STRUCTURAL
BIOLOGY

Volume 78 (2022)

Supporting information for article:

Structure and molecular dynamics of prolyl oligopeptidase from *Microbulbifer arenaceous* provide insights into catalytic and regulatory mechanisms

Ping Huang, Ang Lv, Qiaojuan Yan, Zhengqiang Jiang and Shaoqing Yang

Table S1 Primers used in this study

Primers	Oligonucleotide sequence (5' to 3')[a]	Description
POP-F1	ACCATGGGCAGCAGCACC GATACCCAGAGTTCG	Forward primer for cloning no signal peptide *Ma*POP (*nt* 69-2133) into pET15b via LIC
POP-R1	AAGGGGTTATGCTAGTTAG TGATGATGATGATG	Reverse primer for cloning no signal peptide *Ma*POP (*nt* 69-2133) into pET15b with a C-terminal His_6 tag via LIC
pET15b-F1	CTAGCATAACCCCTTGGGG CC	Forward primer for cloning pET15b backbone to construct pET15b-*mapop* via LIC
pET15b-R1	GCTGCTGCCCATGGTATAT CTCC	Reverse primer for cloning pET15b backbone to construct pET15b-*mapop* via LIC
15bPOP-F	GGTGATGCCGGCCACGAT GCG	Forward primer for sequencing pET15b-*mapop* constructs
15bPOP-R	TTATGCCGGTACTGCCGGG CC	Reverse primer for sequencing pET15b-*mapop* constructs
S562A-F	GGCATCATGGGCCGCgCCA ACGGCGGTCTGC	Forward primer for mutating S562 codon of *Ma*POP (TCC) to Ala (GCC)
S562A-R	cGCGGCCCATGATGCCGAG CTTTTCCGGTG	Reverse primer for mutating S562 codon of *Ma*POP (TCC) to Ala (GCC); Primer for sequencing the D403G and D403C mutations
D646A-F	CCACCGCCGATCGCGcTGA CCGCGTGGTGCC	Forward primer for mutating D646 codon of *Ma*POP (GAT) to Ala (GCT)
D646A-R	gCGCGATCGGCGGTGGTGA TCAAGGTAGCC	Reverse primer for mutating D646 codon of *Ma*POP (GAT) to Ala (GCT)

H681F-F	GAAACCCGTGCGGGCttCG GTGCCGGCAAGCC	Forward primer for mutating H681 codon of *Ma*POP (CAC) to Phe (TTC)
H681F-R	aaGCCCGCACGGGTTTCCA CAGCGAGGTAG	Reverse primer for mutating H681 codon of *Ma*POP (CAC) to Phe (TTC)
Triad-R	GGATATCCGGATATAGTTC CTCC	Primer for sequencing S562A, D646A, and H681F mutations
K106C-F	CGCTACGGCGTGCCCTATtg tAAAGCCGGTC	Forward primer for mutating K106 codon of *Ma*POP (AAA) to Cys (TGT)
K106C-R	acaATAGGGCACGCCGTAG CGCTCGTACTGC	Reverse primer for mutating K106 codon of *Ma*POP (AAA) to Cys (TGT)
D403C-F	CACTGGCGACCCCGATtgtC CGGAAACCTAC	Forward primer for mutating D403C codon of *Ma*POP (GAC) to Cys (TGT)
D403C-C	acaATCGGGGTCGCCAGTGA AACCCTCTACG	Reverse primer for mutating D403C codon of *Ma*POP (GAC) to Cys (TGT)
Y98C-F	CTTGACGGAGCTGTGGCAG TgCGAGCGCTACG	Forward primer for mutating Y98C codon of *Ma*POP (TAC) to Cys (TGC)
Y98C-R	cACTGCCACAGCTCCGTCA AGCGGTCATTGAT	Reverse primer for mutating Y98C codon of *Ma*POP (TAC) to Cys (TGC)
Y437C-F	ATAGTTAAATCGCCGAAAT gTCCTGCGGATT	Forward primer for mutating Y437C codon of *Ma*POP (TAT) to Cys (TGT)
Y437C-R	cATTTCGGCGATTTAACTA TTTCGCTGTTAC	Reverse primer for mutating Y437C codon of *Ma*POP (TAT) to Cys (TGT)
D403G-F	CACTGGCGACCCCGATGgC CCGGAAACCTAC	Forward primer for mutating D403G codon of *Ma*POP (GAC) to Cys (GGC)
D403G-R	cCATCGGGGTCGCCAGTGA AACCCTCTACG	Reverse primer for mutating D403G codon of *Ma*POP (GAC) to Cys (GGC)

T7	TAATACGACTCACTATAGG	Primer for sequencing Y98C and K106C mutations
D263A-F	GCCGATGGTAAATATTTGATTCTGGGAG	Primer for sequencing Y437C mutation

[a] Red lowercase letters designate the mutated nucleotides.

Table S2 Hydrogen bonds between the N-terminal region (residues 29-98) and C-terminal region (residues 436-709) in the α/β-hydrolase domain of chain A[a]

Atoms of N-terminal region	Atoms of C-terminal region	Hydrogen bond distance (Å)
N Ser29	OD1 Asp668	3.24
NE2 His30	O Asp668	2.88
OH Tyr32	N Ile671	2.96
OD1 Asp51	NZ Lys656	3.12
OE1 Glu57	NH2 Arg593	2.97
OE2 Glu57	NZ Lys656	2.78
OE1 Gln71	NE2 Gln662	2.96
NE2 Gln71	OE1 Gln662	2.94
OD1 Asn72	N Val675	2.76
ND2 Asn72	O Val675	3.04
OG Ser75	O Leu673	2.61
NE1 Trp85	OD1 Asn696	2.92
OD1 Asn89	OG1 Thr695	3.17
NE Arg91	OD2 Asp502	2.77
NE1 Trp96	OE1 Gln690	3.22
OH Tyr98	OG1 Thr495	2.67

[a] Distances were calculated using Contact of the CCP4 program suite. The distances of the hydrogen bonds were in the range of 2.38~3.24 Å.

