



## RESEARCH ARTICLE

# P1' specificity of the S219V/R203G mutant tobacco etch virus protease

Mária Golda<sup>1</sup> | Gyula Hoffka<sup>1,2</sup> | Scott Cherry<sup>3</sup> | Joseph E. Tropea<sup>3</sup> |  
 George T. Lountos<sup>4</sup>  | David S. Waugh<sup>3</sup> | Alexander Wlodawer<sup>3</sup> |  
 József Tózsér<sup>1</sup> | János András Mótyán<sup>1</sup> 

<sup>1</sup>Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary

<sup>2</sup>Doctoral School of Molecular Cell and Immune Biology, University of Debrecen, Debrecen, Hungary

<sup>3</sup>Center for Structural Biology, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, Maryland, USA

<sup>4</sup>Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

## Correspondence

János András Mótyán, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Egyetem Square 1, Life Science Building 3.109, 4032 Debrecen, Hungary.  
 Email: [motyjan.janos@med.unideb.hu](mailto:motyjan.janos@med.unideb.hu)

## Funding information

Magyar Tudományos Akadémia; Nemzeti Kutatási, Fejlesztési és Innovációs Alap; NIH Intramural Research Program; National Research, Development and Innovation Fund (Hungary), Grant/Award Number: TKP2021-EGA-20; Hungarian Academy of Sciences, Grant/Award Number: BO/00110/23/5; National Institutes of Health, Grant/Award Number: 75N91019D00024

## Abstract

Proteases that recognize linear amino acid sequences with high specificity became indispensable tools of recombinant protein technology for the removal of various fusion tags. Due to its stringent sequence specificity, the catalytic domain of the nuclear inclusion cysteine protease of tobacco etch virus (TEV PR) is also a widely applied reagent for enzymatic removal of fusion tags. For this reason, efforts have been made to improve its stability and modify its specificity. For example, P1' autoproteolytic cleavage-resistant mutant (S219V) TEV PR was found not only to be nearly impervious to self-inactivation, but also exhibited greater stability and catalytic efficiency than the wild-type enzyme. An R203G substitution has been reported to further relax the P1' specificity of the enzyme, however, these results were obtained from crude intracellular assays. Until now, there has been no rigorous comparison of the P1' specificity of the S219V and S219V/R203G mutants in vitro, under carefully controlled conditions. Here, we compare the P1' amino acid preferences of these single and double TEV PR mutants. The in vitro analysis was performed by using recombinant protein substrates representing 20 P1' variants of the consensus TENLYFQ\*SGT cleavage site, and synthetic oligopeptide substrates were also applied to study a limited set of the most preferred variants. In addition, the enzyme–substrate interactions were analyzed in silico. The results indicate highly similar P1' preferences for both enzymes, many side-chains can be accommodated by the S1' binding sites, but the kinetic assays revealed lower catalytic efficiency for the S219V/R203G than for the S219V mutant.

## KEYWORDS

enzymology, fusion tag removal, molecular dynamics, protein structure, TEV protease, tobacco etch virus

## 1 | INTRODUCTION

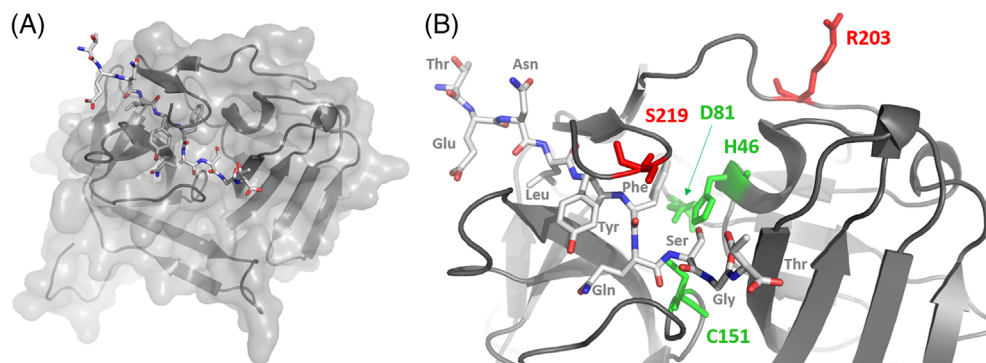
The cysteine protease (PR) of tobacco etch virus (TEV) is among a handful of enzymes that are commonly used in biotechnology and

molecular biology for enzymatic removal of various fusion tags.<sup>1–4</sup>

Due to its stringent sequence specificity, the TEV PR has become one of the most popular reagents for this purpose. The canonical recognition sequence of TEV PR is ENLYFQ\*G/S. A number of

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Proteins: Structure, Function, and Bioinformatics* published by Wiley Periodicals LLC.



**FIGURE 1** Structure of wild-type TEV PR complexed with an oligopeptide substrate. (A) The figure shows the complex of TEV PR and the oligopeptide substrate representing the TENLYFQ\*SGT canonical cleavage site (PDBID: 1LVB).<sup>7</sup> (B) Enlarged view of the active site. The catalytic residues are shown in green while the wild-type Ser219 and Arg203 residues are shown in red.

protein expression vectors that include a TEV PR cleavage site for fusion tag removal are available, such as pDestHis-MBP (MBP: maltose-binding protein)<sup>5</sup> and the pTEV plasmid system.<sup>6</sup>

The TEV PR adopts a two-domain antiparallel  $\beta$ -barrel fold. The substrate binding sites and the catalytic center are located in the interdomain cleft, while His46, Asp81, and Cys151 form the catalytic triad.<sup>7</sup> The Ser219 residue is located close to the C terminus of the PR, in close proximity to the catalytic site and the substrate-binding cleft (Figure 1).

Multiple studies have been performed to design TEV PR variants with altered stability, catalytic activity, and/or specificity.<sup>8–16</sup> Enzyme instability was one of the key drawbacks of the wild-type TEV PR, as the concentration-dependent inactivation of TEV PR via autolysis (occurring between residues Met218 and Ser219 of the catalytic domain) (close to the C terminus of the enzyme) was a major impediment to its utility.<sup>17</sup>

Mutations adjacent to the autolytic cleavage site, such as S219D, S219V<sup>8</sup> and S219N,<sup>18</sup> were demonstrated to have the ability to inhibit self-processing, and thus to increase enzyme stability. Remarkably, the S219V mutant also increased its catalytic efficiency.<sup>8</sup> The crystal structure of a catalytically inactive mutant TEV PR (C151A) revealed that the Ser219 residue forms no direct hydrogen bond(s) or hydrophobic interactions with the peptide substrate bound to the active site.<sup>7</sup> Nevertheless, its replacement was considered to potentially alter P1' amino acid preferences. In vivo intracellular and in vitro processing assays also showed that the S219V mutant is able to accommodate not only Ser or Gly, but also many other residues in the P1' position.<sup>1</sup>

Effects of the R203G mutation on the specificity of TEV PR have also been studied. The R203G mutation was identified in a screening study by random mutagenesis and homologous recombination, which was designed to identify mutants that could process a P1'-Arg-containing substrate with an increased proteolysis rate.<sup>9</sup> It is important to note that the form of the enzyme that was investigated in this study, referred to as pTEV2 PR, contained a truncated N terminus, and thus it encompassed only the catalytic domain of TEV PR rather than the full-length protein. Moreover, in addition to the mutation of Arg203,<sup>9</sup> pTEV2 PR contained amino acid substitutions at six other positions (T17D, N68D, I77V, I83M, T118A, and S219V)<sup>19</sup> including that of Ser219 (S219V) being introduced in order to improve the enzyme's stability (Figure S1). The mutated residues other than Ser219 and Arg203 are distant from the active site and do not

constitute a part of any binding site. Arg203 is located in a loop that connects two  $\beta$ -strands and is part of a lid over the catalytic site.<sup>7</sup> That residue is more distant from the active site than Ser219 (Figure 1), and thus it does not appear to contribute to direct enzyme-substrate interactions. Structural analysis of the R203G mutant TEV PR revealed no changes of the protein structure, whereas a decrease of the bias for the amino acid at P1' position was observed.<sup>9</sup>

The specificities of TEV PRs containing one or more additional mutations besides S219V have already been investigated. Examples include S219V/A169L, S219V/N171D, S219V/Y178V, S219V/V209S, S219V/K220A,<sup>20</sup> as well as S219V/L56V/S135G, S219V/T17S/N68D/I77V, and S219V/T17S/L56V/N68D/I77V/S135G.<sup>21</sup> However, a double mutant TEV PR containing the S219V and R203G mutations has not been studied so far. Therefore, in this work we aimed to determine whether combination of these mutations changes the enzyme's specificity, as an enzyme with more relaxed P1' specificity might be more flexible for fusion tag removal. Here we describe a comparative analysis of the S219V and S219V/R203G mutant TEV PRs, with special emphasis on their P1' amino acid preferences. The specificity was studied in vitro by using a series of His<sub>6</sub>-MBP-mEYFP recombinant proteins and synthetic oligopeptide substrates representing P1' variants of the canonical TEV PR cleavage site. The structural features were also investigated by molecular dynamics (MD) analysis of the wild-type and mutant enzymes.

