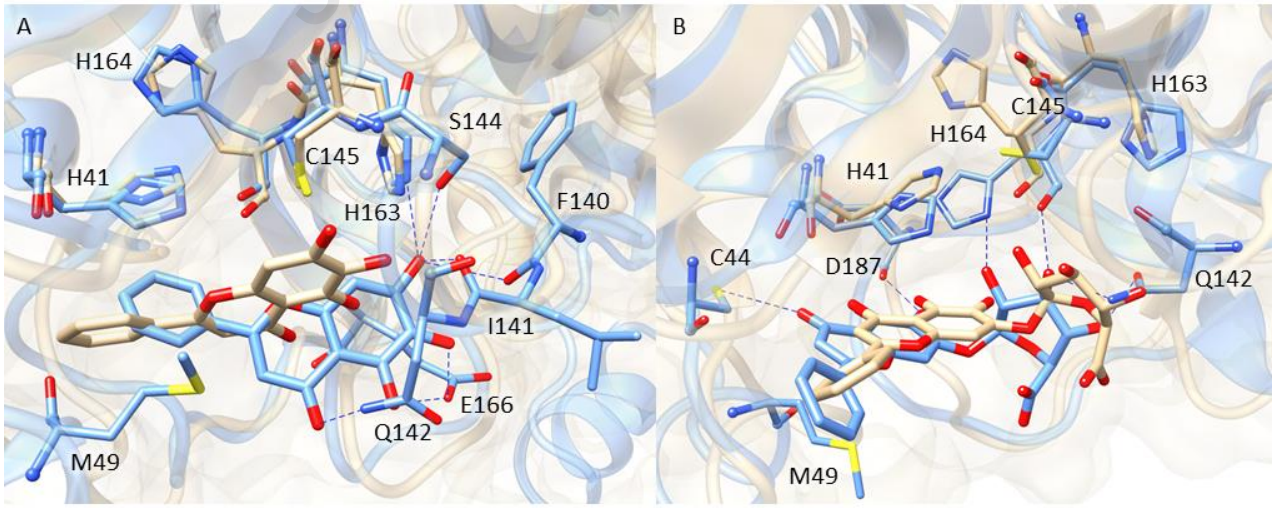


Figure 8. A) Representative structure of the most populated cluster obtained from the MD simulation of baicalein-M^{pro} complex (blue) superimposed to the X-ray structure (beige). B) Representative structure of the most populated cluster obtained from the MD simulation of baicalin-M^{pro} complex (blue) superimposed to the starting coordinates gained from the docking studies (beige).

4. Discussion

The aim of the present paper was to set-up an analytical strategy for the identification of molecules contained in complex matrices, such as natural extracts, able to bind and covalently adduct the catalytic sites of M^{pro} as a target protein. Currently, M^{pro} represents one of the most promising drug targets for SARS-Cov-2, as it plays a crucial role in the maturation of viral polyproteins into functional proteins, which are essential for the completion of the virus replication cycle. Furthermore, targeting M^{pro} would be advantageous for two reasons: (i) the absence of its homologues in human cells; (ii) its high cleavage specificity, which means that all molecules structurally similar to its cleavage sites can be considered potential inhibitors, with little or no impact on host cell proteases.

Numerous high-throughput screening and virtual screening campaigns have identified several natural compounds as M^{pro} inhibitors, indicating that plant matrices are a rich source of bioactive molecules. Most of the *in vitro* and computational models used for inhibitor screening, are based on the use of databases and chemical libraries, consisting of known and isolated natural molecules. If on one hand HTS approaches can be advantageous in terms of reducing costs and study time, on the other they have a limited explorative potential, confined to molecules structurally defined and available as pure compounds in libraries.

Extracts from natural sources represent a cheap source of potential bioactive compounds. However, the use of crude extracts in the well-established phenotypic cell models or target based models is limited due to the fact that even in the presence of a very active molecule, if it is contained in small amounts in respect to the other constituents, its activity is not detectable and where there is an effect, the identity of the active constituent cannot be determined. This is a great limitation to the discovery of hit compounds also taking into account that many bioactive molecules so far reported as effective antiviral agents are natural compounds.