Table S3 Salt bridges between the N-terminal region (residues 29-98) and C-terminal region (residues 436-709) in the α/β-hydrolase domain of chain A

Atoms of the N-terminal region	Atoms of the C-terminal region	Salt bridges distance (Å)
OD2 Asp51	NZ Lys656	3.01
OE2 Glu57	NE Arg593	2.78
NH2 Arg91	OD1 Asp502	2.81

Table S4 C-C contacts between the N-terminal region (residues 29-98) and C-terminal region (residues 436-709) in the α/β-hydrolase domain of chain A[a]

Atoms of the N-terminal region	Atoms of the C-terminal region	C-C contacts (Å)	Atoms of the N-terminal region	Atoms of the C-terminal region	C-C contacts (Å)
CB Ser29	CG Asp668	4.41	CG Phe45	CA Lys614	4.11
CG His30	CB Pro670	4.31	CG Phe45	CB Lys614	4.00
CE1 His30	CA Pro670	4.32	CD1 Phe45	CA Lys614	4.27
CE1 His30	CG Pro670	4.01	CE1 Phe45	CA Lys614	4.14
CE1 His30	CD Pro670	4.37	CE1 Phe45	CB Lys614	3.99
CD2 His30	CA Pro670	4.17	CZ Phe45	CA Asn613	4.1
CD2 His30	CB Pro670	4.24	CZ Phe45	CB Lys614	4.24
CD2 His30	CA Asp668	4.44	CD2 Phe45	CB Lys614	4.22
CD2 His30	CG Asp668	4.13	CD2 Phe45	CG Lys614	4.14
CD2 His30	C Asp668	4.04	CD2 Phe45	CD2 Phe617	4.30
CB Tyr32	CB Gln662	4.20	CD2 Phe45	CB Phe617	4.09
CB Tyr32	CG Gln662	4.19	CD2 Phe45	CG Phe617	4.26
CB Tyr32	CD Gln662	4.06	CE2 Phe45	CA Asn613	4.36
CD1 Tyr32	CB Gln662	4.26	CE2 Phe45	CB Lys614	4.33
CE1 Tyr32	C Ile671	4.19	CE2 Phe45	C Ser611	4.28
CE1 Tyr32	CA Ile671	4.39	CE2 Phe45	CB Phe617	4.19
CE1 Tyr32	CD Gln662	4.20	CB Thr47	CZ Phe617	4.16
CE1 Tyr32	CB Ile671	4.33	CB Thr47	CE2 Phe617	4.21

CZ Tyr32	CG Gln662	4.17	CG2 Thr47	CE2 Tyr621	4.11
CZ Tyr32	CB Ile671	4.26	CB Val49	CD1 Leu592	4.33
CZ Tyr32	CG1 Ile671	4.43	CG1 Val50	CD2 His627	3.96
CE2 Tyr32	CG Gln662	4.33	CG2 Val49	CE1 Phe617	4.34
CE2 Tyr32	CA Gln662	4.42	CG2 Val49	CZ Phe617	4.2
CD2 Tyr32	CB Gln662	4.07	CG2 Val49	CD1 Leu592	4.26
CD2 Tyr32	CG Gln662	4.08	CG2 Val49	CE1 Tyr621	4.01
CD2 Tyr32	CA Gln662	4.29	CG2 Val49	CZ Tyr621	4.10
CD2 Tyr32	C Gln662	4.44	CG Asp51	CE Lys656	4.39
CB Thr35	CB Ala659	4.10	CB Tyr53	CG Lys656	4.01
CB Tyr44	CD2 Phe617	4.20	CZ Tyr53	C Ala659	4.15
CB Tyr44	CG Phe617	4.11	CE2 Tyr53	CA Ser660	3.98
CG Tyr44	CZ Phe617	4.03	CE2 Tyr53	CZ Arg663	4.11
CD1 Tyr44	CD1 Leu592	4.34	C Tyr53	CE Lys656	4.10
CZ Tyr44	CB Leu592	4.20	CA Arg54	CE Lys656	4.29
CZ Tyr44	CD1 Leu592	4.06	CD Arg54	CZ Arg593	4.24
CD2 Tyr44	CB Ser611	4.09	CB Leu56	CB Tyr655	4.20
CD1 Leu56	CA Lys656	4.31	CG Asn72	CA Val675	4.44
CD1 Leu56	CB Tyr655	4.05	CZ Phe74	CE1 Tyr672	4.37
CD1 Leu56	C Tyr655	3.95	CD2 Phe74	CE1 Tyr672	4.04
C Leu56	CG Trp652	4.43	CA Ser75	CG Tyr672	4.34
CA Glu57	CE2 Trp652	4.28	CA Ser75	CD1 Tyr672	4.04
CA Glu57	CD2 Trp652	4.04	CB Ser75	C Leu673	4.26

CB Glu57	CG Trp652	4.20	CG2 Thr78	CZ Tyr672	4.26
CB Glu57	CD1 Trp652	4.14	CD1 Leu79	CD2 Leu639	4.28
CB Glu57	CE2 Trp652	4.08	CD1 Leu79	CB Asn696	4.25
CB Glu57	CD2 Trp652	4.15	CD1 Leu79	CG Asn696	4.18
CG Glu57	CZ2 Trp652	4.26	CD2 Leu79	CG Phe700	4.09
CG Glu57	CZ3 Trp652	4.21	CD2 Leu79	CE2 Tyr672	4.21
CG Glu57	CH2 Trp652	4.35	CD2 Leu79	CD2 Tyr672	4.17
CG Glu57	CG Trp652	4.28	CD1 Leu82	CB Ala699	4.24
CD Glu57	CZ Arg593	4.12	CD1 Leu82	CB Asp703	4.36
CD Glu57	CZ2 Trp652	4.06	CD2 Leu82	CB Asp703	4.25
CD Glu57	CZ3 Trp652	4.22	CB His84	CB Val709	4.09
CD Glu57	CH2 Trp652	4.01	CB His84	CG2 Val709	4.34
CD Glu57	CE Lys656	4.13	CG His84	CG2 Val709	4.37
CD Glu57	CE2 Trp652	4.24	CD2 His84	CB Val709	4.00
C Glu57	CG Trp652	4.37	CD2 His84	CG2 Val709	4.31
CB Gln59	CE2 Tyr655	4.00	CA Trp85	CB Ala699	4.11
CB Gln59	CD1 Trp652	4.34	CB Trp85	CB Ala699	4.39
CG Gln59	CD Pro651	4.10	CG Trp85	CB Ala699	4.40
CG Gln59	CG Pro651	3.97	CE2 Trp85	CB Thr695	4.22
CD Gln59	CD Pro651	4.02	CE3 Trp85	CG2 Thr695	4.10
CD Gln59	CG Pro651	4.43	CZ3 Trp85	CG2 Thr695	4.36
CD Gln59	CD1 Trp652	4.20	CH2 Trp85	CG2 Thr695	4.13
CZ3 Trp67	CD2 Tyr655	4.32	CH2 Trp85	CD Glu692	4.33