## 2 | MATERIALS AND METHODS

### 2.1 | Molecular dynamics

Crystal structures of the active [PDBID: 1LVM] and inactive [PDBID: 1LVB] TEV PR<sup>7</sup> were used for the MD simulations, using protomer A of the active enzyme as a starting structure. The conformation of the Ser219 residue was modeled based on the crystal structure of the inactive protease, and the substrate was also taken from the same coordinate file (PDBID: 1LVB). The S219V and R203G single as well as the S219V/R203G double mutations were introduced by using Chimera software<sup>22</sup> and the dynamoics rotamer library.<sup>23</sup>

For the simulations, the Amber22 software<sup>24</sup> was used, taking advantage of GPU acceleration<sup>25–27</sup> and the AmberTools23's features.<sup>28</sup> The protonation states of titratable residues were determined with

Propka 3.1.<sup>29,30</sup> For the catalytic site, a neutral catalytic dyad was applied, based on earlier computational studies.<sup>31</sup> The protonation state of neutral histidines was determined by using Chimera software.<sup>22</sup> The water molecules of the primary structure were kept. Possible clashes were checked in the case of the mutants. Further solvent water molecules were added to the structures with tleap. The protein and the substrate were parametrized by the FF14SB force field<sup>32</sup> and the surrounding water molecules were described by the TIP3P model.<sup>33</sup> Besides the water molecules that were added to the enzymes using the tleap software, sodium and chloride ions were also added in the proper number order to have a final concentration corresponding to that of 0.15 M NaCl. The proper number of the required ions to be added was determined by considering the charge of the enzyme, making the system neutral, in accordance with the split protocol.<sup>34</sup> The sodium and chloride ions were added by replacing randomly selected water molecules, using the `addionsrand` command. Before the MD simulations, the structures were first minimized with the steepest descent method for 2500 steps, followed by another minimization with the conjugate gradient method, for additional 7500 steps, applying  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  positional restraints on solute heavy atoms. The system was heated to 100 K in a 250 ps simulation in an NVT ensemble, followed by heating to 310 K in a 250 ps simulation, in NPT ensemble. A further minimization was applied, with  $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  positional restraints on C $\alpha$  atoms. Subsequently the same heating protocol was applied, keeping the restraint. The restraint was gradually decreased by  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  in a series of five 100 ps long equilibrations at 310 K in NPT ensemble. The production run was carried out at 310 K, for 250 ns with a 2 fs step size, with no positional restraint. Three independent replicates were run for each protein. During the last equilibration step, `cpptraj`<sup>35</sup> was used to monitor the convergence of the structures, calculating the RMSD compared to the reference structure that was the average of the geometries over the equilibration trajectory. Following completion of the MD simulations, a clustering protocol was applied with `cpptraj`, applying the average-linkage clustering method. The cluster representative of the largest cluster, in all cases representing more than 90% of the trajectory, was used to evaluate the enzyme–substrate interactions with `LigPlot+`.<sup>36</sup> The MMPBSA method was applied to evaluate the binding energy of the substrate, applying the `MMPBSA.py script`.<sup>37</sup>

## 2.2 | Preparation of the plasmids for expression of His<sub>6</sub>-MBP-mEYFP proteins

The pDestHis-MBP-mEYFP expression plasmid was designed previously.<sup>38,39</sup> First, a recognition site for the *Bam*HI restriction endonuclease was introduced by modifying the plasmid using the mutagenesis primers listed in Table S1. Then, the existing sequence coding for the TEV PR cleavage site was exchanged to that of the TVMV PR, using *Bam*HI and *Pac*I restriction endonucleases for cleavage and T4 DNA ligase for ligation.

The oligonucleotide primers coding for the wild-type (TENLYFQ\*SGTRR) and P1' mutant TEV PR cleavage sites were cloned into modified pDestHis-MBP-mEYFP expression plasmids using *Pac*I

and *Nhe*I restriction endonucleases, based on protocols described previously.<sup>38</sup> The oligonucleotide primers are listed in Table S1.

## 2.3 | Expression and purification of TEV PRs

The expression and purification of the S219V and S291V/R203G mutants of the TEV PR catalytic domain containing an N-terminal poly-histidine tag were performed as previously reported,<sup>40</sup> except that the final solution used for size exclusion chromatography contained 150 mM NaCl. The S219V/R203G mutant was prepared with the Quickchange Lightning Site-Directed Mutagenesis kit (Agilent Technologies) using the S219V expression plasmid (pRK793) as a template. The protein concentration was determined by the conventional BCA method (Pierce).

## 2.4 | Expression and purification of His<sub>6</sub>-MBP-mEYFP protein substrates

The His<sub>6</sub>-MBP-mEYFP substrates were prepared, expressed, and purified based on the method described previously.<sup>38,39</sup> After cloning the coding sequences of cleavage sites into the pDestHis-MBP-mEYFP bacterial expression plasmid, the proteins were expressed in BL21 (DE3) *Escherichia coli* cells and were purified using Ni-NTA magnetic affinity beads (Pierce, #78605). The success of cloning was confirmed by sequencing in every case. After purification, the proteins were concentrated by ultrafiltration (10 K Amicon Ultra-0.5 Centrifugal Filter Units), and the buffer was changed to reaction buffer (25 mM sodium phosphate, 100 mM NaCl, 5 mM DTT, 10% glycerol, pH 7.4).

## 2.5 | Cleavage reactions with His<sub>6</sub>-MBP-mEYFP protein substrates

The cleavage reactions were performed in the reaction buffer (25 mM sodium phosphate, 100 mM NaCl, 5 mM DTT, 10% glycerol, pH 7.4). The reaction mixtures contained the enzyme (5 ng/ $\mu$ L,  $\sim$ 175 nM final concentration) and the protein substrate (100 ng/ $\mu$ L,  $\sim$ 1.5  $\mu$ M final concentration) in 20  $\mu$ L final volume and were incubated at 30°C for 60 min. The reactions were stopped by the addition of 6 $\times$  sample loading buffer containing SDS and  $\beta$ -mercaptoethanol, followed by incubation at 95°C for 10 min. The samples were loaded onto 16% polyacrylamide gel and separated at 100 V, the gels were stained with Coomassie dye. The cleavage reactions were performed in triplicates. After imaging, the gels were evaluated using GelAnalyzer2010a software ([www.gelanalyzer.com](http://www.gelanalyzer.com), designed by István Lázár Jr. and Sr.) for densitometry.

## 2.6 | Cleavage reactions with oligopeptide substrates

To determine the kinetic parameters of the S219V and S219V/R203G mutant TEV PRs, oligopeptide substrates representing the wild-type

(P1'S) and P1'C, P1'A, P1'H, P1'G, and P1'M variants of TENLYFQ\*SGTRR TEV PR cleavage site were applied. The oligopeptides were dissolved in water with the exception of the P1'M mutant which was dissolved in water containing 5 mM DTT. Five microliters of purified TEV PR (79–288 nM in 25 mM Na-phosphate, 100 mM NaCl, 5 mM DTT, 10% glycerol, pH 7.4) was incubated with 5  $\mu$ L of oligopeptide substrate (0.017–1.74 mM final concentration) and 10  $\mu$ L incubation buffer (25 mM Na-phosphate, 100 mM NaCl, 5 mM DTT, 10% glycerol, pH 7.4) for 1 h at 30°C. The activity measurements were performed at five different substrate concentrations with the range selected based on the approximate  $K_M$  values. After the incubation, the reactions were stopped by adding 180  $\mu$ L of 1% (v/v) trifluoroacetic acid (TFA), then the samples were loaded onto a Nova-Pak C18 reversed-phase chromatography column (3.9  $\times$  150 mm, Waters, Milford, MA, USA) and mixtures were analyzed by using a Shimadzu Nexera X2 HPLC system. The substrates and reaction products were separated by an increasing water–acetonitrile gradient (0–100%) in the presence of 0.05% TFA. Kinetic parameters were determined at less than 20% substrate conversion, and kinetic parameters ( $v_{max}$ ,  $K_M$ ,  $k_{cat}$ ) were determined by using GraphPad Prism 9.4.1 (for Windows, GraphPad Software, La Jolla, CA, USA; [www.graphpad.com](http://www.graphpad.com)). While no specific inhibitor was available to perform active-site titration, 100% activity was assumed for TEV PRs while calculating the  $k_{cat}$  values.