Hence, we believe that an analytical method able to fish-out bioactive molecules from natural complex matrices with poor or even no information at all on the constituents would be a powerful and innovative tool for the discovery of novel hit compounds such as M^{pro} binders and inhibitors. In light of this, the present work is aimed at optimizing a high-resolution mass spectrometry (HR-MS) based analytical platform, which integrates the principles and potential of metabolomics and proteomics, to identify potential covalent inhibitors of M^{pro} in natural extracts.

The method is based on the search for compounds which covalently react with the active sites of M^{pro}. Most of the inhibitors so far reported for M^{pro} are recognized by the catalytical domain and then the complex is stabilized by a covalent adduction to Cys145. Carmofur, has been shown to inhibit SARS-CoV-2 replication in cells by covalently modifying the catalytic Cys145 of M^{pro}. An oral antiviral compound PF-07321332 from Pfizer, specifically designed to inhibit SARS-CoV-2 M^{pro} modifies the active site Cys145 with its nitrile warhead, is considered a good candidate antiviral and is currently undergoing trials (NCT04756531, NCT04909853, NCT05011513, Clinical-Trials.gov). Very recently, the X-ray crystal structure of M^{pro} in

complex with MG-101 has shown a covalent bond formation between the inhibitor and the active site Cys145 residue, indicating that its mechanism of inhibition is blocking the substrate binding at the active site. Even natural compounds have been reported to inhibit M^{pro} by covalent binding. Compounds containing the pyrogallol group are of interest since this group can act as the warhead that could covalently link to cysteine under oxidative conditions. Myricetin and derivatives have recently been reported as selective covalent inhibitors of cysteine proteinase by using the pyrogallol group as a warhead and the chromone as the reversible binding portion. In particular, the oxidized myricetin is first recognized by the catalytic site in which the specific side-chain conformation of His41 is prone to forming the π–π stacking interactions with the chromone ring, which is followed by the covalent reaction of the pyrogallol moiety with Cys145. Other compounds, even those containing a pyrogallol moiety, act through a non-covalent engagement as in the case of baicalin and baicalein. The orientation of myricetin at the binding site is different from that of baicalein, resulting in distinct ligand-protein interaction patterns. In comparison with myricetin, baicalein forms more H-bonding and hydrophobic interactions with the residues. Notably, the pyrogallol group of baicalein forms multiple H-bonds with main chains of Leu141/Gly143 as well as the side chain of Ser144, fixing the conformation of the oxyanion loop (residues 138–145) which serves to stabilize the tetrahedral transition state of the proteolytic reaction, whereas the pyrogallol group of myricetin acts as an electrophile which covalently binds to Cys145.

The method here proposed was set-up to identify, from natural extracts, not only the covalent inhibitors of M^{pro}, but also the non-covalent inhibitors which in certain conditions are forced to form stable adducts, such as baicalein and baicalin. In other words, the method proposed was designed to detect those compounds which are recognized by a portion of the molecule forming reversible engagement, a complex which is then stabilized by the warhead. The method was tuned also to detect pyrogallol/catechol non-covalent inhibitors by forcing their covalent adduction for example by using long incubation times (up to 12 hours) or oxidizing conditions.

The method was set-up and simultaneously validated by using *S. baicalensis* extract, a valuable plant used in traditional Chinese medicine for the treatment and prophylaxis of various diseases and which contains several components, among which baicalin and baicalein, which are well-established pyrogallol non-covalent inhibitors of M^{pro}.

The proposed method consists of a targeted and an untargeted approach which can be used independently or combined. Targeted analysis is firstly based on a metabolomic approach aimed at identifying the MW of the natural extract constituents and at finding which of them are able to covalently react with a soft nucleophilic substrate, i.e. Cys. Basically, Cys rather than GSH was used as nucleophilic substrate since the former is more acidic than the latter (pKa of Cys and GSH are 8.30 and 8.83, respectively) and is therefore more reactive in the Michael addition reaction. In the case that very weak electrophilic compounds need to be screened, Cysteamine characterized by a pKa of 8.19 could be used or the pH of the incubation mixture could be set at basic values to further convert the thiol to the thiolate form. The electrophilic constituents identified through the metabolomic analysis were set as variable modifications in the proteomic analysis aimed at

identifying adducts to the target nucleophilic sites of Mpro, not only limited to Cys 145 but also to His41, and His 163 or His 164.

