

DOI: 10.1002/cbic.201100267

Papain-Catalyzed Peptide Bond Formation: Enzyme-Specific Activation with Guanidinophenyl Esters

Roseri J. A. C. de Beer,^[a] Barbara Zarzycka,^[b] Helene I. V. Amatdjais-Groenen,^[a] Sander C. B. Jans,^[a] Timo Nuijens,^[c] Peter J. L. M. Quaedflieg,^[c] Floris L. van Delft,^[a] Sander B. Nabuurs,^[b] and Floris P. J. T. Rutjes*^[a]

The substrate mimetics approach is a versatile method for small-scale enzymatic peptide-bond synthesis in aqueous systems. The protease-recognized amino acid side chain is incorporated in an ester leaving group, the substrate mimetic. This shift of the specific moiety enables the acceptance of amino acids and peptide sequences that are normally not recognized by the enzyme. The guanidinophenyl group (OGp), a known substrate mimetic for the serine proteases trypsin and chymotrypsin, has now been applied for the first time in combination with papain, a cheap and commercially available cysteine protease. To provide insight in the binding mode of various Z-X_{AA}-

OGp esters, computational docking studies were performed. The results strongly point at enzyme-specific activation of the OGp esters in papain through a novel mode of action, rather than their functioning as mimetics. Furthermore, the scope of a model dipeptide synthesis was investigated with respect to both the amino acid donor and the nucleophile. Molecular dynamics simulations were carried out to prioritize 22 natural and unnatural amino acid donors for synthesis. Experimental results correlate well with the predicted ranking and show that nearly all amino acids are accepted by papain.

Introduction

The importance of peptides in the fields of healthcare and nutrition renders the amide bond probably the most synthesized chemical bond.^[1] In light of this, both chemocatalytic and enzymatic strategies for amide-bond formation are being developed. The use of enzymes is advantageous, particularly on an industrial scale, because they are usually selective, prevent racemization, and require minimal or no side-chain protection. The typical enzymes of choice for peptide synthesis are proteases, readily available enzymes that normally hydrolyze peptidic amide bonds, but also show the capacity to effect amide bond formation. The typical selection criterion for a specific protease is based on the specificity for a particular amino acid residue on either side of the scissile bond. However, the main drawback of such an approach is the need for a different enzyme for nearly each desired peptide bond, as is nicely illustrated by the enzymatic synthesis of octapeptide CCK-8 for which three different proteases were required.^[2] Secondary hydrolysis of the products necessitated a judicious choice in the order of the fragment couplings, and prudent fine-tuning of the reaction conditions for each individual step. These drawbacks in combination with the limited recognition of many amino acids, both natural and unnatural, severely restrict the universal application of enzymes.

In the so-called substrate mimetics strategy, the problem of limited applicability is overcome by incorporating the enzyme-recognized amino acid side chain moiety in the ester leaving group, thereby making the enzymatic recognition independent of the side chain. A well-known example is the guanidinophenyl (OGp) group, a mimic of the arginine side chain, which is naturally recognized by the protease trypsin.^[3] The catalytic

mechanism^[4] of a substrate mimetic (SM; Scheme 1 B) is analogous to the natural situation (Scheme 1 A). First the substrate mimetic binds to the S₁ pocket; this is followed by nucleophilic attack of the catalytic serine residue on the carbonyl of the substrate. Kinetic and computational studies have indicated that the substrate mimetic should bind in a reverse orientation to enable this attack.^[5] In this way, the developing negative charge can still be stabilized by the oxyanion hole. Because the deacylation step needs unoccupied S' subsites, the non-specific acyl residue has to flip from the S' to the S subsites prior to nucleophilic attack in order to liberate both product and enzyme. Water as the nucleophile will result in the hydrolysis product, while an amino acid nucleophile will lead to dipeptide synthesis.

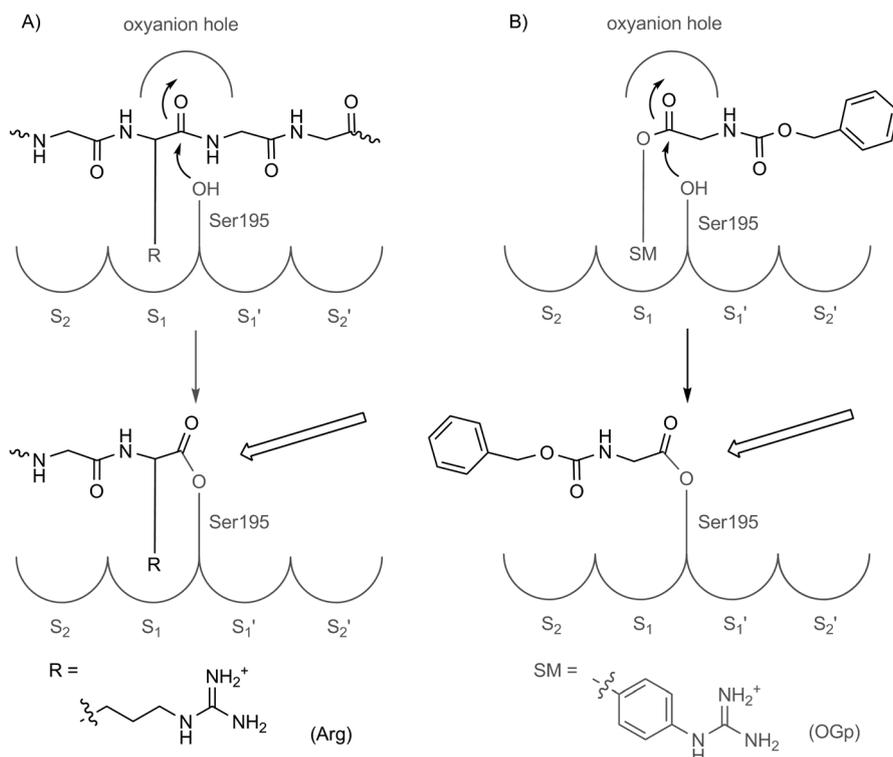
Besides fundamental studies on the mechanism, much research on substrate mimetics has been devoted to method improvement and to solving difficulties related to enzymatic pep-