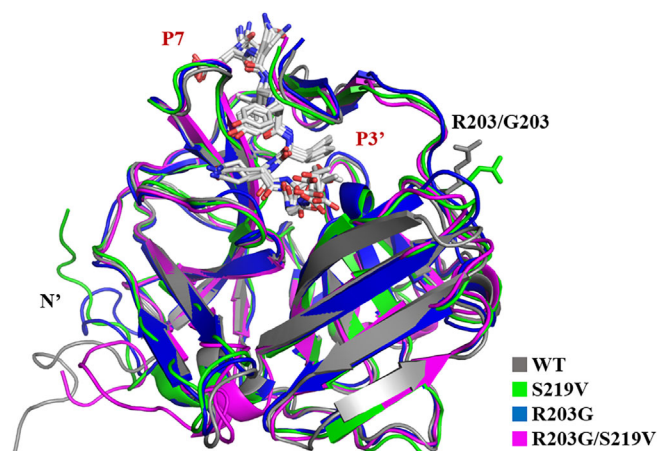
### 3 | RESULTS

#### 3.1 | In silico analysis of the TEV PR structures

The possible effects of the single (S219V or R203G) and combined mutations (S219V/R203G) on the enzyme structure and enzyme–substrate interactions were analyzed in silico. Unlike the pTEV2 PR<sup>9</sup> (Figure S1), the TEV PRs studied in this work contained only the S219V and/or R203G mutations of the canonical sequence (UniProt ID: P04517).

The structural impacts of the single and double mutations were investigated by MD analysis. After the MD simulations, the cluster representative structures were aligned (Figure 2), and the effects of the mutations were determined by calculating the RMSD values for the overall structures (Table 1). The analysis of RMSD values revealed that the structures of the wild-type and mutant TEV PRs are highly similar (Table 1) and neither the single nor the double mutations induced noticeable conformational changes to the active site or to the overall structure. The substrate also showed no significant conformational changes. The terminal P7 and P3' residues of the decapeptide were the most flexible while the P4–P2' residues showed no such movement upon mutations. The most remarkable changes were observed at the N terminus of the enzyme, although this flexible region is quite distant from the active site (Figure 2).

Renicke et al. reported that the R203G mutation alters the surface charge distribution of the enzyme.<sup>9</sup> Such remarkable change of the surface electrostatics was not observed for the S219V mutant, but the positive charge was reduced upon the R203G mutation and a positively charged surface patch was not present in the R203G mutation-containing enzymes (Figure 3).



**FIGURE 2** Structural alignment of wild-type and mutant TEV PRs. The cluster representatives of the complexes were aligned by using PyMOL. The obtained RMSD values are shown in Table 1. The oligopeptide substrate bound to the active site represents the canonical TENLYFQ\*SGT cleavage site sequence.

**TABLE 1** Comparison of the TEV PR structures.

Enzyme	RMSD
S219V	1.20 Å
R203G	1.10 Å
S219V/R203G	1.08 Å

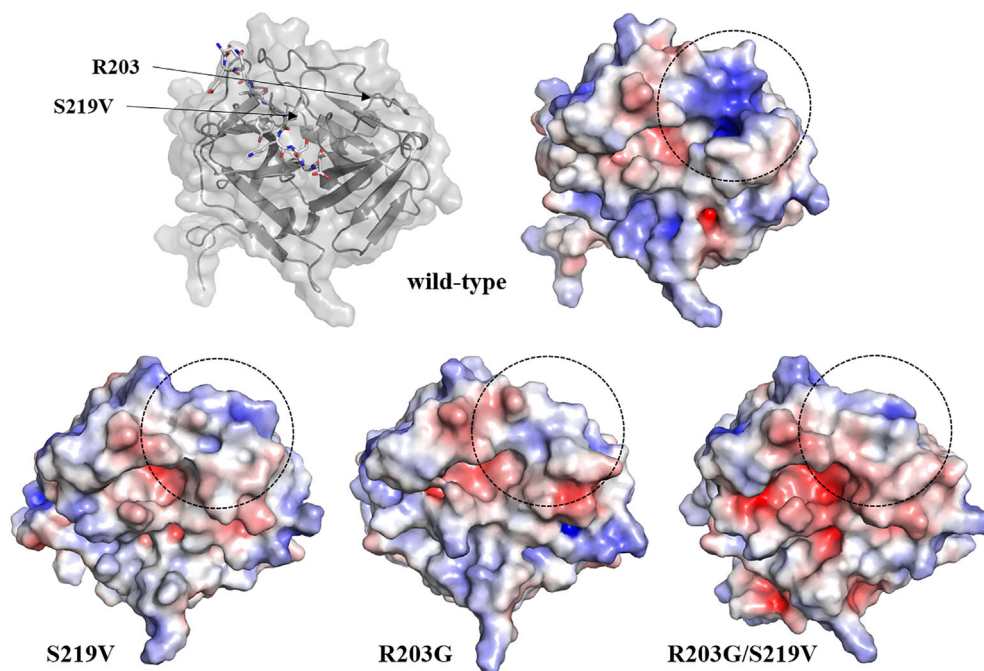
Note: The RMSD values (Å) were determined by aligning the clustered structures of the mutants to that of the wild-type, by using PyMOL.

The cluster representative structures were used to map the enzyme–substrate contacts. Compared to the wild-type enzyme, the single and double mutants that contained the R203G mutation showed a smaller number of hydrogen bonds and a higher number of non-bonded contacts, whereas the total number of hydrogen bonds was identical for the S219V single mutant and wild-type TEV PRs (Table 2). Nevertheless, the prevalence of these contacts can frequently vary during the MD trajectory.

The enzyme–substrate interactions that may directly contribute to interactions with the P1' residue were also mapped (Tables 3 and 4 and S2). In all structures (wild-type, S219V, R203G, and S219V/R203G), the H-bond-mediated interactions at the S6–S3' sites were almost identical and most of the unique interactions were observed at the S3, S1, and S1' sites of the R203G mutant (Table 3).

The number of side-chain-mediated non-bonded contacts of the P1'–Ser side chain was similar in the S219V and S219V/R203G mutants, while the value was lower for the R203G single mutant, being more similar to the value obtained for the wild-type enzyme (Table 4). This is in agreement with the fact that Arg203 is located more distantly from the S1' site than Ser219. Overall, the number of non-bonded interactions was smaller in the enzyme containing only the R203G mutation, while the contacts were more prevalent in the wild-type enzyme, as well as in the enzymes containing the S219V mutation.

**FIGURE 3** Electrostatic surface representation of wild-type and mutant TEV PR structures. The electrostatic surfaces are represented based on the cluster representative structures from the MD simulations. The regions containing the 203th residue (Arg or Gly) are circled.



**TABLE 2** Total number of hydrogen bonds and non-bonded contacts in the enzyme-substrate complexes.

	Wild-type	S219V	R203G	S219V/R203G
Total number of hydrogen bonds	18	18	17	15
Total number of non-bonded contacts	129	139	140	141

Note: The interactions were mapped by using “Generate” module of PDBSum.

**TABLE 3** Hydrogen bonds between TEV PR and an oligopeptide substrate.

Residue	Wild-type	S219V	R203G	S219V/R203G
P7-Thr	-	-	-	-
P6-Glu	N171, N176, T178, H214	N171, N176, T178, H214	N171, N176, T178, H214	N171, N176, T178, H214
P5-Asn	-	-	-	-
P4-Leu	K215, F217	K215, F217	K215, F217	K215, F217
P3-Tyr	S170 (2), N174	S170 (2), N174	S170 (2), N174, <b>D148</b>	S170 (2), N174
P2-Phe	F217	F217	-	F217
P1-Gln	T146 (2), G149, H167, S168	T146 (2), G149, H167, S168	T146 (2), <b>D148</b> , G149, H167, S168	T146, G149, H167, S168
P1'-Ser	-	-	<b>T30</b>	-
P2'-Gly	T30, S31 (2)	T30, S31	-	S31
P3'-Thr	-	<b>T29</b>	-	-

Note: Interactions are shown for each substrate residue. The numbers in parentheses denote the number of hydrogen bonds formed by the given residue, otherwise there was only a single hydrogen bond. The interactions were determined based on cluster representatives of each ensemble by using LigPlot+. The interactions that are formed by the residues shown in bold are unique for the given mutant at the given site.