This approach easily identified among the components of natural extract, baicalin and baicalein which are known as non-covalent binders despite the presence of a pyrogallol moiety. However, by modulating the experimental conditions, baicalin and baicalein were forced to form covalent adducts. The method also permitted the identification of the aa involved, mainly His163 / 164 and His41 and Cys145, the latter only after a long incubation time. Hence, the present method, besides making it easy to fish-out covalent binders of M^{pro}, also permits the identification of the nucleophilic sites involved, providing information on the mechanism of action of the inhibitor. Currently we do not know whether, in *in vivo* conditions, baicalein and baicalin, having interacted with M^{pro} by a non-covalent engagement and been exposed to an oxidative stress environment (such as that induced by an inflammatory response which activates the electrophilicity of the pyrogallol ring by forming the quinone intermediate), form the covalent bond as here reported. Further analysis needs to be done by searching for the adducted peptide of M^{pro} with baicalein/baicalin in *ex vivo* conditions.

We then proposed an untargeted approach permitting the identification without any prior knowledge of the adducted peptides. The method is based on the selection of the target aa and the proteolytic peptide containing it, such as Cys145 and the peptide GSFLNGSC_{145}GSVGF. By calculating the difference of the MW between the adducted and the corresponding native peptide, the MW of the covalent binder is determined. Finally, using a metabolomic approach, the component of the natural extract identified on the basis of the MW is characterized on the basis of elemental composition, isotopic pattern and MS/MS fragmentation. The method easily identified the Michael adduct between baicalin and Cys145 in a quinoid form.

The binding mechanism of baicalein and baicalin within the M^{pro} active site was elucidated by MD simulations, allowing the clarification of the role of the nucleophilic residues involved in the formation of the covalent adducts in the protein-ligand recognition process. The results showed that baicalein might conveniently approach Cys145 and His163 through its pyrogallol ring, assuming distances conducive to a nucleophilic attack. Concerning baicalin, during the simulation the side chains of His41 drew near the electrophilic warhead of the ligand, while His163 and His164, while showing some conformational fluctuations, remain to a distance not conducive to a nucleophilic attack, reasonably due to the steric hindrance caused by the presence of the sugar moiety. Altogether, the obtained results confirms the marked flexibility of the M^{pro} binding cavity where small fluctuations of few side chains have remarkable effects on its interaction capacities. Notably, the reported results evidence that such a flexibility plays a key role which goes far beyond the mutual adaptability between enzyme and inhibitor since it can modulate the reactivity/exposure of the surrounding nucleophilic side chains thus involving unexpected residues such as His163 and His164.

In conclusion, the method here proposed represents a suitable tool to screen natural extracts, in particular those containing compounds with catechol or pyrogallol moieties which are recognized as potential warheads and which, under certain conditions, can react covalently with the activated nucleophilic sites of M^{pro}. The suitable compound to be screened would contain a moiety highly selective for the catalytic site of

740 M^{pro} and with a warhead which should be activated or becomes more reactive when binding the catalytical site
741 and/or when present in an oxidizing environment such as that occurring at the site of virus proliferation.
742 Clearly, a rational-drug design approach would be suitable but as an alternative or in parallel to it, with the aim
743 of finding hit compounds, the present analytical approach can be used to screen extracts containing a chemical
744 variety of compounds containing pyrogallol/catechol moieties. Many natural and safe extracts containing such
745 a class of compound covering a wide chemical space of potential derivatives are known, including for example
746 berries rich in glycosides and acetyl glycosides.

747

AUTHOR CONTRIBUTIONS: Conceptualization, AA, GB and GA; Formal analysis, AA, GB, SB, LDV, and SV; Funding acquisition, MC and GA; Investigation, AA, GB, SV, GA and GV; Methodology, AA, SV, and GB; Project administration, AA, GV and GA; Resources, MC and GA; Supervision, AA, GB, GV and GA; Validation, AA, SV and GB; Writing – original draft, AA, SV and GA; Writing – review & editing, AA, GB, GV and GA.