[a] R. J. A. C. de Beer, H. I. V. Amatdjais-Groenen, S. C. B. Jans, Dr. F. L. van Delft, Prof. F. P. J. T. Rutjes
Institute for Molecules and Materials, Radboud University Nijmegen
Heyendaalseweg 135, 6525 AJ Nijmegen (The Netherlands)
E-mail: f.rutjes@science.ru.nl

[b] B. Zarzycka, Dr. S. B. Nabuurs
Computational Drug Discovery, Center for Molecular and Biomolecular Informatics, Radboud University Nijmegen Medical Centre
Nijmegen (The Netherlands)

[c] T. Nuijens, Dr. P. J. L. M. Quaedflieg
DSM Innovative Synthesis BV
P.O. Box 18, 6160 MD Geleen (The Netherlands)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cbic.201100267>.



Scheme 1. Catalytic mechanism of the serine protease trypsin with A) a natural substrate and B) a substrate mimetic. Recognition takes place in the S_1 subsite, with subsequent nucleophilic attack of Ser195 onto the carbonyl. The substrate mimetic is bound in a reversed orientation compared to the orientation of the natural substrate. The covalent intermediate can only be liberated by nucleophilic attack from the S' region (double arrow).

tide synthesis. For example, several variations of the OGp ester have been evaluated as alternative substrate mimetics for trypsin,^[6] while the OGp ester has been applied to different enzymes such as chymotrypsin^[7] and clostripain,^[8] but also to trypsin variants with diminished hydrolytic activity^[9] or purified from different exotic sources.^[6c,d,10] The undesired enzymatic hydrolysis reaction of specific peptide bonds can be successfully suppressed by freezing the reaction mixture^[11] or by using ionic liquids as solvents.^[12] Furthermore, it has been shown that new substrate mimetics could be designed for two enzymes that recognize negatively charged amino acids,^[13] thus indicating that more-general application of the strategy might be feasible. Despite the considerable amount of existing research, the substrate mimetics approach has never been applied to the cysteine protease papain. This cheap and commercially available enzyme would also be a good candidate for industrial applications because it has been effectively used in enzymatic peptide synthesis before and cysteine proteases are

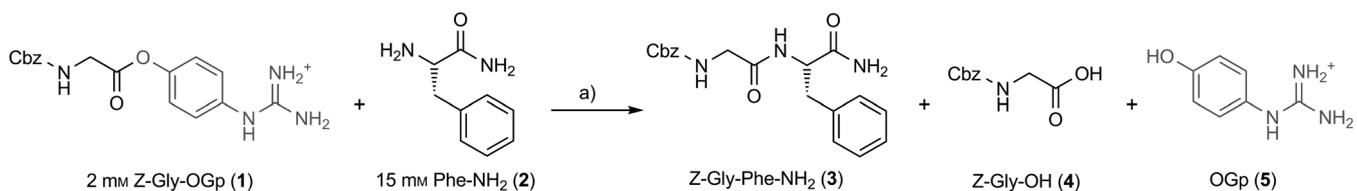
known to result in a better synthesis-over-hydrolysis (S/H) ratio than serine proteases.^[14] The slight preference of papain for arginine and lysine suggested that the existing OGp ester could be used as a mimetic. Furthermore, papain has a broader specificity than trypsin, and this could provide relevant information about the applicability of the substrate mimetics approach to less specific enzymes. So far only proteases that bear a narrow specificity towards distinct amino acid side chains have been used; hence this research could potentially widen the scope of the substrate mimetics approach.

We used the OGp ester as a potential substrate mimetic for papain-induced dipeptide synthesis under aqueous conditions. Subsequent docking studies provided insight into the binding mode, and, based on these results, MD simulations served to predict a set of suitable amino acid donors. The resulting ranking was experimentally verified, and the scope of the amino acid acceptor was determined.

Results and Discussion

Initial dipeptide synthesis reaction

Z-Gly-OGp (**1**) was prepared as a test substrate in order to study the OGp ester as a mimetic applied to papain. Z-Gly-OH was esterified with *p*-[*N,N'*-di(Boc)guanidino]phenol^[15] by using *N,N'*-dicyclohexylcarbodiimide (DCC) as coupling reagent. Glycine was chosen to be incorporated in the acyl donor because it is the simplest amino acid and is not recognized by papain. Next, substrate **1** was subjected to the enzymatic reaction with papain in which H-Phe-NH₂ was used as the acyl acceptor because of its clear UV visibility at 254 nm, which simplifies the HPLC analyses (Scheme 2). The dipeptide was formed quickly (20 min) and in high yield (92%) with hardly



Scheme 2. Enzymatic synthesis with papain. a) 3.5 mM papain, 2 mM pTSA, 10% (v/v) DMF, 0.2 M HEPES buffer (pH 8), 0.2 M NaCl, 20 mM CaCl₂.

any enzymatic hydrolysis (2.4%). This compares favorably to results obtained with trypsin.^[9a] Concurrently, we observed that chemical hydrolysis of the starting material under exactly the same reaction conditions but in the absence of papain was substantial (5.6%). This spontaneous hydrolysis is clearly undesirable, because it will increase with longer reaction times.

Prediction of the binding mode and validation of Z- X_{AA} -OGp esters

Having shown that papain can recognize Z-Gly-OGp (**1**), we developed an increased molecular understanding of the substrate mimetics approach in papain. A molecular-modeling study was performed by using the flexible docking program Fleksy.^[16] The results were visualized and analyzed by using the YASARA program.