CZ3 Trp67	CA Tyr655	4.06	CZ2 Trp85	CB Thr695	4.22
CZ3 Trp67	CD1 Tyr655	3.96	CZ2 Trp85	CD Glu692	4.07
CH2 Trp67	C Tyr655	4.12	C Trp85	CG2 Thr695	4.36
CZ2 Trp67	CB Ala659	4.36	CD Lys87	CB Val709	4.37
CG1 Val68	CG1 Val675	4.05	CD Lys87	C Val709	4.33
CB Gln71	CD2 Leu673	4.06	CG1 Ile88	CE Met503	4.41
CB Gln71	CB Leu673	4.22	CG1 Ile88	CG1 Val709	4.42
CG Gln71	CD2 Leu673	4.10	CG2 Ile88	CB Tyr698	4.41
CD Gln71	CD Gln662	4.13	CG2 Ile88	CA Thr695	4.11
CG2 Ile88	CG2 Thr695	4.30	CG Trp96	CG1 Val687	4.18
CD1 Ile88	CE Met503	4.09	CD1 Trp96	CG2 Thr495	3.97
CD1 Ile88	CG2 Val709	4.41	CD1 Trp96	CG2 Val687	4.40
CD1 Ile88	C Tyr698	4.15	CE2 Trp96	CB Thr495	4.21
CA Asn89	CB Thr695	4.30	CE2 Trp96	CB Gln690	4.25
CG Asn89	CG2 Thr695	4.13	CZ3 Trp96	CG2 Thr495	4.25
CB Leu92	CG1 Val691	4.11	CH2 Trp96	CG2 Thr495	4.17
CD1 Leu92	CA Phe694	4.37	CH2 Trp96	CB Arg496	4.08
CD1 Leu92	CB Phe694	4.03	CH2 Trp96	CA Arg496	4.00
CD1 Leu92	CD2 Tyr698	4.29	CH2 Trp96	CB Phe694	4.36
CD1 Leu92	CB Thr695	4.42	CH2 Trp96	CG Phe694	4.44
CD2 Leu92	CA Gly499	4.07	CH2 Trp96	CD2 Phe694	4.13
C Leu92	CG2 Val691	4.40	CZ2 Trp96	CB Thr495	4.27
C Leu92	CG1 Val691	4.13	CZ2 Trp96	CB Arg496	4.31

CA Thr93	CG2 Val691	4.31	CZ2 Trp96	CA Arg496	4.43
CA Thr93	CG1 Val691	4.05	CZ2 Trp96	CB Gln690	4.23
CB Thr93	CG1 Val691	4.43	CB Tyr98	CD1 Tyr437	4.41
C Glu94	CG Pro438	4.39	CB Tyr98	CD Pro438	4.12
CB Leu95	CZ Tyr437	4.16	CG Tyr98	CE1 Tyr437	4.22
CD1 Leu95	CG Asp502	4.01	CE1 Tyr98	CG2 Thr495	4.09
CD2 Leu95	CZ Tyr437	4.27	CZ Tyr98	CE2 Phe493	4.04
CD2 Leu95	CE2 Tyr437	4.29	CZ Tyr98	CG2 Thr495	4.11
CD2 Leu95	CB Ala439	4.16	CE2 Tyr98	CE2 Phe493	4.41
C Leu95	CE1 Tyr437	4.31	CE2 Tyr98	CE1 Tyr437	4.37
CB Trp96	CG1 Val687	4.01			

[a] Distances were calculated using Contact of the CCP4 program suite. The distances of C-C contacts were in the range of 3.95~4.44 Å.

Table S5 Hydrogen bonds between β-propeller domain and α/β-hydrolase domain of chain A in the native structure[a]

Atoms of β-propeller domain	Atoms of α/β-hydrolase domain	Hydrogen bond distance (Å)
NH1 Arg100	O Trp96	3.04
OD1 Asp121	N Trp688	3.16
NE2 Gln122	O Gly684	2.64
OG1 Thr149	NH2 Arg678	3.19
OG1 Thr149	NE Arg678	3.20
OD1 Asp170	NH1 Arg678	3.05
O Asp170	NE Arg645	3.14
O Asp170	NH1 Arg645	2.72
OG1 Thr173	OD2 Asp647	2.61
N Asp174	OD1 Asp647	2.82
OD2 Asp174	NE Arg648	2.69
O Asp225	N Ala599	3.00
O Thr254	NE2 Gln603	2.86
OD1 Asp274	NE2 Gln603	2.82
N Tyr276	OG Ser606	3.04
O Thr323	N Gly520	2.93
OE1 Gln352	OG Ser516	2.92
OD1 Asp368	OH Tyr518	2.58
O Asp368	NE2 Gln448	2.79
NZ Lys370	OE1 Gln448	3.16

OG Ser411	N Phe493	2.84

[a] Distances were calculated using Contact of the CCP4 program suite. The distances of the hydrogen bonds were in the range of 2.38~3.24 Å.

Table S6 Salt bridges between the β-propeller domain and α/β-hydrolase domain of chain A in the native structure

Atoms of β-propeller domain	Atoms of α/β-hydrolase domain	Salt bridges distance (Å)
OD1 Asp170	NH2 Arg678	2.51
OD1 Asp174	NH2 Arg648	3.26

Table S7 C-C contacts between the β-propeller domain and the α/β-hydrolase domain of chain A in the native structure[a]

Atoms of propeller domain	Atoms of hydrolase domain	C-C contacts (Å)	Atoms of propeller domain	Atoms of hydrolase domain	C-C contacts (Å)
C Asp117	C Tyr98	4.35	C Thr173	CG Asp647	4.43
CA Gly118	CA Gln97	4.36	CB Asp174	CG Asp647	4.17
CA Gly118	C Gln97	4.19	CG Asp174	CZ Arg648	4.04
CA Gly118	CA Tyr98	4.28	CG Asp174	CD Arg648	4.40
C Gly118	CA Gln97	4.27	CG Asp174	C Asp647	4.34
CA Asp121	CG2 Val687	4.26	CG Asp174	CA Arg648	4.20
CA Asp121	CB Val687	4.10	CG Asp174	CB Arg648	4.06
CB Asp121	CA Pro686	4.36	CE1 Phe198	CZ Arg648	4.11
CB Asp121	CD1 Trp688	4.39	CZ Phe198	CB Asn600	4.41
CG Asp121	CB Pro686	4.27	CZ Phe198	CD Arg648	4.40
CG Asp121	CG Trp688	4.36	CE2 Phe198	CB Asn600	4.01
C Asp121	CG2 Val687	4.27	CG Asp225	CB Ser598	4.23
CB Asp147	CZ2 Trp688	4.35	C Asp225	CB Ala599	4.25
CA Gly148	CB Pro686	4.36	CA Ser226	CB Ala599	4.00
C Gly148	CB Pro686	4.03	CZ Phe273	CB Gln603	4.26
C Gly148	CG Pro686	4.01	CE2 Phe273	CG Gln603	4.06
C Gly148	CD Pro686	4.40	CA Asp274	CG Gln603	4.41
CB Thr149	CB Arg678	4.31	CA Asp274	CD Gln603	4.32