ACKNOWLEDGMENTS:

LDV is supported as Ph.D student by Ministero dell'Università e della Ricerca PON "Ricerca e Innovazione" 2014-2020, Azione IV.4 – "Dottorati e contratti di ricerca su tematiche dell'innovazione" and Azione IV.6 – "Contratti di ricerca su tematiche Green" .

This research is part of the project "MIND FoodS HUB (Milano Innovation District Food System Hub): Innovative concept for the eco-intensification of agricultural production and for the promotion of dietary patterns for human health and longevity through the creation in MIND of a digital Food System Hub", cofunded by POR FESR 2014-2020_BANDO Call HUB Ricerca e Innovazione, Regione Lombardia.

CONFLICT OF INTERESTS: The authors declare that no competing interests exist.

DATA AVAILABILITY STATEMENT: Data available on request from the authors.

ORCID:

Giovanna Baron https://orcid.org/0000-0002-9335-6318

Alessandra Altomare https://orcid.org/0000-0002-9906-6098

SUPPORTING INFORMATION: Additional supporting information may be found in the online version of the article at the publisher's website.

References

[1] A.C. Kalil, Treating COVID-19-Off-Label Drug Use, Compassionate Use, and Randomized Clinical Trials During Pandemics, JAMA 323(19) (2020) 1897-1898. https://doi.org/10.1001/jama.2020.4742.

[2] J.M. Sanders, M.L. Monogue, T.Z. Jodlowski, J.B. Cutrell, Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review, JAMA 323(18) (2020) 1824-1836. https://doi.org/10.1001/jama.2020.6019.

[3] R. Poduri, G. Joshi, G. Jagadeesh, Drugs targeting various stages of the SARS-CoV-2 life cycle: Exploring promising drugs for the treatment of Covid-19, Cell Signal 74 (2020) 109721. https://doi.org/10.1016/j.cellsig.2020.109721.

[4] D. Batlle, J. Wysocki, K. Satchell, Soluble angiotensin-converting enzyme 2: a potential approach for coronavirus infection therapy?, Clin Sci (Lond) 134(5) (2020) 543-545. https://doi.org/10.1042/CS20200163.

[5] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for SARS-CoV-2, Nat Rev Microbiol 19(3) (2021) 155-170. https://doi.org/10.1038/s41579-020-00468-6.

[6] Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L.W. Guddat, Q. Wang, Z. Lou, Z. Rao, Structure of the RNA-dependent RNA polymerase from COVID-19 virus, Science 368(6492) (2020) 779-782. https://doi.org/10.1126/science.abb7498.

[7] C.J. Gordon, E.P. Tchesnokov, E. Woolner, J.K. Perry, J.Y. Feng, D.P. Porter, M. Götte, Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency, J Biol Chem 295(20) (2020) 6785-6797. https://doi.org/10.1074/jbc.RA120.013679.

[8] A. Shannon, B. Selisko, N. Le, J. Huchting, F. Touret, G. Piorkowski, V. Fattorini, F. Ferron, E. Decroly, C. Meier, B. Coutard, O. Peersen, B. Canard, Favipiravir strikes the SARS-CoV-2 at its Achilles heel, the RNA polymerase, bioRxiv (2020). https://doi.org/10.1101/2020.05.15.098731.

[9] A. Shulla, T. Heald-Sargent, G. Subramanya, J. Zhao, S. Perlman, T. Gallagher, A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry, J Virol 85(2) (2011) 873-82. https://doi.org/10.1128/JVI.02062-10.

[10] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.H. Wu, A. Nitsche, M.A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, Cell 181(2) (2020) 271-280.e8.

[11] M.D. Sacco, C. Ma, P. Lagarias, A. Gao, J.A. Townsend, X. Meng, P. Dube, X. Zhang, Y. Hu, N. Kitamura, B. Hurst, B. Tarbet, M.T. Marty, A. Kolocouris, Y. Xiang, Y. Chen, J. Wang, Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against M, Sci Adv 6(50) (2020). https://doi.org/10.1126/sciadv.abe0751.