Crystallization and subsequent structure determination by X-ray methods revealed that papain is a single-chain polypeptide containing three disulfide bridges, which is folded into two domains with the active site in a groove between them.^[17] The most important residues in the active site are Cys25, which provides the nucleophilic thiol, and His159, which completes the catalytic diad (Figure 1A). The oxyanion hole is formed by the side chain NH_2 group of Gln19 and the backbone amide of Cys25. Residues Asp158, Gly66, and Trp177 are involved in a conserved hydrogen-bonding network that is required for substrate affinity.^[18]

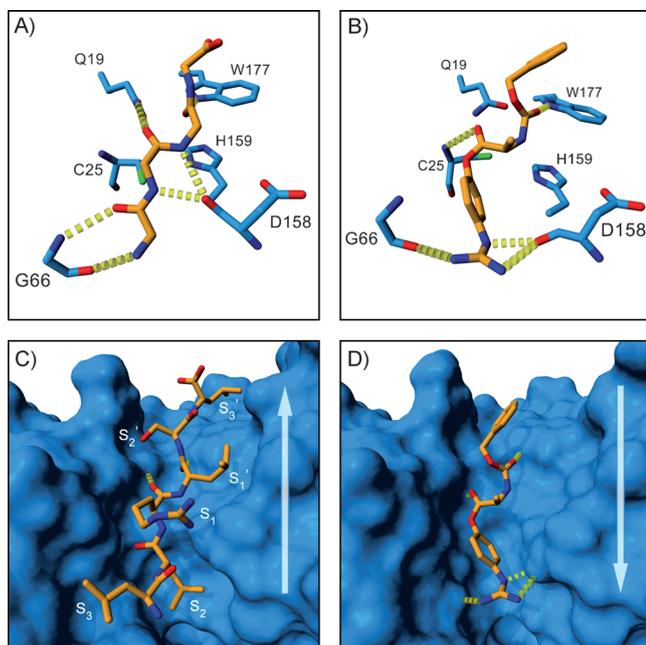


Figure 1. Molecular modeling of papain substrates. Hydrogen bonding interactions to functionally important amino acids in the papain active site are shown for A) the backbone of a peptide substrate and B) the compound Z-L-Ala-OGp. The overall orientation in the papain peptide binding groove is shown for C) the hexapeptide LLRLSL and D) the compound Z-L-Ala-OGp. The different papain subsites occupied by the hexapeptide are indicated. The arrows indicate the N- to C-terminal direction of the scissile bond, thus illustrating the “reversed” binding mode of Z-L-Ala-OGp.

To gain insight into the influence of the configuration of the amino acids and sterically demanding substrates, in addition to the achiral glycine ester, the OGp esters of L-alanine (**1A**), D-alanine (**1a**), β -alanine (**1 β A**) and L-proline (**1P**) were considered in the docking study. Figure 1B shows that Z-L-Ala-OGp perfectly fits in the active site, interacting in a similar manner to the natural substrate (Figure 1A). Based on the obtained Z-L-Ala-OGp binding mode, we rationalized that the oppositely configured D-alanine ester would be a much poorer substrate, as its methyl side chain is predicted to clash with the catalytic His159. This effect is even more profound and disadvantageous for the secondary amino acid L-proline. In the case of the β -alanine ester, the interactions with the oxyanion hole and the hydrogen bonding network with the guanidino group could be maintained, but the beneficial interaction with Trp177 was disrupted by the extra carbon atom in the ester backbone. Experimental verification of these modeling-based hypotheses confirmed that reaction rates for **1a** and **1 β A** are indeed around 100 times lower than those for **1** and **1A**, and that **1P** does not react at all. This all provides experimental support for the binding mode of **1A** proposed in Figure 1B.

Figure 1C shows that the natural substrate is positioned in the groove with all the amino acid side chains, including the specificity-determining arginine, exposed. The substrate mimetic is also located in this cleft, but is in the reverse orientation (Figure 1D), similar to what Bordusa described for the OGp ester in trypsin.^[5] It is remarkable though, that the guanidino group present in the mimetic is situated in the groove, unlike the natural substrate’s arginine side chain. This suggests that, in the case of papain, the OGp ester is not actually mimicking the natural substrate in the sense that the guanidino group binds at the same position. Nonetheless, the OGp ester is specifically recognized by the enzyme, and therefore we propose to call it an enzyme-specific activating ester.

Molecular-dynamics simulations of Z- X_{AA} -OGp esters in papain

To further probe the viability of enzyme-specific activating esters as a more general enzymatic method for peptide synthesis, we investigated whether the successful dipeptide synthesis with Z-Gly-OGp and Phe- NH_2 could be extended to a variety of other natural and unnatural amino acids. Figure 1D clearly shows that the side chain of Z-L-Ala-OGp points up and out of the groove, thereby implying there should be sufficient space to accommodate the side chains of other natural amino acids as well. Rather than determining the scope of the acyl donor directly, we incorporated a computational approach as an intermediate step. A reliable model would be of great use in predicting which amino acids will be accepted by papain in order to prioritize them for synthesis and for reaction planning. All 20 natural and the two previously evaluated unnatural amino acids were docked, and the stabilities of the resulting complexes were assessed with the help of molecular-dynamics simulations.