CB Thr149	CZ Arg678	4.40	C Asp274	CG Gln603	4.23
CG2 Thr149	CD Pro686	4.34	CA Gly275	CB Ser606	4.34
CG2 Thr149	CB Arg678	4.17	CB Tyr276	CE1 His523	4.05
CB Ala151	C Gly684	4.19	CD1 Tyr276	CE2 Phe485	3.98
CG Asp170	CZ Arg678	4.38	CE1 Tyr276	C Gln603	4.38
C Asp170	CZ Arg645	4.42	CE1 Tyr276	CA Trp604	4.30
CA Gly171	C Arg678	3.97	CE1 Tyr276	CD1 Trp604	3.99
CA Gly171	CA Arg678	4.37	CZ Tyr276	CD1 Trp604	4.25
CA Gly171	CZ Arg678	4.19	CZ Tyr276	CE2 Phe485	4.34
C Gly171	CA Ala679	4.09	CE2 Tyr276	CD2 Phe485	4.11
C Gly171	C Ala679	4.24	CD2 Tyr276	CD2 Phe485	4.06
C Gly171	CD Arg645	4.37	CA Asp277	CE1 His523	4.21
CA Gly172	C Ala679	4.22	CB Asp277	CG Lys524	4.36
CA Gly172	CA Gly680	4.12	CB Asp277	CE1 His523	4.13
C Gly172	CA Gly680	4.02	CG Asp277	CE Lys524	4.11
CA Thr173	CG Asp647	4.17	CB Thr323	CA Gly520	4.03
CB Thr173	CG Asp647	4.33	CG2 Thr323	CA Gly520	4.22
CG2 Thr323	C Gly520	4.33	CA Asn412	CA Phe493	4.42
CG2 Thr323	CB Ala521	4.05	C Asn412	CA Arg492	4.33
CG2 Thr323	CA Ala521	4.39	C Asn412	CB Arg492	4.15
C Thr323	CA Gly519	4.38	CB Phe413	CG2 Val446	4.16
CB Ala350	CA Gly519	4.07	CB Phe413	CG2 Val464	4.03
CB Ala350	C Tyr518	4.35	CB Phe413	CG1 Val464	4.32

CB Ala350	C Glu517	4.32	CD1 Phe413	CE1 Tyr494	4.14
CD2 Leu366	CD2 Leu490	4.13	CD1 Phe413	CE2 Phe507	4.41
CA Asp368	CE1 Tyr518	4.31	CE1 Phe413	CA Pro491	4.02
CA Asp368	CZ Tyr518	4.21	CE1 Phe413	C Pro491	4.42
CB Asp368	CD Gln448	4.04	CE1 Phe413	CE Met509	3.98
CB Asp368	CZ Phe450	4.37	CE1 Phe413	CE2 Phe507	4.02
CG Asp368	CE2 Phe450	4.13	CZ Phe413	C Pro491	4.26
CG Asp368	CZ Phe450	3.98	CZ Phe413	CB Phe462	4.06
CG Asp368	CE1 Tyr518	4.41	CZ Phe413	CG Phe462	4.24
CG Asp368	CZ Tyr518	4.43	CZ Phe413	CD2 Phe462	4.12
CB Ala369	CZ Phe462	3.96	CZ Phe413	CE Met509	4.39
CB Ala369	CE2 Phe462	3.96	CZ Phe413	CB Met509	4.22
CB Ala369	CD Pro491	4.43	CD2 Phe413	CA Pro491	4.21
CB Ala369	CG Glu517	4.14	CD2 Phe413	CB Pro491	4.05
CB Ala369	CG Asn511	4.02	CD2 Phe413	CG1 Val446	4.16
C Ala369	CG Pro491	4.39	CD2 Phe413	CG2 Val446	4.06
C Ala369	CD Pro491	4.04	CE2 Phe413	C Pro491	4.31
CG Lys370	CE1 Phe462	4.24	CE2 Phe413	CD1 Phe462	4.35
CG Lys370	CZ Phe462	4.16	CE2 Phe413	CG2 Val464	4.11
CD Lys370	CE1 Phe462	4.33	CE2 Phe413	CD2 Phe462	3.99
CE Met391	CB Val446	4.24	C Phe413	CD1 Phe493	4.01
CA Gly392	CG Pro491	4.39	CA Leu414	CB Phe441	4.42
C Gly392	CG Pro491	4.32	CA Leu414	CD2 Phe441	4.12

CB Thr393	C Pro491	4.31	CB Leu414	CB Phe441	4.14
CG2 Thr393	CB Arg492	4.03	CG Leu414	CB Tyr444	4.35
CG2 Thr393	CB Leu490	4.12	CG Leu414	CG1 Val464	3.99
CG2 Thr393	CD1 Leu490	4.01	CD1 Leu414	CA Val445	3.96
CB Ser411	CA Phe493	4.29	CD1 Leu414	CA Tyr444	4.22
C Ser411	CA Arg492	4.29	CD1 Leu414	CG1 Val464	4.34
C Ser411	C Arg492	4.44	CD2 Leu414	CD1 Leu501	4.17
CA Asn412	C Pro491	4.16	CD2 Leu414	CD1 Phe441	4.12
CA Asn412	CA Arg492	4.00	CD2 Leu414	CD2 Phe441	4.34
CD2 Leu414	CG Tyr444	4.39	CA Pro416	CD2 Phe493	4.06
C Leu414	CD2 Phe441	4.16	CB Pro416	CB Tyr437	4.00
C Thr415	CB Phe493	4.29	CG Pro416	CE1 Phe493	4.13
C Thr415	CG Phe493	4.34	CG Pro416	CZ Phe493	4.30
C Thr415	CD1 Phe493	3.99	CG Pro416	CG Tyr437	4.00
CA Pro416	CB Phe493	4.08	CD Pro417	CB Phe493	4.17
CA Pro416	CE1 Phe493	4.29	CD Pro417	CG Phe493	4.23

[a] Distances were calculated using Contact of the CCP4 program suite. The distances of C-C contacts were in the range of 3.95~4.44 Å.