The structures of the enzyme-substrate complexes were used to calculate binding free energies by using the MMPBSA method (Table 5). The binding free energy was the lowest for the R203G single mutant and implied a slightly destabilizing effect of the R203G mutation. The values obtained for the S219V and S219V/R203G mutant enzymes were highly similar in case of both different approaches and indicated a more preferred binding of the oligopeptide substrate to these enzymes than for the R203G mutant.

### 3.2 | Modification of the pDestHis-MBP-mEYFP expression plasmid

For our specificity studies, we chose a His<sub>6</sub>-MBP-mEYFP recombinant fusion protein substrate-based assay that has already been applied successfully to measure the activity of TEV PR.<sup>38,39</sup> In addition, this substrate system enables the easy and cheap generation of substrate libraries, such as we used it to prepare a series of His<sub>6</sub>-MBP-mEYFP

Residue of S1' site	wild-type	S219V	R203G	S219V/R203G
T30	-	1	-	1*
S31	-	3*	-	3*
L32	-	1*	-	-
H46	2*	2*	4*	3*
G149	4	-	1	-
C151	1	2	-	1
Total (side-chain-mediated)	2	6	3	5
Total (main-chain-mediated)	7	9	5	8

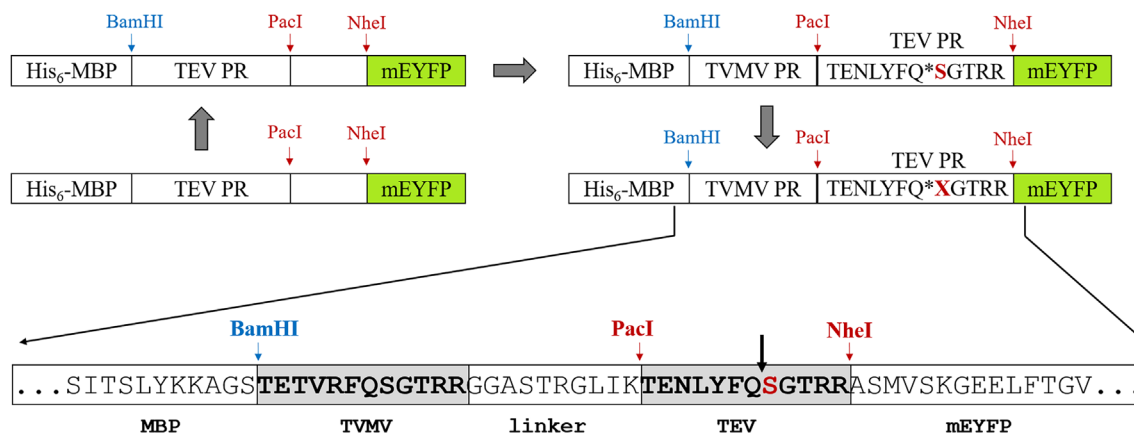
**TABLE 4** Non-bonded contacts at S1' site of TEV PR.

Note: The table shows the non-bonded interactions formed by the S1' residues, the total number of the contacts are also represented. The interactions that are mediated by the P1'-Ser side-chain are marked with asterisks, while interactions mediated by the P1'-Ser main-chain atoms are not highlighted. The individual non-bonded contacts are listed in Table S2.

	S219V	R203G	R203G/S219V
$\Delta G_{\text{bind,solv}}^0$ , generalized Born (kcal/mol)	-5.96	1.00	-2.77
$\Delta G_{\text{bind,solv}}^0$ , Poisson Boltzman (kcal/mol)	-6.85	2.48	-3.17

**TABLE 5** Calculated binding free energies of the oligopeptide substrate.

Note: The binding free energy values in water were calculated by using two different approaches of the MMPBSA method. The values obtained for the mutants were compared the value of the wild-type.



**FIGURE 4** Modification of pDestHis-MBP-mEYFP expression plasmid for the preparation of P1' variant substrates of TEV PR. The main steps of plasmid modification are represented. The sequence between the MBP and mEYFP fusion partners—encompassing the TVMV and TEV PR cleavage sites—is enlarged. The P1' residue of the TEV PR cleavage site is red, the black arrow shows the cleavage position.

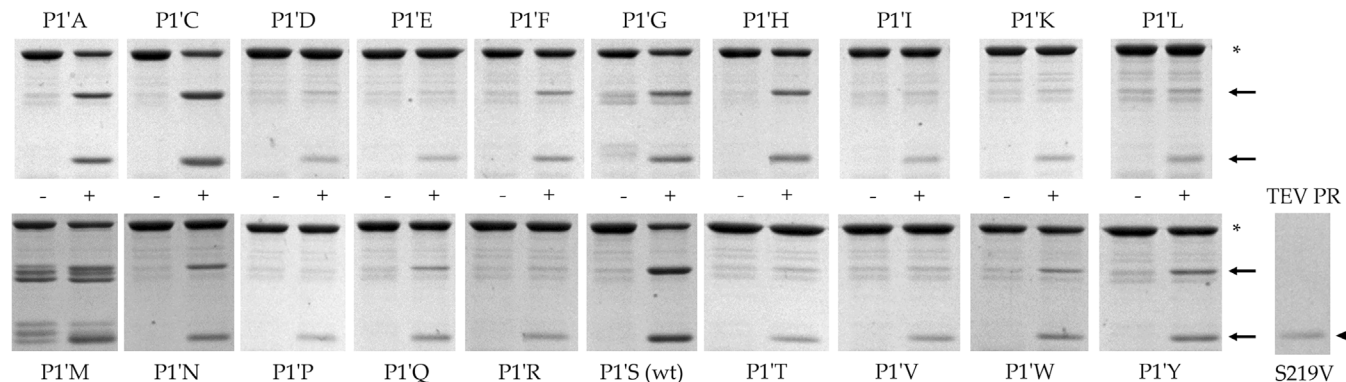
substrates and study the P1' specificity of the Venezuelan equine encephalitis virus (VEEV) non-structural protein 2 protease (nsP2pro).<sup>41</sup>

The previously designed pDestHis-MBP-mEYFP plasmid<sup>39</sup> was modified by PCR-based mutagenesis (Figure 4A). The original expression construct contained only a single cloning cassette—enabling cloning of an insert by *PacI* and *NheI* restriction endonucleases—but an additional cloning cassette was prepared by the introduction of a *BamHI* recognition site. Then, the coding sequences of the TVMV and TEV PR consensus cleavage sites were cloned into the modified plasmid containing two cloning cassettes (Figure 4B). This plasmid was used for the preparation of a series of substrates representing P1'-modified versions of the TEV PR cleavage site

(TENLYFQ\*SGTRR). The complementary primer pairs coding for the P1'-modified sequences were ligated into the plasmid that was linearized with *PacI* and *NheI* restriction enzymes.

### 3.3 | Screening of P1' amino acid preferences using His<sub>6</sub>-MBP-mEYFP substrates

To examine the P1' amino acid preferences of the S219V and S219V/R203G mutants, a series of His<sub>6</sub>-MBP-mEYFP substrates were screened in cleavage reactions. The substrates represented 20 different P1' variants of the natural TEV PR cleavage site including the wild-type sequence (TENLYFQ\*SGTRR).



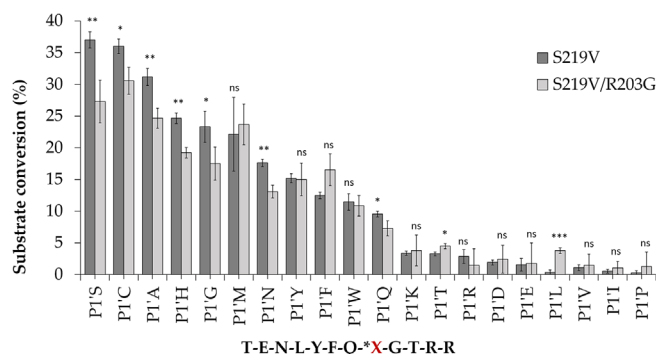
**FIGURE 5** Processing of His<sub>6</sub>-MBP-mEYFP protein substrates with S219V mutant TEV PR. The figure shows SDS-PAGE analysis of representative cleavage reactions based on three independent experiments. Asterisks and arrows indicate the full-length substrates and cleavage products, respectively. The enzyme control is also represented, the arrowhead shows the band of TEV PR (S219V mutant).