This was done on the assumption that the aforementioned interactions with the oxyanion hole are essential to the forma-

tion of a productive enzyme–substrate complex. Both the distance to oxyanion-hole residue Gln19 and the percentage of simulation time for which this residue forms a hydrogen bond with the substrate were used as measures of complex stability. Initially the oxyanion-hole Cys25 was also taken into account, but these interactions did not affect the results. The ranking is mostly determined by the specific interactions that the side chains make with papain during the simulation (Figure 2). As a

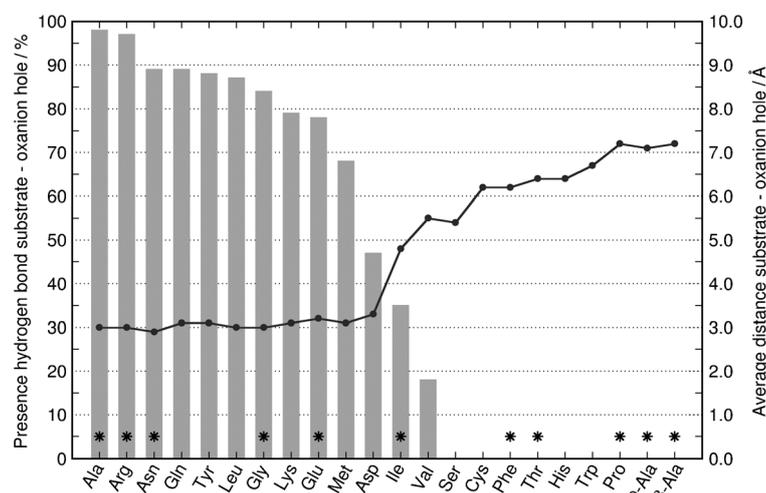


Figure 2. Stability of modeled papain-Z-X_{AA}-OGp complexes as assessed by molecular dynamics simulations. The presence of a hydrogen bond between Gln19 of the papain oxyanion hole and carbonyl of the scissile bond of the substrate throughout the simulation time is shown by bars (left axis), and the average distance between the oxyanion hole (Gln19) and carbonyl of the scissile bond of the substrate throughout the simulation is shown by a line (right axis). The amino acids that were selected for synthesis and testing are indicated with asterisks. All amino acids were considered in the L-conformation, unless otherwise indicated.

result, amino acids with similar characteristics generally show a comparable ranking. For instance, asparagine and glutamine are positioned close to each other, as are the β-branched amino acids isoleucine and valine. The observed discrepancy between phenylalanine and tyrosine can be explained by the fact that, although the docking showed a nice fit, phenylalanine is drawn into the nearby hydrophobic S₁' pocket during the MD simulation, an event that is probably prevented for tyrosine by the presence of the polar hydroxyl group. Based on the docking studies alone, one would expect a higher ranking for serine, but this is presumably prohibited by the formation of an intramolecular hydrogen bond during the simulation.

Verification of the MD-based predictions

To verify the MD-based rankings and to determine the scope of the acyl donor experimentally, a representative set of amino acids was selected for validation (marked with an asterisk in Figure 2). The corresponding Z-X_{AA}-OGp esters were synthesized from amino acids with appropriate side-chain protection. Yields of the coupling reactions varied between 63 and 98%. The acidic deprotection appeared to be troublesome in some cases, because chemical hydrolysis of the ester occurred as a

side reaction (<10%). Nevertheless, the crude product obtained after Boc deprotection was used directly in the enzymatic reactions, and a correction was performed afterwards. The enzymatic reaction was monitored for three hours. The identity of the products formed in the enzymatic reaction was confirmed by chemical synthesis of reference compounds and LC-MS analysis. Table 1 presents either the time to reach full conversion of the OGp esters, or the conversion after three hours. The indicated percentages of enzymatic synthesis and hydrolysis remained constant over time, as measured after 24 h unless stated otherwise. The S/H ratio for the various amino acids gives an ambiguous impression. While excellent in the case of glycine and threonine (entries 4 and 8), proline (entry 9) does not react with papain at all. With the OGp esters of both arginine and asparagine (entries 2 and 3), nonenzymatic cyclic side products were formed, a piperidone^[19] and succinimide,^[20] respectively; this was facilitated by the guanidinophenyl leaving group. Over time, the piperidone side product was converted to Z-Arg-OH, while the succinimide side product concentration remained constant.

It is not trivial to assess the correlation between the computational results in Figure 2 and the experimental results in Table 1. Given the large variation in the S/H ratios for the various amino acids, the analysis should be restricted to conversion rates only. This can be rationalized by realizing that the simulations address only the effective formation and stability of the enzyme–substrate complex, which is merely the first step in a cascade of events. The stability of the acyl–enzyme intermediate as well as the velocity of deacylation are important factors in defining the S/H ratio, but cannot be predicted from our modeling

Table 1. The various Z-AA-OGp esters tested experimentally.^[a]

$\text{Z-X}_{\text{AA}}\text{-OGp} + \text{Phe-NH}_2 \xrightarrow[3.5 \mu\text{M papain}]{\text{HOGp}} \text{Z-X}_{\text{AA}}\text{-Phe-NH}_2 + \text{Z-X}_{\text{AA}}\text{-OH}$						
Amino acid	t [min]	Conv. [%]	Background Z-X _{AA} -OH [%]	Enzymatic Z-X _{AA} -Phe-NH ₂ [%]	Z-X _{AA} -OH [%]	
1	L-Ala	2	100	1.9	77.5	20.6
2	L-Arg	1	100	3.4	45.9 ^[b]	17.2
3	L-Asn	25	100	1.4	6.1 ^[c]	30.2
4	Gly	20	100	5.6	92.0	2.4
5	L-Glu	15	100	3.8	68.5	27.7
6	L-Ile	90	100	–	24.5 ^[d]	12.2
7	L-Phe	25	100	3.4	29.8	66.8
8	L-Thr	20	100	4.3	90.7	5.0
9	L-Pro	180	5	5.0	–	–
10	D-Ala	180	82	13.5	11.6	5.4
11	β-Ala	180	39	7.4	31.6	–

[a] Conditions: 2 mM Z-X_{AA}-OGp, 15 mM Phe-NH₂, 0.2 M HEPES buffer (pH 8.0), 0.2 M NaCl, 20 mM CaCl₂, 10% (v/v) DMF. [b] Nonenzymatic piperidinone side-product formation (33.5%). [c] Nonenzymatic succinimide side-product formation (62.3%). [d] Dipeptide product precipitates during reaction, yield is probably 87.8%

studies. For example, during deacylation, unforeseen steric hindrance might occur between the acyl donor residue and the groove of the active center in papain. Furthermore, the type of nucleophile that attacks the enzyme–acyl intermediate is not taken into account in our modeling experiments.