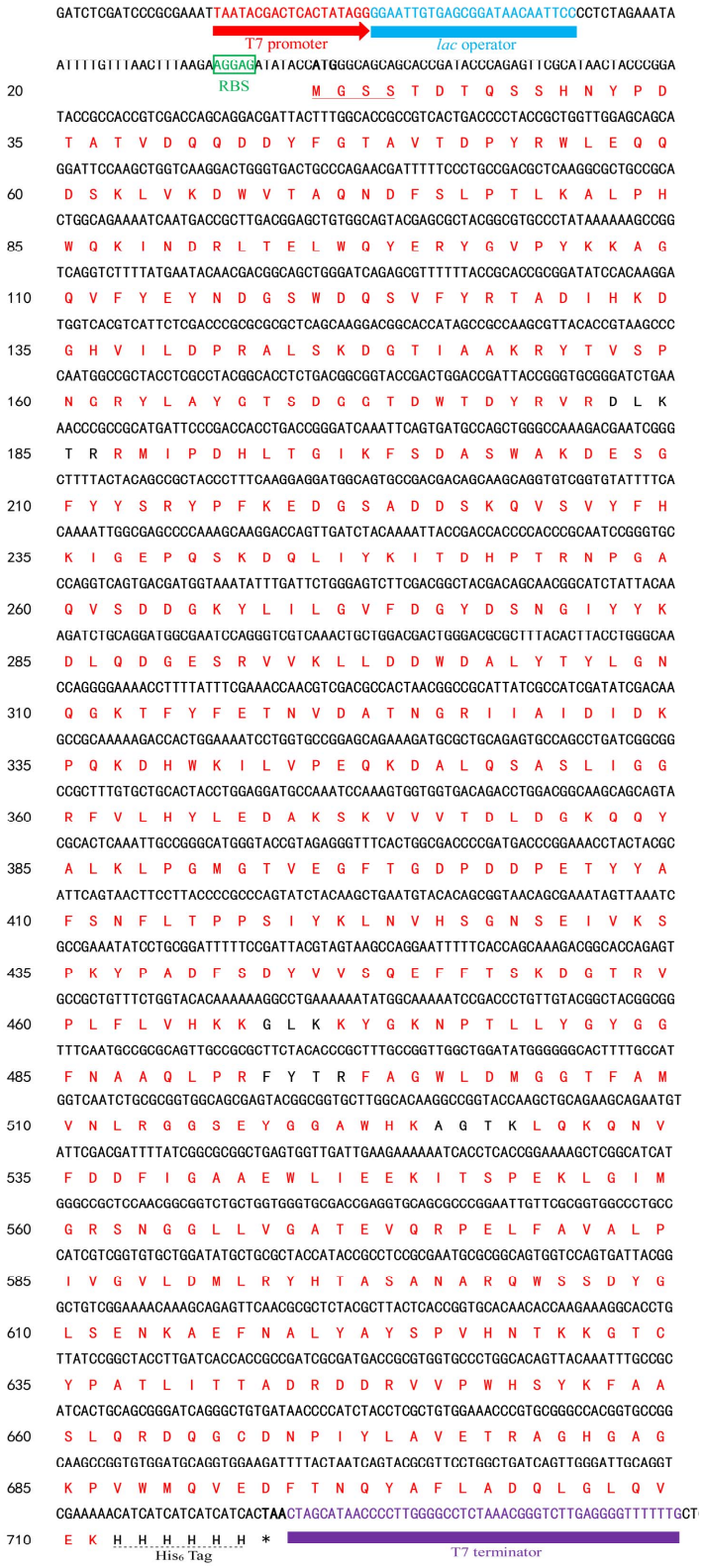
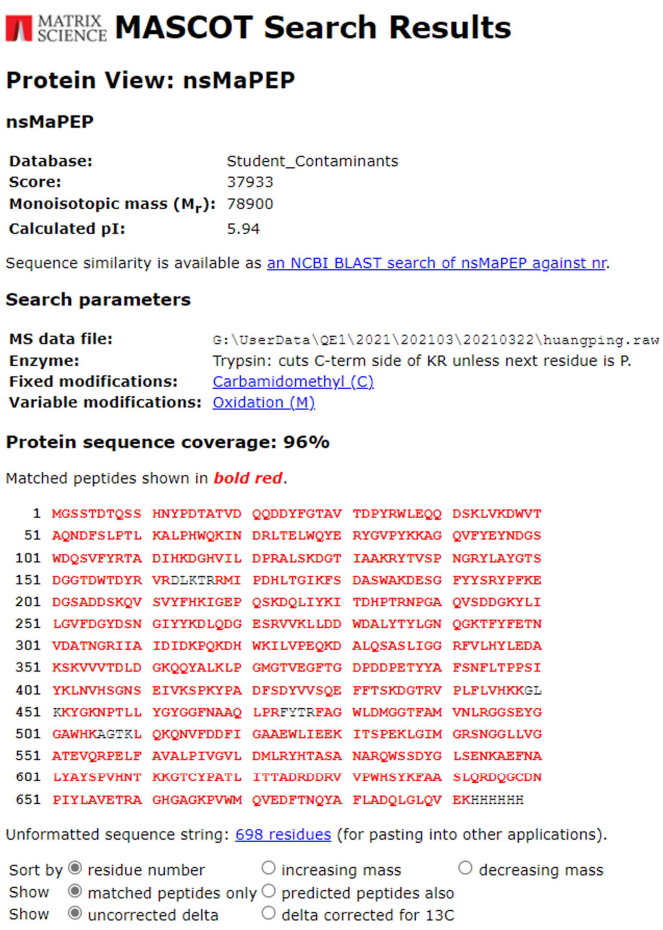


Figure S1 The expression cassette of *Ma*POP and the deduced amino acid sequence of the recombinant enzyme. To be consistent with the original sequence, the numbering of amino acids was started at the amino terminus, Met, which is coded by the initiation codon and

numbered 20. The recombinant protein was fused with a four-amino acid (MGSS), which was specified by underlines, at N-terminus, and a His_6 tag, which was specified by a dotted line, at C-terminus, respectively. Endoprotease trypsin digests of the enzyme, determined by nanoLC-MS/MS, are shown in red. The catalytic triad (Ser^{562}, Asp^{646} and His^{681}) are indicated with green triangles. Red arrow, the T7 promoter region; lightblue rectangle, the lac operator; green box, the ribosome-binding site; the violet rectangle, transcription terminator region.



MATRIX SCIENCE **MASCOT Search Results**

Protein View: nsMaPEP

nsMaPEP

Database:	Student_Contaminants
Score:	37933
Monoisotopic mass (M_r):	78900
Calculated pI:	5.94

Sequence similarity is available as an NCBI BLAST search of nsMaPEP against nr.

Search parameters

MS data file:	G:\UserData\QE1\2021\202103\20210322\huangping.raw
Enzyme:	Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Oxidation (M)