The His<sub>6</sub> affinity tag enables the purification of the recombinant protein by immobilized metal affinity chromatography, the MBP fusion partner enhances the water-solubility of the protein substrate and improves its folding, while the C-terminal mEYFP tag can be detected based on its fluorescence. In this work, both the TEV PRs and the recombinant substrates were fused to an N-terminal His<sub>6</sub> tag, therefore it was not possible to apply a heterogeneous assay format, that is, perform the cleavage reactions by using substrates immobilized onto nickel-chelate affinity surface and follow the product formation by fluorimetry.<sup>38,39</sup> Accordingly, we applied a homogenous assay for screening because the cleavage efficiencies determined by the homogenous assay were found to correlate with those determined by the heterogeneous assay format.<sup>41</sup> After separation of the substrates and cleavage products by SDS-PAGE (Figure 5), the cleavage efficiencies were determined by calculating the substrate conversions based on band intensities.

Based on the apparent molecular weights of the cleavage products, all processed substrates were cleaved only at the incorporated TEV PR cleavage site but not at the TVMV PR cleavage site. This is in agreement with the previously established specificity of TEV PR, it was unable to cleave the natural cleavage site of TVMV PR.<sup>20</sup>

The cleavage efficiencies were determined based on the relative amounts of the His<sub>6</sub>-MBP-containing cleavage product and substrate. The mEYFP-containing product and the enzyme had highly similar molecular weights, and thus it was not possible to separate properly these proteins from each other by SDS-PAGE, even using more dense gels for separation. However, it is important to note that the amounts of cleavage products are stoichiometrically equivalent; therefore, both of the products can be used to estimate substrate turnover. The P1'M substrate was found to have relatively lower stability in the applied conditions (Figure 5), while no such instability was observed for the other substrates.

The highest turnover was observed for P1'S, P1'C, P1'A, P1'H, P1'G, and P1'M-modified substrates in the case of both enzymes (Figure 6). Lower cleavage efficiency was observed for the P1' residues containing aromatic side chains, while the P1' mutants with charged residues showed remarkably lower turnover (<5%) (Figure 6).

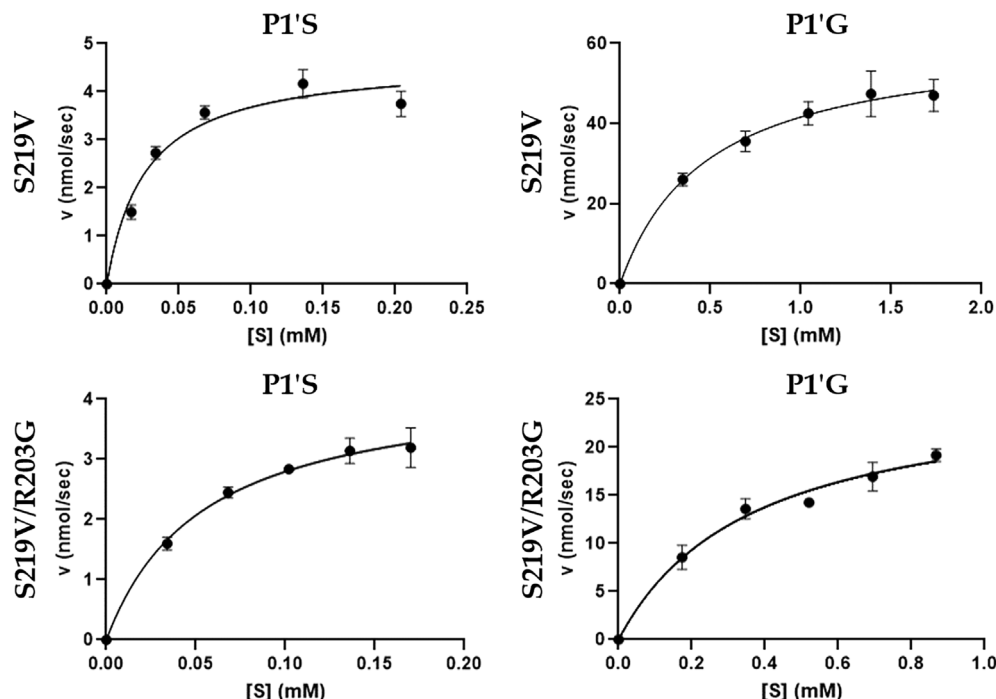


**FIGURE 6** Screening of His<sub>6</sub>-MBP-mEYFP protein substrate series containing P1' mutations. Substrate conversion (%) was determined based on product formation. Error bars represent SD of average values,  $n = 3$ . The red X indicates P1' residue in the cleavage site sequence of TEV PR. The substrate conversions were determined by densitometry of the gel images (Figure 5). Statistical analysis was performed by using the QuickCalcs unpaired t test calculator online tool of GraphPad (available at <https://www.graphpad.com/quickcalcs/ttest2>, date of last accession: August 16, 2023). ns, non-significant ( $p > .05$ ); \* $p \leq .05$ ; \*\* $p \leq .01$ ; \*\*\* $p \leq .001$ .

Both enzymes showed the slowest cleavage of the substrates that contained Pro, Ile, Val, Leu, Glu, Asp, Arg, Thr, or Lys residues in P1' position; less than 5% substrate conversion was observed for these substrates (Figure 6).

The P1' amino acid preferences of the S219V/R203G mutant are in agreement with those of the R203G mutant (Figure 6), as it was found previously to cleave those substrates with the lowest efficiency for substrates that contain Arg, Ile, Leu, Lys or Val residues in the P1' position, while the P1'-Pro mutant was not processed.<sup>9</sup> Intracellular processing of fusion protein substrates also revealed that the P1'-Pro variant substrate is not cleaved by the S219V mutant TEV PR as well, and the  $\beta$ -branched hydrophobic residues were also among those less preferred in this position.<sup>1</sup>

The substrate conversions determined for the S219V single and S219V/R203G double mutants were highly similar (Figure S2A). The observed processing efficiencies were comparable with the  $k_{cat}/K_M$



**FIGURE 7** Representative kinetic curves. The graphs show the results of kinetic analysis of S219V and S219V/R203G mutant TEV PRs using P1'S- and P1'G-modified oligopeptide substrates. Error bars represent SD of average values,  $n = 3$ .

values that were determined previously for the S219V mutant in case of the corresponding P1' mutants<sup>1</sup> (Figure S2B).

In order to determine the dependence of *in vitro* cleavage efficiencies of the fluorescent substrates on the hydrophobicities and volumes of the P1' residues, linear regression analysis was performed in the case of both mutant enzymes (Figure S3). The Kyte and Doolittle hydrophathy scale<sup>42</sup> and the residue volumes were also retrieved from the literature.<sup>43</sup> The analysis revealed that there is no correlation between the substrate conversions and the volumes or hydrophobicities of P1' residues. Nevertheless, the highest substrate turnover was determined for the substrates containing a small volume P1' residue, such as Ser, Gly, Ala, Cys, or Asn, both in the case of the S219V and the S219V/R203G mutant enzyme (Figure S3A).

### 3.4 | Kinetic analysis of TEV PRs using synthetic oligopeptide substrates

Based on the results of protein substrate-based screening, the most preferred P1'-variants were selected for downstream kinetic analysis. The synthetic oligopeptide substrates represented the wild-type and P1'C, P1'A, P1'H, P1'G, and P1'M mutants of the TEV PR cleavage site (TENLYFQ\*SGTRR). However, it is important to note that the reaction buffer we used in this work (incubation buffer: 25 mM Naphosphate, 100 mM NaCl, 5 mM DTT, 10% glycerol, pH 7.4) was slightly different from the buffer environment that was used to determine the kinetic parameters of the S219V mutant TEV PR (TEV PR reaction buffer: 50 mM Tris-HCl, 0.5 mM EDTA, 1 mM DTT, pH 8.0).<sup>1</sup> Some representative kinetic curves are shown in Figure 7.