[12] T.J. El-Baba, C.A. Lutomski, A.L. Kantsadi, T.R. Malla, T. John, V. Mikhailov, J.R. Bolla, C.J. Schofield, N. Zitzmann, I. Vakonakis, C.V. Robinson, Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays*, Angew Chem Int Ed Engl 59(52) (2020) 23544-23548. https://doi.org/10.1002/anie.202010316.

[13] C. Ma, M.D. Sacco, B. Hurst, J.A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M.T. Marty, Y. Chen, J. Wang, Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease, Cell Res 30(8) (2020) 678-692. https://doi.org/10.1038/s41422-020-0356-z.

[14] Y.W. Chen, C.B. Yiu, K.Y. Wong, Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL, F1000Res 9 (2020) 129. https://doi.org/10.12688/f1000research.22457.2.

[15] X. Xue, H. Yu, H. Yang, F. Xue, Z. Wu, W. Shen, J. Li, Z. Zhou, Y. Ding, Q. Zhao, X.C. Zhang, M. Liao, M. Bartlam, Z. Rao, Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design, J Virol 82(5) (2008) 2515-27. https://doi.org/10.1128/JVI.02114-07.

[16] O. Sheik Amamuddy, G.M. Verkhivker, Ö. Tastan Bishop, Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 M, J Chem Inf Model 60(10) (2020) 5080-5102. https://doi.org/10.1021/acs.jcim.0c00634.

[17] J. Ziebuhr, E.J. Snijder, A.E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the Nidovirales, J Gen Virol 81(Pt 4) (2000) 853-79. https://doi.org/10.1099/0022-1317-81-4-853.

[18] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, Design of wide-spectrum inhibitors targeting coronavirus main proteases, PLoS Biol 3(10) (2005) e324. https://doi.org/10.1371/journal.pbio.0030324.