Protein sequence coverage: 96%

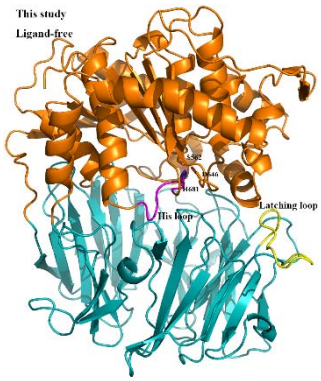
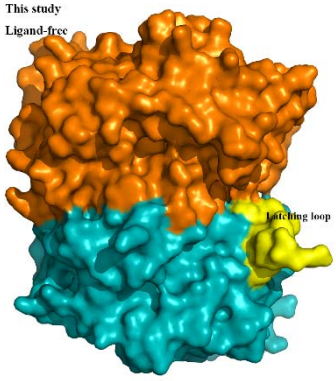
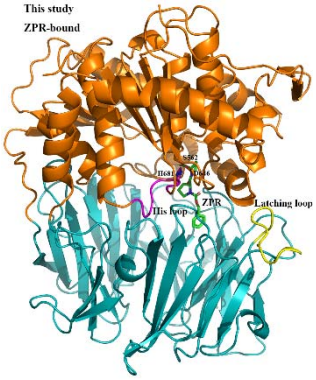
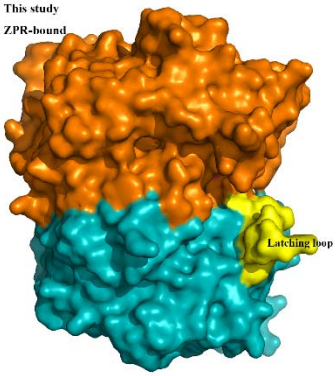
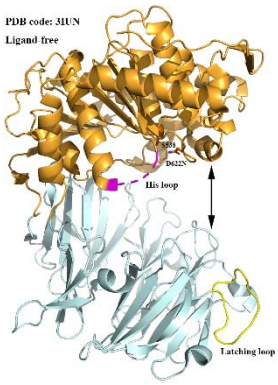
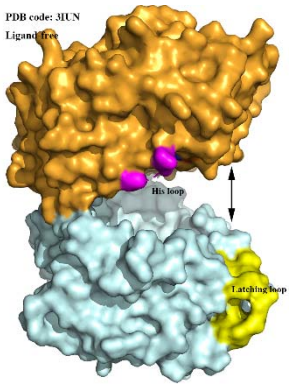
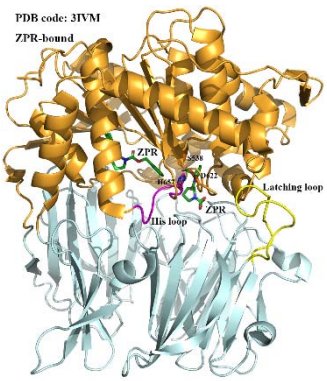
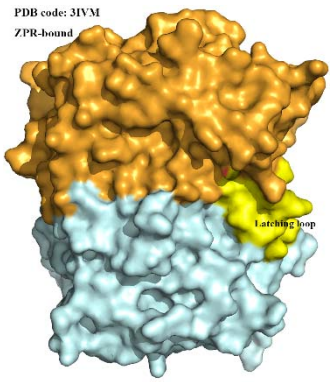
Matched peptides shown in ***bold red***.

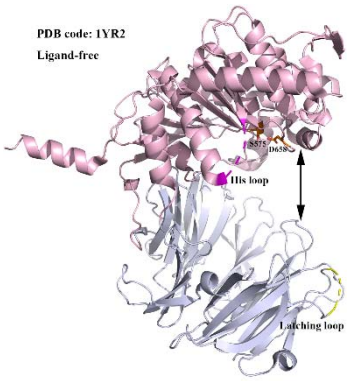
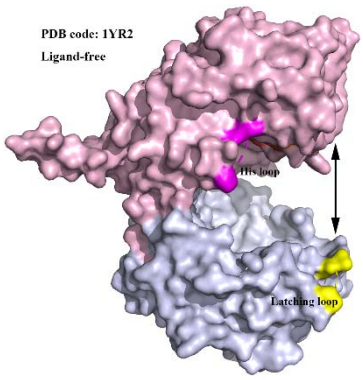
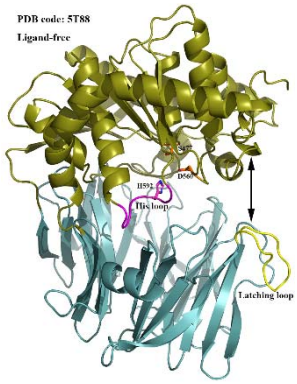
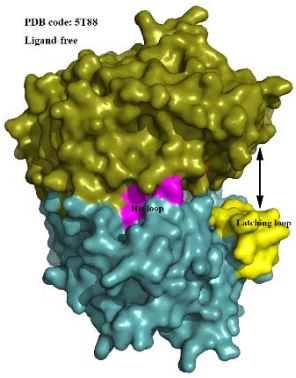
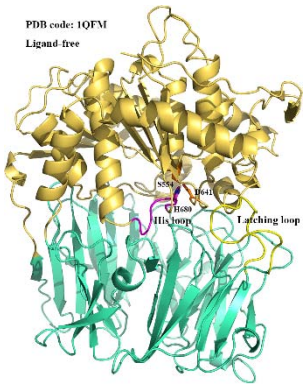
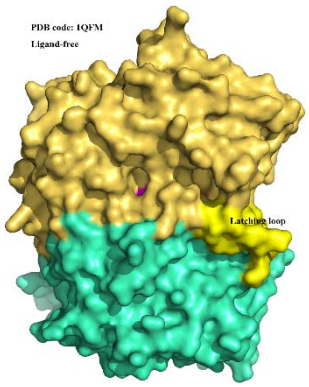
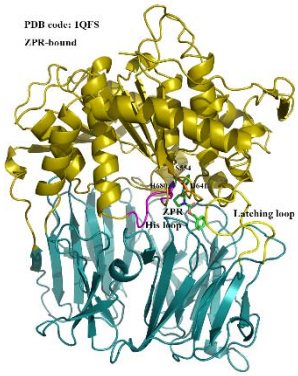
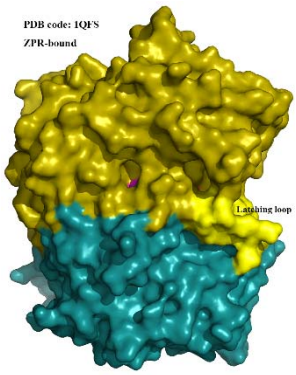
```
  1 MGSSTDTQSS HNYPDTATVD QQDDYFGTAV TDPYRWLEQQ DSKLVKDWVT
 51 AQNDFSLPTL KALPHWQKIN DRLTELWQYE RYGVPYKKAG QVFYEYNDGS
101 WDQSVFYRTA DIHKDGHVIL DPRALSKDGT IAAKRYTVSP NGRYLAYGTS
151 DGGTDWTDYR VRDLKTRRMI PDHLTGIKFS DASWAKDESG FYYSRYPFKE
201 DGSADDSKQV SVYFHKIGEP QSKDQLIYKI TDHPTRNPGA QVSDDGKYLI
251 LGVFDGYDSN GIYYKDLQDG ESRVVKLLDD WDALYTYLGN QGKTFYFETN
301 VDATNGRIIA IDIDKPQKDH WKILVPEQKD ALQSASLIGG RFVLHYLEDA
351 KSKVVVTDLD GKQQYALKLP GMGTVEGFTG DPDDPETYYA FSNFLTPPSI
401 YKLNVHSGNS EIVKSPKYPA DFSDYVVSQE FFTSKDGTRV PLFLVHKKGL
451 KKYGKNPTLL YGYGGFNAAQ LPRFYTRFAG WLDMGGTFAM VNLRGGSEYG
501 GAWHKAGTKL QKQNVFDDFI GAAEWLIEEK ITSPEKLGIM GRSNGGLLVG
551 ATEVQRPELF AVALPIVGVL DMLRYHTASA NARQWSSDYG LSENKAEFNA
601 LYAYSPVHNT KKGTCYPATL ITTADRDDRV VPWHSYKFAA SLQRDQGCDN
651 PIYLAVETRA GHGAGKPVWM QVEDFTNQYA FLADQLGLQV EKHHHHHH
```