Despite the aforementioned shortcomings, it is clear that the high-ranking amino acids require the shortest reaction times, and that the reaction time generally increases with a decreasing predicted stability. However, phenylalanine does not seem to fit the correlation, with a much faster experimentally determined reaction time than suggested by its ranking. Apparently the hydrophobic pocket that determines the outcome of the MD simulation is of less influence on substrate mimetic binding than expected. Threonine also performed much better than expected from our simulations. To explain this observation, we evaluated the different rotamers accessible to the threonine side chain. In the docking experiment, it was positioned in the *trans* χ_1 rotamer, which is not the typically preferred rotamer for threonine.^[21] Therefore, we also assessed the stability of the more preferred *gauche* χ_1 rotamers and found that the *g* rotamer indeed resulted in a more stable complex, with an oxyanion hole–substrate hydrogen bond present for 41% of the simulation time. This is comparable to what is observed in Figure 2 for the other β -branched amino acids isoleucine and valine, which were both also predicted to bind in their preferred *g* and *trans* rotamers.

It would not have been possible to decide a priori what percentage of hydrogen bond formation would be sufficient for activity. With the experimental results in hand, we can see that only a minor percentage of hydrogen bond formation and proximity is already enough to lead to enzymatic activity. Basically every experimentally tested amino acid in Table 1, except for proline, is accepted by papain. Even the two unnatural amino acids *D*-alanine and β -alanine, correctly predicted to react considerably more slowly, do react. The difference between predicted and experimentally determined reactivity may be taken as an indication that papain is more flexible than modeling suggests.

Probing the influence of the amino acid in the nucleophile–acyl acceptor

Contrary to the acyl donor amino acids, the scope of the nucleophile was not computationally studied. As mentioned previously, papain displays broad specificity as a result of its wide-open peptide binding groove, which allows for many interactions upon substrate binding. The acyl donor can be modeled in this network by utilizing known interactions with the oxyanion hole, but no such requirement is known for the incoming nucleophile. Hence, the influence of the acyl acceptor on the enzyme-specific activation was only probed experimentally with a restricted set of amino acid derivatives (Table 2).

Switching the configuration from *L*-Phe-NH₂ to *D*-Phe-NH₂ shifted the enzymatic reaction from primarily synthesis to mainly hydrolysis (entry 1). The exchange of an amide for a *tert*-butyl ester did not really affect the reaction time, although the dipeptide product precipitated (entry 2). The *p*NA-amide of

Table 2. The various nucleophiles tested experimentally.^[a]

$\text{Z-Gly-OGp} + \text{Nu} \xrightarrow[3.5 \mu\text{M papain}]{\text{HOGp}} \text{Z-Gly-Nu} + \text{Z-Gly-OH}$					
Nucleophile	t [min]	Conv. [%]	Background Z-Gly-OH [%]	Enzymatic Z-Gly-Nu [%]	Z-Gly-OH [%]
1 <i>D</i> -Phe-NH ₂	20	100	2.8	22.9	74.3
2 <i>L</i> -Phe-OtBu	25	100	3.8	11.8 ^[b]	–
3 <i>L</i> -Tyr-NH ₂	25	100	6.2	91.3	2.5
4 <i>L</i> -Tyr- <i>p</i> NA	25	100	5.1	90.6	4.3
5 <i>L</i> -Ala-NH ₂	20	100	5.3	87.0	7.7
6 <i>L</i> -Ala- <i>p</i> NA	20	100	4.0	96.0	–
7 <i>L</i> -Ser- <i>p</i> NA	30	100	6.0	94.0	–
8 <i>L</i> -Pro- <i>p</i> NA	30	100	9.7	–	90.3
9 –	20	100	5.2	–	94.8

[a] Conditions: 2 mM Z-X_{AA}-OGp, 15 mM Phe-NH₂, 0.2 M HEPES buffer (pH 8.0), 0.2 M NaCl, 20 mM CaCl₂, 10% (v/v) DMF. [b] Dipeptide product precipitated.

phenylalanine did not dissolve under these reaction conditions and was therefore excluded from the set. As alternatives, the Tyr-NH₂ and Tyr-*p*NA pair were tested for the influence of the *p*NA group, prior to testing the less visible amino acids with this chromophore attached. The results were very much comparable, with or without *p*NA, also for the small amino acids alanine and serine. Proline, a worse nucleophile, only gave hydrolysis and that at a comparable rate to the one seen without an additional nucleophile being present.

Conclusions

In this paper, we have shown that the OGp ester can be successfully applied in papain-catalyzed dipeptide synthesis under aqueous conditions. Our docking studies, which were performed to increase our molecular understanding of the system, resulted in an unexpected binding mode for the Z-*L*-Ala-OGp ester, which is supported by experimentally observed structure–activity relationships. In contrast to the anticipated function of a substrate mimetic, the OGp ester showed an unprecedented enzyme-specific activation in papain. Prior to determining the scope of the acyl donor experimentally, we used a molecular-dynamics-simulations approach to prioritize 22 natural and unnatural amino acids for synthesis. The resulting ranking was in good agreement with the experimental data. A representative set of Z-X_{AA}-Phe-NH₂ dipeptides was obtained in moderate to excellent yields. The scope of the incoming nucleophile was relatively broad, ranging from the small Ala-NH₂ to the much larger Tyr-*p*NA. Since the OGp ester exhibits rather unexpected enzyme-specific activation, we hypothesize that it is not strictly required to have an arginine-like activating ester. Future investigations will be directed towards confirming the predicted binding mode of the OGp ester by using X-ray studies and to replacing it by a simpler moiety in order to render the described enzymatic approach to peptide-bond formation even more accessible.