The P1'C variant was found to be the best substrate for both enzymes, while the His residue was the least preferred in the P1'

position. The order of preferences was slightly different for the other substrates (Table 6). The catalytic constants of the S219V/R203G mutant were almost identical for the P1'C and P1'M substrates, and for the P1'S and P1'A substrates, respectively. The amino acid preferences established in this work agree with those determined by a former kinetic analysis of the S219V mutant,<sup>1</sup> the P1'C variant was less efficient substrate as compared to the P1'S, P1'G and P1'A. The highest catalytic constants were obtained for the P1'C substrate, the  $k_{cat}/K_M$  values were 0.899 and 0.548  $\text{mM}^{-1} \text{s}^{-1}$  for the S219V and S219V/R203G mutants, respectively. The P1'M mutant was more preferred by the S219V/R203G than by the S219V mutant TEV PR (in agreement with the screening of recombinant substrates, see Figure 6), this was the only substrate for which the catalytic constant was higher for the double mutant as compared to the single mutant enzyme.

The data obtained from the screening of protein substrates and by the oligopeptide-based kinetic assays were in good agreement in case of both enzymes (Figure S4A,B).

A possible limitation of this study is that the enzyme-substrate complexes were investigated *in silico* only in the context of an oligopeptide that represents the canonical TENLYFQ\*SGT cleavage site, but a series of P1' mutant substrates were not prepared and analyzed. Consequently, the amino acid preferences obtained for the selected substrate variants *in vitro* (Table 6) could not be corroborated with structural studies.

The MD simulations revealed a greater number of hydrogen bonds formed with the oligopeptide substrate (representing the canonical TENLYFQ\*SGT cleavage site sequence) in the case of the S219V mutant TEV PR as compared to the S219V/R203G mutant (18 and 15, respectively) (Table 2). The number of the side-chain- and main-chain-mediated non-bonded contacts of the P1'-Ser side chain



**TABLE 6** Kinetic parameters determined for S219V and S219V/R203G mutant TEV PRs.

Enzyme	Substrate	$v_{\max}$ (nM s <sup>-1</sup> )	$K_M$ (mM)	$k_{\text{cat}}$ (s <sup>-1</sup> )	$k_{\text{cat}}/K_M$ (mM <sup>-1</sup> s <sup>-1</sup> )	$k_{\text{cat}}/K_M$ (mM <sup>-1</sup> s <sup>-1</sup> ) <sup>a</sup>
S219V	P1'A	21.026 ± 1.575	0.189 ± 0.049	0.120 ± 0.009	0.635 ± 0.171	3.01 ± 0.54
	P1'G	61.478 ± 3.872	0.493 ± 0.094	0.352 ± 0.022	0.714 ± 0.143	3.08 ± 0.67
	P1'S	4.667 ± 0.222	0.027 ± 0.005	0.021 ± 0.001	0.778 ± 0.149	4.51 ± 0.65
	P1'H	7.061 ± 0.698	0.069 ± 0.024	0.024 ± 0.002	0.348 ± 0.124	n.d.
	P1'C	19.601 ± 1.166	0.249 ± 0.034	0.224 ± 0.013	0.899 ± 0.133	n.d.
	P1'M	18.304 ± 1.063	0.315 ± 0.036	0.150 ± 0.009	0.476 ± 0.062	2.37 ± 0.23
S219V/R203G	P1'A	23.893 ± 1.848	0.241 ± 0.048	0.109 ± 0.008	0.452 ± 0.096	
	P1'G	26.304 ± 1.876	0.364 ± 0.063	0.120 ± 0.009	0.329 ± 0.062	
	P1'S	4.312 ± 0.218	0.054 ± 0.008	0.025 ± 0.001	0.463 ± 0.071	
	P1'H	4.304 ± 0.406	0.090 ± 0.026	0.025 ± 0.002	0.278 ± 0.083	
	P1'C	16.137 ± 1.440	0.374 ± 0.064	0.205 ± 0.018	0.548 ± 0.105	
	P1'M	18.072 ± 1.190	0.342 ± 0.049	0.187 ± 0.012	0.547 ± 0.086	

Note: Oligopeptides representing the TENVLYFQ\**X*GTRR cleavage site were used as substrates, the *X* denotes the P1' residue in the sequence (Ala, Gly, Ser, His, Cys, or Met). Error bars represent SD of average values,  $n = 3$ .

Abbreviation: n.d., not determined.

<sup>a</sup>The  $k_{\text{cat}}/K_M$  values were determined previously for the S219V mutant TEV PR by using a series of P1'-modified oligopeptide substrates.<sup>1</sup>

was also higher for the S219V mutant (Table 4). The slightly higher density of enzyme-substrate interactions is consistent with the two-fold higher  $K_M$  obtained for the double mutant TEV PR.

The  $k_{\text{cat}}/K_M$  values determined for the S219V/R203G mutant were lower than 0.6 mM<sup>-1</sup> s<sup>-1</sup>, while were between 0.6 and 0.9 for the S219V mutant TEV PR in the case of four of the six studied substrates (Table 6). On average, the  $k_{\text{cat}}/K_M$  values indicate higher catalytic efficiency of the S219V mutant enzyme. The catalytic constants reported here were lower than the previously determined ones<sup>1</sup> due to differences in the conditions of the protease assays (such as buffer conditions, especially lower ionic strength). Nevertheless, the P1' specificities of the two mutants were successfully compared, and the preferences determined for the P1'A, P1'G, P1'S and P1'M were identical: Ser > Gly > Ala > Met (Table 1), indicating consistency of the obtained results. Based on the  $k_{\text{cat}}/K_M$  values that were obtained from the oligopeptide-based kinetic analysis, the following P1' amino acid preferences can be invoked: S219V: Cys > Ser > Gly > Ala > Met > His; S219V/R203G: Cys/Met > Ser/Ala > Gly > His.

The correlation of the experimental data with the physical parameters of P1' residues was carried out in the case of  $k_{\text{cat}}/K_M$  values, as well. Linear regression analyses revealed only moderate correlation of the catalytic constants with the volumes of the P1' residues in the case of the single mutant enzyme, but no correlation was observed for the double mutant (Figure S4C,D). Interestingly, a strong relationship was observed between the  $k_{\text{cat}}/K_M$  values and the P1' residue hydrophobicity only in the case of the S219V/R203G mutant enzyme, the highest catalytic constants were obtained for the most hydrophobic P1' variants (Figure S4E,F). Nonetheless, such a strong correlation was not observed if the series of His<sub>6</sub>-MBP-mEYFP protein substrates was screened (Figure S3B). The lowest catalytic constant was obtained for the P1'-His variant for both enzymes, this residue is less hydrophobic and has a greater volume than the other five studied P1' variants.

The P1' substrate variants were compared in the context of the activation energies ( $\Delta G_{\text{exp}}^\ddagger$ ), as well. The energies were calculated based on the  $k_{\text{cat}}$  values determined in this study (Table 6) or reported previously for the S219V mutant.<sup>1</sup> The experimentally derived  $k_{\text{cat}}$  values were converted into activation energies based on the Eyring-Polanyi equation (Equation (1)).

$$k = \frac{\kappa k_B T}{h} e^{-\frac{\Delta G^\ddagger}{RT}} \quad (1)$$

In this equation,  $k$  represents the rate constant ( $k_{\text{cat}}$  in the case of an enzyme reaction),  $\kappa$  is the transmission coefficient, generally accepted as 1,  $k_B$  is the Boltzmann constant,  $h$  is the Planck constant,  $T$  is temperature,  $R$  is the gas constant, and  $\Delta G^\ddagger$  is the Gibbs energy of activation. The calculated activation energy values were highly comparable and showed <0.5 kcal/mol difference, excepting the P1'S substrate (Table 7). The activation energies were highly comparable in most cases, the highest values were determined for the P1'S and P1'H variants. The high activation energy in the case of the P1'H mutant is consistent with the low cleavage efficiency of this substrate (Table 6).