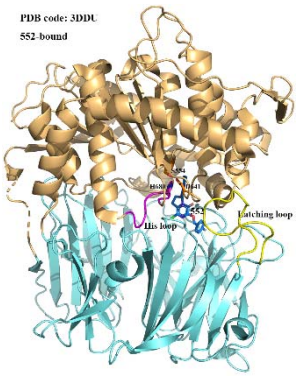
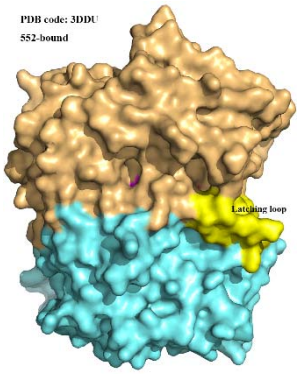
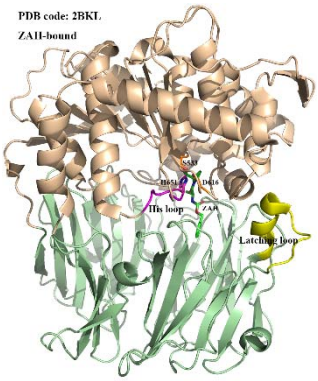
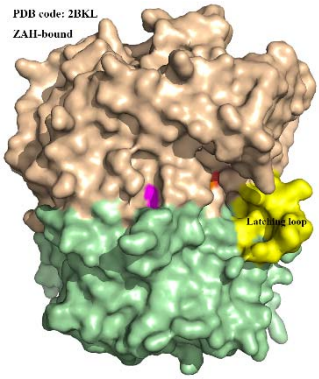
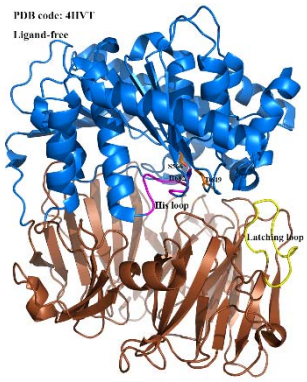
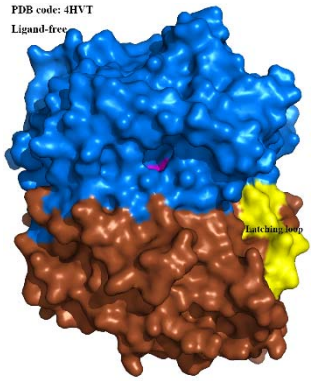
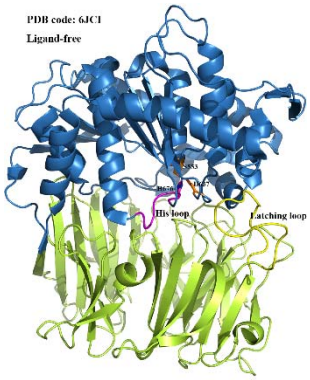
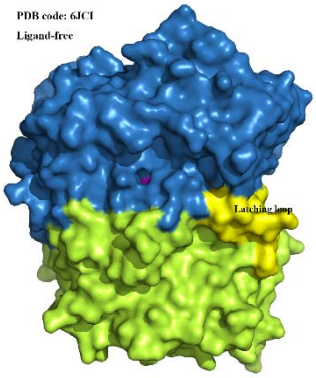
Unformatted sequence string: 698 residues (for pasting into other applications).

Sort by ◉ residue number ○ increasing mass ○ decreasing mass
Show ◉ matched peptides only ○ predicted peptides also
Show ◉ uncorrected delta ○ delta corrected for 13C

Figure S2 Analysis of protein sequence by nanoLC-MS/MS. A Mascot search yielded 96 % sequence coverage, as highlighted in red. The identified peptides originated from the whole region of the protein sequence deduced from the nucleotide sequence. Only five short peptides, including ^{182}DLKTR^{186}, ^{468}GLK^{470}, ^{493}FYTR^{496}, ^{525}AGTK^{528}, and ^{712}HHHHHH^{717} (the C-terminal His6 tag) were not identified.

Sources	Cartoon representation	Surface representation
Source: PDB code: 7VGB Ligand-free Interdomain: close *Ma*POP Reference: This study		
Source: PDB code: 7VGC Ligand-bound Interdomain: close *Ma*POP-ZPR Reference: This study		
Source: *Aeromonus punctata* PDB code: 3IUN Ligand-free Interdomain: open		
Source: *Aeromonus punctata* PDB code: 3IVM Ligand-bound Interdomain: close		

Source: *Sphingomonas capsulata* PDB code: 1YR2 Ligand-free Interdomain: open		
Source: *Pyrococcus furiosus* PDB code: 5T88 Ligand-free Interdomain: open		
Source: *Sus scrofa* PDB code: 1QFM Ligand-free Interdomain: close		
Source: *Sus scrofa* PDB code: 1QFS Ligand-bound Interdomain: close		

Source: *Homo sapiens* PDB code: 3DDU Ligand-bound Interdomain: close		
Source: *Myxococcus xanthus* PDB code: 2BKL Ligand-bound Interdomain: close		
Source: *Rickettsia typhi* PDB code: 4HVT Ligand-free Interdomain: close		
Source: *Haliotis discus hannai* PDB code: 6JCI Ligand-free Interdomain: close		

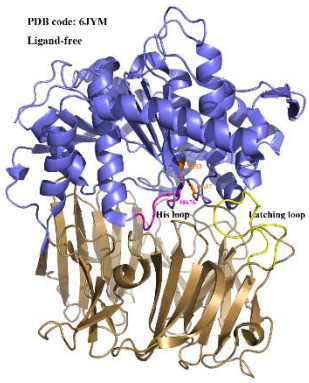
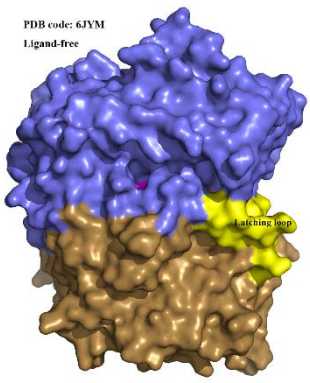
Source: *Haliotis discus hannai* PDB code: 6JYM Ligand-free Interdomain: close		

Figure S3 The structures of *Ma*POP and other POPs determined previously. Each structure is shown in cartoon and surface representations. It is worth noting that the unbound forms of *A. punctata* POP (*Ap*POP, PDB id: 3IUN), *S. capsulata* (*Sc*POP, PDB id: 1YR2), and *P. furiosus* POP (*Pfu*POP, PDB id: 5T88) are in an open conformation in which the α/β-hydrolase and the β-propeller domains are widely separated. While, in the presence of inhibitor, the *Ap*POP (PDB id: 3IVM) are in a closed state that the two domains are packed tightly together. It is observed that the "latching loop" (shown as a yellow cartoon) and "His loop" (shown as a magenta cartoon) were modeled only in the closed state and omitted from the open state. All these closed structures expect for *Ma*POP (this study), show the "latch loop" extends to the α/β-hydrolase domain. The "latching loop" of *Ma*POP extends to the outside and makes minimal contact with the α/β-hydrolase domain, entirely solvent-accessible.