Experimental Section

Synthesis: see the Supporting Information for a detailed description of the synthetic procedures and product characterization.

***p*-[*N,N'*-Di(Boc)guanidino]phenol (7):** *N,N'*-Di(Boc)-*S*-methylisothiourea (2.90 g, 10.0 mmol, 1 equiv) and *p*-aminophenol (1.64 g, 15.0 mmol, 1.5 equiv) were dissolved in dry THF (60 mL), and this mixture was cooled to 0 °C before HgCl₂ (2.99 g, 11.0 mmol, 1.1 equiv) was added. After the mixture had been stirred for 20 min under argon, its temperature was raised to 25 °C, and it was stirred for 20 h. The white precipitate that was formed during the reaction was filtered off, and the filtrate was concentrated under reduced pressure. Recrystallization from methanol yielded 1.53 g (43%) of the pure product. The mother liquor was then evaporated to dryness, and the remaining solid was purified by column chromatography (EtOAc/heptane, 1:9→1:2) to afford **7** (753 mg, 65% yield) as an off-white solid. *R*_f = 0.37 (EtOAc/heptane 1:2); m.p. decomp. > 240 °C; ¹H NMR (CDCl₃, 300 MHz): δ = 11.61 (s, NH), 10.04 (s, NH), 7.14 (d, *J* = 8.6 Hz, 2H), 6.68 (d, *J* = 8.3 Hz, 2H), 6.22 (brs, OH), 1.54 (s, 9H), 1.46 (s, 9H); ¹³C NMR (CDCl₃, 75 MHz): δ = 163.2, 155.9, 155.3, 153.2, 127.0, 126.3, 116.1, 83.7, 79.8, 28.1; IR (film): $\tilde{\nu}$ = 3264, 2979, 2737, 1720, 1647, 1517, 1409, 1227, 1152, 1112, 1059 cm⁻¹; HRMS (ESI) *m/z* calcd for C₁₇H₂₆N₃O₅: 352.1873 [*M*+H]⁺, found: 352.1875.

General procedure for the DCC coupling of Cbz-protected amino acids with phenol 7 as illustrated by the synthesis of *N'*-Cbz-glycine *p*-[*N,N'*-di(Boc)guanidino]phenyl ester (8G): DCC (867 mg, 4.21 mmol, 1.4 equiv) was added slowly to a cooled (0 °C) solution of *Z*-Gly-OH (879 mg, 4.20 mmol, 1.4 equiv), *p*-[*N,N'*-di(Boc)guanidino]phenol (**7**, 1.05 g, 3.00 mmol, 1 equiv), and *p*-(dimethylamino)pyridine (73 mg, 0.61 mmol, 0.2 equiv) in EtOAc (10 mL). The reaction mixture was stirred at 0 °C for 1 h and for an additional 2 h at room temperature. The solid *N,N'*-dicyclohexylurea was filtered off, and the solvent was evaporated in vacuo. The product was obtained as a white solid after purification by column chromatography (1→4% MeOH in CH₂Cl₂). Yield: 1.33 g (82%); *R*_f = 0.67 (4% MeOH in CH₂Cl₂); m.p. 111 °C; ¹H NMR (CDCl₃, 300 MHz): δ = 11.62 (s, NH), 10.35 (s, NH), 7.62 (d, *J* = 8.5 Hz, 2H), 7.41–7.28 (m, 5H), 7.07 (d, *J* = 8.3 Hz, 2H), 5.34 (m, NH), 5.15 (s, 2H), 4.22 (d, *J* = 5.3 Hz, 2H), 1.53 (s, 9H), 1.50 (s, 9H); ¹³C NMR (CDCl₃, 75 MHz): δ = 168.6, 163.4, 156.2, 153.5, 153.3, 146.8, 136.1, 134.8, 128.6, 128.2, 128.1, 123.2, 121.6, 83.9, 79.8, 67.3, 42.9, 28.2, 28.1; IR (film): $\tilde{\nu}$ = 2978, 2928, 1779, 1720, 1640, 1508, 1412, 1240, 1154, 1114, 1057 cm⁻¹; HRMS (ESI) *m/z* calcd for C₂₇H₃₄N₄NaO₈: 565.2274 [*M*+Na]⁺, found: 565.2274.

General procedure for the acidic Boc deprotection of guanidino-phenyl esters (8) with TFA as illustrated by the synthesis of *N'*-Cbz-Glycine *p*-guanidinophenyl ester (1G): Boc-protected compound **8** (100 mg) was dissolved in CH₂Cl₂ (2 mL), and TFA (0.5 mL) was added. The reaction mixture was stirred at room temperature overnight. The solvents were removed under reduced pressure and co-evaporated with *t*BuOH (3 × 10 mL). The obtained oil was lyophilized from H₂O/dioxane (10 mL) in the presence of HCl (2 M, 0.5 mL) to give the product as a sticky oil (quant.). *R*_f = 0.50 (CHCl₃/MeOH/NH₄OH, 65:45:20); ¹H NMR ([D₆]DMSO, 300 MHz): δ = 10.03 (s, NH), 7.89 (t, *J* = 5.9 Hz, NH), 7.56 (brs, 4NH), 7.40–7.25 (m, 7H), 7.21–7.15 (m, 2H), 5.08 (s, 2H), 4.07 (d, *J* = 6.1 Hz, 2H); ¹³C NMR ([D₆]DMSO, 75 MHz): δ = 169.0, 156.5, 156.0, 148.2, 136.8, 132.9, 128.3, 127.8, 127.7, 125.7, 122.7, 65.6, 42.4; IR (film): $\tilde{\nu}$ = 3309, 3166, 2950, 1770, 1706, 1671, 1629, 1588, 1507, 1455, 1280, 1201, 1166, 1053 cm⁻¹; HRMS (ESI) *m/z* calcd for C₁₇H₁₈N₄NaO₄: 365.1226 [*M*+Na]⁺, found: 365.1231.