[19] M. Mei, X. Tan, Current Strategies of Antiviral Drug Discovery for COVID-19, Front Mol Biosci 8 (2021) 671263. https://doi.org/10.3389/fmolb.2021.671263.
[20] D.W. Kneller, G. Phillips, H.M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates, A. Kovalevsky, Structural plasticity of SARS-CoV-2 3CL M pro active site cavity revealed by room temperature X-ray crystallography, Nat Commun 11(1) (2020) 3202. https://doi.org/10.1038/s41467-020-16954-7.
[21] H.M. Mengist, T. Dilnessa, T. Jin, Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease, Front Chem 9 (2021) 622898. https://doi.org/10.3389/fchem.2021.622898.
[22] M. Hassanzadeganroudsari, A.H. Ahmadi, N. Rashidi, M.K. Hossain, A. Habib, V. Apostolopoulos, Computational Chemistry to Repurposing Drugs for the Control of COVID-19, Biologics 1 (2021) 111-128. https://doi.org/10.3390/biologics1020007.
[23] V.R. Naik, M. Munikumar, U. Ramakrishna, M. Srujana, G. Goudar, P. Naresh, B.N. Kumar, R. Hemalatha, Remdesivir (GS-5734) as a therapeutic option of 2019-nCOV main protease -, J Biomol Struct Dyn 39(13) (2021) 4701-4714. https://doi.org/10.1080/07391102.2020.1781694.
[24] Z. Su, Y. Wu, A Multiscale and Comparative Model for Receptor Binding of 2019 Novel Coronavirus and the Implication of its Life Cycle in Host Cells, bioRxiv (2020). https://doi.org/10.1101/2020.02.20.958272.
[25] G. Baron, A. Altomare, M. Mol, J.L. Garcia, C. Correa, A. Raucci, L. Mancinelli, S. Mazzotta, L. Fumagalli, G. Trunfio, L. Tucci, E. Lombardo, D. Malara, E. Janda, V. Mollace, M. Carini, E. Bombardelli, G. Aldini, Analytical Profile and Antioxidant and Anti-Inflammatory Activities of the Enriched Polyphenol Fractions Isolated from Bergamot Fruit and Leave, Antioxidants (Basel) 10(2) (2021). https://doi.org/10.3390/antiox10020141.
[26] L. Hu, Y. Xiong, Z. Zou, Y. Yang, J. He, L. Zhong, Y. Wang, M. Yang, Identifying the chemical markers in raw and wine-processed Scutellaria baicalensis by ultra-performance liquid chromatography/quadrupole time-of-flight mass spectrometry coupled with multiple statistical strategies, Biomed Chromatogr 34(8) (2020) e4849. https://doi.org/10.1002/bmc.4849.
[27] T. Zhao, H. Tang, L. Xie, Y. Zheng, Z. Ma, Q. Sun, X. Li, Scutellaria baicalensis Georgi. (Lamiaceae): a review of its traditional uses, botany, phytochemistry, pharmacology and toxicology, J Pharm Pharmacol 71(9) (2019) 1353-1369. https://doi.org/10.1111/jphp.13129.
[28] Y.-Y. Zhang, Y.-Z. Guo, M. Onda, K. Hashimoto, Y. Ikeya, M. Okada, M. Maruno, Four flavonoids from S*cutellaria baicalensis*, Phytochemistry 35(2) (1994) 511-514. https://doi.org/10.1016/S0031-9422(00)94792-7.[29] T. Tomimori, H. Jin, Y. Miyaichi, S. Toyofuku, T. Namba, [Studies on the constituents of Scutellaria species. VI. On the flavonoid constituents of the root of Scutellaria baicalensis Georgi (5). Quantitative analysis of flavonoids in Scutellaria roots by high-performance liquid chromatography], Yakugaku Zasshi 105(2) (1985) 148-55. https://doi.org/10.1248/yakushi1947.105.2_148.
[30] T. Tomimori, Y. Miyaichi, H. Kizu, [On the flavonoid constituents from the roots of Scutellaria baicalensis Georgi. I], Yakugaku Zasshi 102(4) (1982) 388-91. https://doi.org/10.1248/yakushi1947.102.4_388.
[31] M. Takido, M. Aimi, S. Yamanouchi, K. Yasukawa, H. Torii, S. Takahashi, [Studies on the constituents in the water extracts of crude drugs. II. On the leaves of scutellaria baicalensis Georgi (author's transl)], Yakugaku Zasshi 96(3) (1976) 381-3. https://doi.org/10.1248/yakushi1947.96.3_381.
[32] T. Tomimori, Y. Miyaichi, Y. Imoto, H. Kizu, Y. Tanabe, [Studies on the constituents of Scutellaria species. III. On the flavonoid constituents of the root of Scutellaria baicalensis Georgi (3)], Yakugaku Zasshi 104(5) (1984) 524-8. https://doi.org/10.1248/yakushi1947.104.5_524.
[33] H.L. Long, G.Y. Xu, A.J. Deng, Z.H. Li, L. Ma, Y. Lu, Z.H. Zhang, F. Wu, H.L. Qin, Two new flavonoids from the roots of Scutellaria baicalensis, J Asian Nat Prod Res 17(7) (2015) 756-60. https://doi.org/10.1080/10286020.2014.999048.
[34] M. Kubo, H. Matsuda, M. Tanaka, Y. Kimura, H. Okuda, M. Higashino, T. Tani, K. Namba, S. Arichi, Studies on Scutellariae radix. VII. Anti-arthritic and anti-inflammatory actions of methanolic extract and flavonoid components from Scutellariae radix, Chem Pharm Bull (Tokyo) 32(7) (1984) 2724-9. https://doi.org/10.1248/cpb.32.2724.
[35] S. Takagi, M. Yamaki, K. Inoue, [Studies on the water-soluble constituents of the roots of Scutellaria baicalensis Georgi (Wogon) (author's transl)], Yakugaku Zasshi 100(12) (1980) 1220-4. https://doi.org/10.1248/yakushi1947.100.12_1220.