## 4 | DISCUSSION

This study was performed with the aim of comparing the P1' specificity of the S219V/R203G mutant TEV PR with that of the S219V mutant in silico and in vitro. The S219V mutation was designed previously to improve the stability of TEV PR by preventing enzyme inactivation via autolysis,<sup>1</sup> while the R203G mutation was identified by random mutagenesis and purported to have even more relaxed specificity in the P1' position.<sup>9</sup> In this work we studied the S219V/R203G double mutant TEV PR with special emphasis on its P1' amino acid preferences. Structural features were investigated by molecular

**TABLE 7** Activation energy values with various substrates.

Substrate	$\Delta G_{\text{exp}}^{\ddagger}$ (kcal/mol)	$\Delta G_{\text{exp}}^{\ddagger}$ (kcal/mol) <sup>a</sup>
P1'A	18.84	18.36
P1'G	18.20	18.32
P1'S	19.88	18.56
P1'H	19.80	-
P1'C	18.47	-
P1'M	18.71	18.60

Note: The activation energies were calculated by the Eyring–Polányi equation based on the experimentally determined  $k_{\text{cat}}$  values (Table 6).

<sup>a</sup>These values were calculated based on the results published previously by Kapust et al.<sup>1</sup> The  $k_{\text{cat}}$  values were not determined previously for P1'H and P1'C substrate variants; thus, it was not possible to calculate activation energy in these cases.

dynamics (MD) analysis in the case of the wild-type, single, and double mutant enzymes, while the P1' specificities were determined in vitro only for the S219V and S219V/R203G mutants. This choice was due to the importance of the S219V mutation in the stabilization of TEV PR.<sup>8</sup>

First, the P1' specificities of both enzymes were established in vitro by screening recombinant protein substrates. For this assay, we modified a pDestHis-MBP-mEYFP expression plasmid by preparing a new cloning cassette (Figure 4), followed by the preparation of 20 His<sub>6</sub>-MBP-mEYFP substrates that contained the wild-type TENLYFQ\*SGT TEV PR cleavage site and all of its P1' variants (Figure 5). The S219V single and S219V/R203G double mutants exhibited similar P1' specificity, with highest preferences for P1'S, P1'C, P1'A, P1'H, P1'G, and P1'M residues (Figure 6). Lower tolerance was observed for the aromatic side chains in the P1' position, while residues containing charged P1' variants were among the least efficient substrates, and the turnover of the P1'P substrate was negligible. Overall, the in vitro assays revealed no remarkable differences between the specificities of the S219V and S219V/R203G mutant enzymes. In accordance with this, the MD simulation-based analysis of the enzyme–substrate complexes also implied comparable pattern of the enzyme–substrate interactions at the active site of the S219V and S219V/R203G mutant, including the hydrogen bonds (Table 2) and the non-bonded contacts (Table 4).

The Arg203 residue was deduced to form a positively charged surface patch together with another two arginine residues (Figure 3). Thus, the R203G mutation was predicted to reduce this charge, allowing easier binding of substrates with positively charged P1' residues.<sup>9</sup> In contrast, the in vitro screening of P1' mutants revealed that the substrates containing a charged residue in the P1' position are among the least tolerated ones and are inefficient substrates of both S219V and S219V/R203G mutant TEV PRs.

The TEV PR mutants we designed contained either the S219V or S219V/R203G mutations of the canonical catalytic domain sequence (UniProt ID: P04517). Although the previously described pTEV2 PR enzyme also contained both point mutations, five other mutations were also present in that enzyme (Figure S1), but it has not been elucidated unequivocally so far whether the additional five mutations

influence the specificity. The absence of the S219V mutation may be a limiting factor for the biotechnological and molecular biological application of TEV PR, as this mutation remarkably improves the global stability of the enzyme.<sup>1</sup> Therefore, the R203G single mutant—lacking any stabilizing mutation—was omitted from our in vitro study. Not only the sequence of the applied protease but some other conditions may also influence the obtained phenotypes, such as application of crude intracellular or in vitro assays, or investigation of the cleavage reactions at different substrate turnover or depletion. Hence, the P1' amino acid preferences obtained for the S219V and S219V/R203G mutants in this work are not directly comparable with those determined for pTEV2 PR.<sup>9</sup>

Based on the screening of the recombinant substrates (Figure 6), the top six cleavage site variants were further investigated through in vitro kinetic assays performed with synthetic oligopeptide substrates (Figure 7). The highest catalytic constants were determined for the P1'C-containing substrates in the case of both enzymes (Table 6). The calculated binding free energies were higher for the S219V mutant as compared to S219V/R203G mutant (Table 5), indicating more preferred binding of the P1'S variant substrate (TENLYFQ\*SGT) to the single mutant enzyme. This results of the computational analysis was in good agreement with that of the kinetic analysis (Table 6) and corroborate the higher catalytic constants  $K_M$  and  $k_{\text{cat}}/K_M$  obtained for the S219V mutant. A possible limitation of this study is that the kinetic parameters were not determined for P1'-Arg mutant oligopeptide substrate. Nevertheless, this substrate was among the less-preferred ones in the protein substrate-based screening (Figure 6).

Our hypothesis was that the R203G mutation might further relax the P1' specificity of the S219V mutant TEV PR. Proteolysis of the model fusion protein substrates was considered to resemble realistic enzyme–substrate interactions as compared to the oligopeptide substrates, as TEV PR is most typically used for the processing of recombinant proteins during their production and purification. Overall, the in vitro assays revealed no remarkable differences between the specificities of the S219V and S219V/R203G mutant enzymes. The Arg203 residue is located farther from the catalytic site than Ser219, so the effect of the mutation at the 219th position may be more dominant. We assume that the five additional mutations that were introduced into pTEV2 PR in order to improve its solubility<sup>9</sup> also influenced the P1' preferences of the enzyme and potentially contributed to the more relaxed specificity of pTEV2 in the intracellular assays. In agreement with this, we did not observe such a relaxed specificity if only the S219V and R203G mutations were present in the canonical sequence of the TEV PR's catalytic domain.

Addition of the R203G mutation along with S219V did not change the overall P1' specificity compared to the single mutant TEV PR but it caused a slight decrease its catalytic efficiency. Due to the relatively lower catalytic efficiency of the double mutant, the S219V mutant TEV PR remains the most efficient variant for fusion tag removal, irrespective of the P1' residue. It remains to be determined how the mutations of pTEV2 PR other than S219V<sup>8</sup> and R203G<sup>9</sup> influence the amino acid preferences, which may help to

identify the residues whose modification might further relax the P1' specificity.

#### AUTHOR CONTRIBUTIONS

**János András Mótyán:** Writing – original draft; visualization; validation; supervision; software; methodology; investigation; funding acquisition; formal analysis; data curation; conceptualization. **Mária Golda:** Writing – original draft; visualization; validation; methodology; investigation; formal analysis; data curation; conceptualization. **Gyula Hoffka:** Writing – original draft; visualization; software; methodology; investigation; formal analysis; data curation; conceptualization. **Scott Cherry:** Investigation; data curation. **Joseph E. Tropea:** Writing – original draft; investigation; data curation. **George T. Lountos:** Writing – original draft; project administration. **David S. Waugh:** Writing – original draft; supervision; resources; project administration; methodology; funding acquisition; conceptualization. **Alexander Wlodawer:** Writing – original draft; supervision; resources; funding acquisition. **József Tózsér:** Writing – original draft; validation; supervision; resources; methodology; funding acquisition; conceptualization.

#### ACKNOWLEDGMENTS

The authors are grateful to members of the Laboratory of Retroviral Biochemistry for the technical support and to Dr. Károly Jambrovics (Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen) for critical reading of the manuscript. We thank the Biophysics Resource of the NCI/CCR Center for Structural Biology for use of the LC/ESMS instrument. We acknowledge KIFÜ for awarding us access to resource based in Hungary.