General procedure for the enzymatic reactions: Enzymatic acyl-transfer reactions were performed at 25 °C in a total volume of 375 μL containing HEPES buffer (0.2 M, pH 8.0), NaCl (0.2 M), CaCl₂ (20 mM), 10% DMF and *p*-toluenesulfonic acid (pTSA; 2 mM) as an internal standard. Stock solutions of *Z*-X_{AA}-OGp esters (50 mM) in DMF and nucleophiles (30 mM) in buffer were prepared. The final concentrations of acyl donor and acyl acceptor were 2 and 15 mM, respectively. The latter was calculated as the free, *N'*-unprotonated nucleophile concentration [HN]₀ according to the Henderson–Hasselbalch Equation:

$$[\text{HN}]_0 = [\text{N}]_0 / (1 + 10^{\text{pK} - \text{pH}})$$

Papain (4 mg) was activated before use by adding dithiothreitol (0.6 mg) and phosphate buffer (1 mL, 0.1 M, pH 6.5) containing EDTA (2.5 mM) and shaking the mixture for 10 min at 25 °C. This solution was stored on ice and was freshly prepared daily. Following thermal equilibration of assay mixtures, the enzymatic reactions were started by the addition of papain at a final concentration of 3.5 μM. Blank reactions with Milli-Q instead of papain were run in parallel. From this control experiment, the spontaneous ester hydrolysis could be determined, as well as nonenzymatic aminolysis of the acyl donor esters; the latter could be ruled out. At regular intervals, aliquots (20 μL) were withdrawn and quenched with glacial acetic acid (20 μL). The reactions were monitored for 3 h by HPLC and checked once more for changes in reaction mixture composition after 24 h. The values reported are the average of at least two separate experiments. The identity of the peptide products was established by chemical synthesis of reference compounds and LC-MS.

HPLC analyses: Samples were analyzed on a Shimadzu LC 2010 analytical HPLC system equipped with a RP C18 column (Varian, Inertsil ODS-3, 5 μm, 150 × 4.6 mm) and eluted with various mixtures of acetonitrile/water containing 0.1% trifluoroacetic acid under isocratic and gradient conditions at a flow rate of 1.0 mL min⁻¹. The wavelength of detection was 254 nm. Product yields were calculated from peak areas of the substrate esters and the hydrolysis and aminolysis products.

Molecular modeling of the papain–peptide complex: The molecular model of papain bound to the hexapeptide LLRLSL was constructed on the basis of the crystal structure of a papain–leupeptin complex (PDB ID: 1POP) solved at 2.1 Å resolution.^[17] This structure contains an LLR peptide bound only to the *S* subsites. To gain more insight into peptide binding to the *S'* subsites, a hybrid model was built by using an LSL peptide fragment bound to the *S'* subsites of another papain crystal structure (PDB ID: 2CIO) solved at 1.5 Å resolution.^[22] First, the two crystal structures were aligned by using the MOTIF algorithm,^[23] after which the coordinates of the LSL peptide were transferred to the papain–leupeptin complex. Subsequently, a peptide bond between the LLR and LSL peptide fragment was added manually by using the YASARA program^[24] and finally the resulting complex was energy minimized by using the YASARA2 force field.^[25]

Molecular docking of papain substrates: All molecular-docking studies in papain were performed by using the flexible docking program Fleksy.^[16] The crystal structure of a papain–leupeptin complex (PDB ID: 1POP) solved at 2.1 Å resolution^[17] was used as the receptor structure. The structure was prepared for docking by removing leupeptin and all water molecules from the complex. Subsequently, hydrogen atoms were added to the structure, and their positions were optimized by using the YASARA program.^[24] In the applied docking protocol only those docking poses in which the

scissile bond of the docked Z-X_{AA}-OGp compound was aligned with the scissile bond of the natural peptide substrate were taken forward. Otherwise, default parameters as described previously^[16] were applied.

Molecular dynamics simulations: The highest-ranking docking poses obtained from the docking studies were used as starting complexes for MD simulations with the Yasara program.^[24] The complex was first solvated in a simulation cell two times 10 Å larger than the protein along each axis, then the cell was neutralized by replacing water molecules with counter ions. The resulting system was first minimized with the Amber03 force field^[26] by using a 7.86 Å force cutoff and the Particle Mesh Ewald algorithm to treat long-range electrostatic interactions. Simulated annealing was used (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached. Subsequently, 3 ns MD simulations were conducted at 298 K for each of the 22 substrates with periodic boundary conditions and 1.25 fs time steps. Intermolecular forces were recalculated every two simulation steps, and pressure control was employed to maintain a water density of 0.997 g cm⁻³.

Acknowledgements

This research was performed under the ACTS–NWO Integration of Biosynthesis & Organic Synthesis (IBOS) program. DSM N.V. (Geleen, the Netherlands) is gratefully acknowledged for financial support. S.B.N. is supported by the Netherlands Organization for Scientific Research (NWO) through a VENI grant (700.58.410).