[36] H.G. Park, S.Y. Yoon, J.Y. Choi, G.S. Lee, J.H. Choi, C.Y. Shin, K.H. Son, Y.S. Lee, W.K. Kim, J.H. Ryu, K.H. Ko, J.H. Cheong, Anticonvulsant effect of wogonin isolated from Scutellaria baicalensis, Eur J Pharmacol 574(2-3) (2007) 112-9. https://doi.org/10.1016/j.ejphar.2007.07.011.
[37] Z.-H. Zhou, Y.-J. Zhang, C.-R. Yang, New flavonoid glycosides from Scutellaria amoena, Studies in Plant Science 6 (1999) 305-310. https://doi.org/10.1016/S0928-3420(99)80040-8.
[38] H.X. Su, S. Yao, W.F. Zhao, M.J. Li, J. Liu, W.J. Shang, H. Xie, C.Q. Ke, H.C. Hu, M.N. Gao, K.Q. Yu, H. Liu, J.S. Shen, W. Tang, L.K. Zhang, G.F. Xiao, L. Ni, D.W. Wang, J.P. Zuo, H.L. Jiang, F. Bai, Y. Wu, Y. Ye, Y.C. Xu, Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients, Acta Pharmacol Sin 41(9) (2020) 1167-1177. https://doi.org/10.1038/s41401-020-0483-6.
[39] A. Pedretti, A. Mazzolari, S. Gervasoni, L. Fumagalli, G. Vistoli, The VEGA suite of programs: an versatile platform for cheminformatics and drug design projects, Bioinformatics 37(8) (2021) 1174-1175. https://doi.org/10.1093/bioinformatics/btaa774.
[40] J.C. Gordon, J.B. Myers, T. Folta, V. Shoja, L.S. Heath, A. Onufriev, H++: a server for estimating pKas and adding missing hydrogens to macromolecules, Nucleic Acids Res 33(Web Server issue) (2005) W368-71. https://doi.org/10.1093/nar/gki464.
[41] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E.E. Bolton, PubChem in 2021: new data content and improved web interfaces, Nucleic Acids Res 49(D1) (2021) D1388-D1395. https://doi.org/10.1093/nar/gkaa971.
[42] J.J. Stewart, Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, J Mol Model 19(1) (2013) 1-32. https://doi.org/10.1007/s00894-012-1667-x.
[43] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J Mol Biol 267(3) (1997) 727-48. https://doi.org/10.1006/jmbi.1996.0897.
[44] S. Vittorio, L. Ielo, S. Mirabile, R. Gitto, A. Fais, S. Floris, A. Rapisarda, M.P. Germanò, L. De Luca, 4-Fluorobenzylpiperazine-Containing Derivatives as Efficient Inhibitors of Mushroom Tyrosinase, ChemMedChem 15(18) (2020) 1757-1764. https://doi.org/10.1002/cmdc.202000125.
[45] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J Comput Chem 26(16) (2005) 1668-88. https://doi.org/10.1002/jcc.20290.
[46] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J Comput Chem 25(9) (2004) 1157-74. https://doi.org/10.1002/jcc.20035.
[47] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, J Mol Graph Model 25(2) (2006) 247-60. https://doi.org/10.1016/j.jmgm.2005.12.005.
[48] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J Chem Theory Comput 11(8) (2015) 3696-713. https://doi.org/10.1021/acs.jctc.5b00255.
[49] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, J Chem Theory Comput 9(7) (2013) 3084-95. https://doi.org/10.1021/ct400341p.
[50] T. Tubiana, J.C. Carvaillo, Y. Boulard, S. Bressanelli, TTClust: A Versatile Molecular Simulation Trajectory Clustering Program with Graphical Summaries, J Chem Inf Model 58(11) (2018) 2178-2182. https://doi.org/10.1021/acs.jcim.8b00512.
[51] A. Rajagopalan, W. Kroutil, Biocatalytic reactions: Selected highlights, Materials Today 14(4) (2011) 144-152. https://doi.org/10.1016/S1369-7021(11)70086-4.

- SARS-CoV-2 main protease (M^{pro}) is a potential target of antiviral drugs;
- The OMICS - based analytical platform (HR-MS) integrates metabolomics and proteomics principles to identify new potential covalent inhibitors of M^{pro} in natural extracts;
- The proposed method consists of a targeted and an untargeted approach, which can be used independently or combined, expanding the explorative potential;
- Compounds with catechol or pyrogallol moieties can bind the catalytic site of M^{pro};

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Giancarlo Aldini reports financial support was provided by Lombardy Region.