#### FUNDING INFORMATION

Project no. TKP2021-EGA-20 has been implemented with the support provided by the National Research, Development and Innovation Fund (Hungary) financed under the TKP2021-EGA funding scheme. This project was supported by the ÚNKP-23-5-DE-486 New National Excellence Program of the Ministry of Culture and Innovation from the source of the National Research, Development and Innovation Fund (to J.A.M.). János A. Mótyán is supported by Hungarian Academy of Sciences (BO/00110/23/5). This work was also supported in part by the NIH Intramural Research Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health, and the Frederick National Laboratory for Cancer Research, National Institutes of Health under contract 75N91019D00024. (This contract number represents work performed within the scope of the severable FFRDC Bridge contract.) The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

#### DATA AVAILABILITY STATEMENT

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

#### ORCID

George T. Lountos  <https://orcid.org/0000-0003-3463-8881>

János András Mótyán  <https://orcid.org/0000-0002-6079-5621>

#### REFERENCES

- Kapust RB, Tózsér J, Copeland TD, Waugh DS. The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun.* 2002; 294(5):949-955. doi:10.1016/S0006-291X(02)00574-0
- Mótyán JA, Tóth F, Tózsér J. Research applications of proteolytic enzymes in molecular biology. *Biomolecules.* 2013;3(4):923-942. doi:10.3390/biom3040923
- Cesaratto F, Burrone OR, Petris G. Tobacco etch virus protease: a shortcut across biotechnologies. *J Biotechnol.* 2016;231:239-249. doi:10.1016/j.jbiotec.2016.06.012
- Raran-Kurussi S, Cherry S, Zhang D, Waugh DS. Removal of affinity tags with TEV protease. *Methods Mol Biol.* 2017;1586:221-230. doi:10.1007/978-1-4939-6887-9\_14
- Nallamsetty S, Austin BP, Penrose KJ, Waugh DS. Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. *Protein Sci.* 2005;14(12):2964-2971. doi:10.1110/ps.051718605
- Rocco CJ, Dennison KL, Klenchin VA, Rayment I, Escalante-Semerena JC. Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from *Escherichia coli*. *Plasmid.* 2008;59(3):231-237. doi:10.1016/j.plasmid.2008.01.001
- Phan J, Zdanov A, Evdokimov AG, et al. Structural basis for the substrate specificity of tobacco etch virus protease. *J Biol Chem.* 2002; 277(52):50564-50572. doi:10.1074/jbc.M207224200
- Kapust RB, Tózsér J, Fox JD, et al. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* 2001;14(12):993-1000. doi:10.1093/protein/14.12.993
- Renicke C, Spadaccini R, Taxis C. A tobacco etch virus protease with increased substrate tolerance at the P1' position. *PLoS One.* 2013; 8(6):e67915. doi:10.1371/journal.pone.0067915
- Yi L, Gebhard MC, Li Q, Taft JM, Georgiou G, Iverson BL. Engineering of TEV protease variants by yeast ER sequencing screening (YESS) of combinatorial libraries. *Proc Natl Acad Sci U S A.* 2013;110(18): 7229-7234. doi:10.1073/pnas.1215994110
- Packer MS, Rees HA, Liu DR. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat Commun.* 2017;8(1): 956. doi:10.1038/s41467-017-01055-9
- Sanchez MI, Ting AY. Directed evolution improves the catalytic efficiency of TEV protease. *Nat Methods.* 2020;17(2):167-174. doi:10.1038/s41592-019-0665-7
- Mohammadian H, Mahnam K, Sadeghi HM, Ganjalikhany MR, Akbari V. Rational design of a new mutant of tobacco etch virus protease in order to increase the *in vitro* solubility. *Res Pharm Sci.* 2020; 15(2):164-173. doi:10.4103/1735-5362.283816
- Nam H, Hwang BJ, Choi DY, Shin S, Choi M. Tobacco etch virus (TEV) protease with multiple mutations to improve solubility and reduce self-cleavage exhibits enhanced enzymatic activity. *FEBS Open Bio.* 2020;10(4):619-626. doi:10.1002/2211-5463.12828
- Bayar E, Ren Y, Chen Y, et al. Construction, investigation and application of TEV protease variants with improved oxidative stability. *J Microbiol Biotechnol.* 2021;31(12):1732-1740. doi:10.4014/jmb.2106.06075

16. Enríquez-Flores S, De la Mora-De la Mora JI, Flores-López LA, et al. Improved yield, stability, and cleavage reaction of a novel tobacco etch virus protease mutant. *Appl Microbiol Biotechnol*. 2022;106(4):1475-1492. doi:10.1007/s00253-022-11786-5
17. Parks TD, Howard ED, Wolpert TJ, Arp DJ, Dougherty WG. Expression and purification of a recombinant tobacco etch virus N1a proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form. *Virology*. 1995;210(1):194-201.
18. Lucast LJ, Batey RT, Doudna JA. Large-scale purification of a stable form of recombinant tobacco etch virus protease. *Biotechniques*. 2001;30(3):544-546, 548, 550 passim. doi:10.2144/01303st06
19. Taxis C, Stier G, Spadaccini R, Knop M. Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. *Mol Syst Biol*. 2009;5:267. doi:10.1038/msb.2009.25
20. Tözsér J, Tropea JE, Cherry S, et al. Comparison of the substrate specificity of two potyvirus proteases. *FEBS J*. 2005;272(2):514-523. doi:10.1111/j.1742-4658.2004.04493.x
21. Fang J, Chen L, Cheng B, Fan J. Engineering soluble tobacco etch virus protease accompanies the loss of stability. *Protein Exp Purif*. 2013;92(1):29-35. doi:10.1016/j.pep.2013.08.015
22. Pettersen EF, Goddard TD, Huang CC, et al. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612. doi:10.1002/jcc.20084
23. Scouras AD, Daggett V. The dynamo rotamer library: amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Protein Sci*. 2011;20(2):341-352. doi:10.1002/pro.565
24. Case DA, Aktulga HM, Belfon K, et al. AMBER 2023. University of California; 2023.
25. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J Chem Theory Comput*. 2012;8(5):1542-1555. doi:10.1021/ct200909j
26. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J Chem Theory Comput*. 2013;9(9):3878-3888. doi:10.1021/ct400314y
27. Le Grand S, Götz AW, Walker RC. SPFP: speed without compromise - a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput Phys Commun*. 2013;184(2):374-380. doi:10.1016/j.cpc.2012.09.022
28. Case DA, Aktulga HM, Belfon K, et al. AmberTools. *J Chem Inf Model*. 2023;63(20):6183-6191. doi:10.1021/acs.jcim.3c01153
29. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical pK<sub>a</sub> predictions. *J Chem Theory Comput*. 2011;7(2):525-537. doi:10.1021/ct100578z
30. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK<sub>a</sub> values. *J Chem Theory Comput*. 2011;7(7):2284-2295. doi:10.1021/ct200133y
31. Zlobin A, Golovin A. Between protein fold and nucleophile identity: multiscale modeling of the TEV protease enzyme-substrate complex. *ACS Omega*. 2022;7(44):40279-40292. doi:10.1021/acsomega.2c05201
32. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput*. 2015;11(8):3696-3713. doi:10.1021/acs.jctc.5b00255
33. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79:926-935. doi:10.1063/1.445869
34. Machado MR, Pantano S. Split the charge difference in two! A rule of thumb for adding proper amounts of ions in MD simulations. *J Chem Theory Comput*. 2020;16(3):1367-1372. doi:10.1021/acs.jctc.9b00953
35. Roe DR, Cheatham TE. 3rd PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput*. 2013;9:3084-3095. doi:10.1021/ct400341p
36. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model*. 2011;51(10):2778-2786. doi:10.1021/ci200227u
37. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: an efficient program for end-state free energy calculations. *J Chem Theory Comput*. 2012;8(9):3314-3321. doi:10.1021/ct300418h
38. Bozóki B, Gazda L, Tóth F, Miczi M, Mótyán JA, Tözsér J. A recombinant fusion protein-based, fluorescent protease assay for high throughput-compatible substrate screening. *Anal Biochem*. 2018;540-541:52-63. doi:10.1016/j.ab.2017.11.001
39. Bozóki B, Mótyán JA, Miczi M, Gazda LD, Tözsér J. Use of recombinant fusion proteins in a fluorescent protease assay platform and their in-gel renaturation. *J Vis Exp*. 2019;143. doi:10.3791/58824
40. Tropea JE, Cherry S, Waugh DS. Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol Biol*. 2009;498:297-307. doi:10.1007/978-1-59745-196-3\_19
41. Bozóki B, Mótyán JA, Hoffka G, Waugh DS, Tözsér J. Specificity studies of the Venezuelan equine encephalitis virus non-structural protein 2 protease using recombinant fluorescent substrates. *Int J Mol Sci*. 2020;21(20):7686. doi:10.3390/ijms21207686
42. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157:105-132. doi:10.1016/0022-2836(82)90515-0
43. Zamyatnin A. Protein volume in solution. *Prog Biophys Mol Biol*. 1972;24:107-123. doi:10.1016/0079-6107(72)90005-3

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Golda M, Hoffka G, Cherry S, et al. P1' specificity of the S219V/R203G mutant tobacco etch virus protease. *Proteins*. 2024;1-12. doi:10.1002/prot.26693