Keywords: enzymatic peptide synthesis • esters • molecular dynamics • papain • substrate mimetics

- [1] a) K. Adermann, *Curr. Protein Pept. Sci.* **2005**, *6*, 205–206; b) H. Korhonen, A. Pihlanto, *Curr. Pharm. Des.* **2003**, *9*, 1297–1308; c) H. Korhonen, A. Pihlanto, *Curr. Pharm. Des.* **2007**, *13*, 829–843; d) M. B. Roberfroid, *J. Nutr.* **2007**, *137*, 2493–2502; e) B. Steffansen, C. U. Nielsen, S. Frokjaer, *Eur. J. Pharm. Biopharm.* **2005**, *60*, 241–245; f) J. Torres-Llanez, B. Vallejo-Cordoba, A. F. González-Córdova, *Arch. Latinoam. Nutr.* **2005**, *55*, 111–117; g) P. Vlieghe, V. Lisowski, J. Martinez, M. Khrestchatsky, *Drug Discovery Today* **2010**, *15*, 40–56.
- [2] a) R. Joshi, L. Meng, H. Eckstein, *Helv. Chim. Acta* **2008**, *91*, 983; b) H. Xiang, G. Y. Xiang, Z. M. Lu, L. Guo, H. Eckstein, *Amino Acids* **2004**, *27*, 101.
- [3] V. Schellenberger, H.-D. Jakubke, N. P. Zapevalova, Y. V. Mitin, *Biotechnol. Bioeng.* **1991**, *38*, 104–108.
- [4] F. Bordusa, D. Ullmann, C. Elsner, H. D. Jakubke, *Angew. Chem.* **1997**, *109*, 2583–2585; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2473–2475.

- [5] M. Thormann, S. Thust, W. H. Hofmann, F. Bordusa, *Biochemistry* **1999**, *38*, 6056–6062.
- [6] a) K. Itoh, H. Sekizaki, E. Toyota, N. Fujiwara, K. Tanizawa, *Bioorg. Chem.* **1996**, *24*, 59–68; b) H. Sekizaki, K. Itoh, A. Shibuya, E. Toyota, K. Tanizawa, *Chem. Pharm. Bull.* **2007**, *55*, 1514–1517; c) H. Sekizaki, K. Itoh, E. Toyota, K. Tanizawa, *Chem. Pharm. Bull.* **1998**, *46*, 846–849; d) H. Sekizaki, K. Itoh, E. Toyota, K. Tanizawa, *Amino Acids* **1999**, *17*, 285–291.
- [7] R. Günther, S. Thust, H. J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, *267*, 3496–3501.
- [8] R. Günther, A. Stein, F. Bordusa, *J. Org. Chem.* **2000**, *65*, 1672–1679.
- [9] a) R. Grünberg, I. Domgall, R. Günther, K. Rall, H. J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, *267*, 7024–7030; b) S. Xu, K. Rall, F. Bordusa, *J. Org. Chem.* **2001**, *66*, 1627–1632.
- [10] a) T. Fuchise, H. Kishimura, H. Sekizaki, Y. Nonami, G. Kanno, S. Klomklo, S. Benjakul, B.-S. Chun, *Food Chem.* **2009**, *116*, 611–616; b) H. Sekizaki, E. Toyota, T. Fuchise, H. Zhou, Y. Noguchi, K. Horita, *Amino Acids* **2008**, *34*, 149–153; c) E. Toyota, D. Iyaguchi, H. Sekizaki, M. Tateyama, K. K. S. Ng, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2009**, *65*, 717–723.
- [11] a) N. Wehofskey, M. Haensler, S. W. Kirbach, J.-D. Wissmann, F. Bordusa, *Tetrahedron: Asymmetry* **2000**, *11*, 2421–2428; b) N. Wehofskey, S. W. Kirbach, M. Haensler, J.-D. Wissmann, F. Bordusa, *Org. Lett.* **2000**, *2*, 2027–2030.
- [12] N. Wehofskey, C. Wespe, V. Cerovsky, A. Pech, E. Hoess, R. Rudolph, F. Bordusa, *ChemBioChem* **2008**, *9*, 1493–1499.
- [13] a) N. Wehofskey, F. Bordusa, *FEBS Lett.* **1999**, *443*, 220–224; b) N. Wehofskey, J.-D. Wissmann, M. Alisch, F. Bordusa, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **2000**, *1479*, 114–122.
- [14] M. Philipp, M. L. Bender, *Mol. Cell. Biochem.* **1983**, *51*, 5–32.
- [15] S. Thust, B. Koksche, *J. Org. Chem.* **2003**, *68*, 2290–2296.
- [16] S. B. Nabuurs, M. Wagener, J. de Vlieg, *J. Med. Chem.* **2007**, *50*, 6507–6518.
- [17] E. Schroder, C. Phillips, E. Garman, K. Harlos, C. Crawford, *FEBS Lett.* **1993**, *315*, 38–42.
- [18] P. J. Berti, C. H. Faerman, A. C. Storer, *Biochemistry* **1991**, *30*, 1394–1402.
- [19] L. Zervas, T. T. Otani, M. Winitz, J. P. Greenstein, *J. Am. Chem. Soc.* **1959**, *81*, 2878–2884.
- [20] T. Geiger, S. Clarke, *J. Biol. Chem.* **1987**, *262*, 785–794.
- [21] R. J. Dunbrack, M. Karplus, *Nat. Struct. Biol.* **1994**, *1*, 334–340.
- [22] M. S. Alphey, W. N. Hunter, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2006**, *62*, 504–508.
- [23] G. Vriend, C. Sander, *Proteins* **1991**, *11*, 52–58.
- [24] E. Krieger, G. Koraimann, G. Vriend, *Proteins Struct. Funct. Bioinf.* **2002**, *47*, 393–402.
- [25] a) E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, G. Vriend, *Proteins Struct. Funct. Bioinf.* **2004**, *57*, 678–683; b) E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker, K. Karplus, *Proteins Struct. Funct. Bioinf.* **2009**, *77*, 114–122.
- [26] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.* **2003**, *24*, 1999–2012.

Received: April 22, 2011

Published online on August 8, 2011