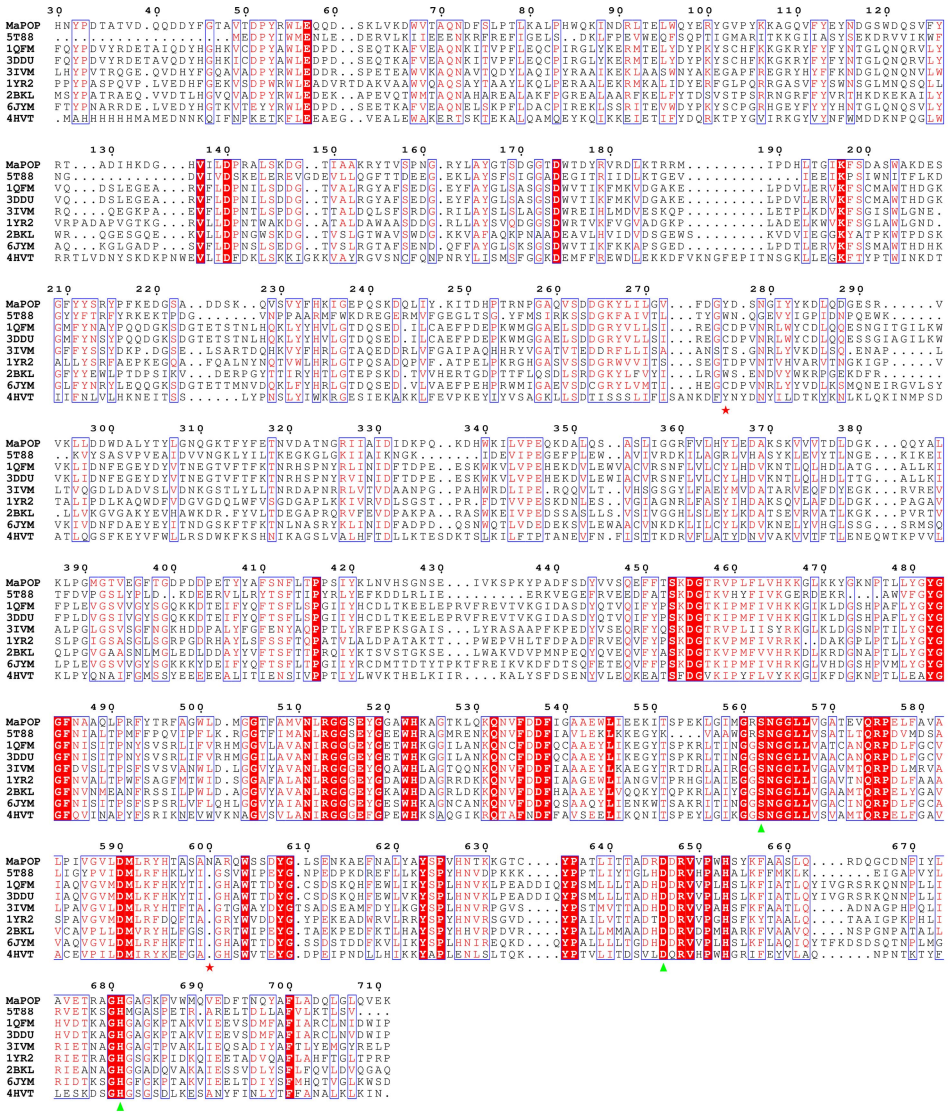


Figure S4 Sequence alignment of *Ma*POP and crystal structures reported POPs. *Microbulbifer arenaceous* POP (*Ma*POP, This study); *Pyrococcus furiosus* POP (PDB ID code 5T88); *Sus scrofa* POP (PDB ID code 1QFM); *Homo sapiens* POP (PDB ID code 3DDU); *Aeromonas punctate* POP (PDB ID code 3IVM); *Sphingomonas capsulata* POP (PDB ID code 1YR2); *Myxococcus xanthus* POP (PDB ID code 2BKL); *Haliotis discus hannai* POP (PDB ID code 6JYM); Rickettsia typhi POP (PDB ID code 4HVT). Amino acid residues with homology ≥ 75% are shaded in blue frames and the strictly conserved regions are highlighted with red background. The deduced catalytic triad (Ser^{562}, Asp^{646}, and His^{681}) of *Ma*POP and the corresponding residues of other serine-dependent peptidases are indicated with green triangles. The unique Tyr^{276} and Asn^{600} of MaPOP are indicated with a red star.

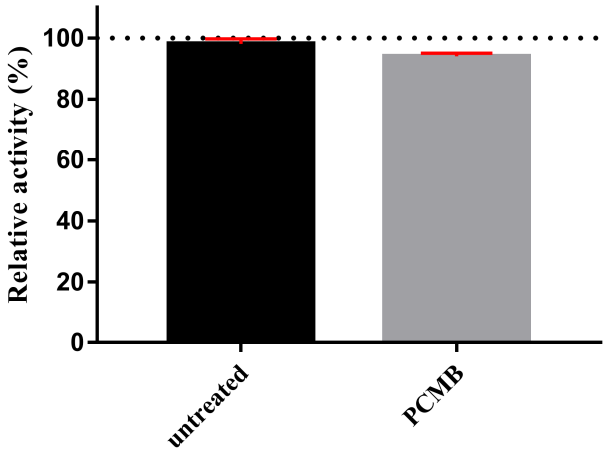


Figure S5 Effects of p-chloromercuribenzoate on enzyme activity. *Ma*POP was dissolved in 100 µl of 50 mM Tris-HCl buffer, pH 7.5, in the presence of 1 mM p-chloromercuribenzoate (PCMB) and incubated at 30 °C for 30 min. Residual peptidase activity was determined after incubation with Z-Gly-Pro-pNA following the standard method. Control experiments were carried out under identical conditions in the absence of the PCMB. All experiments were repeated three times.

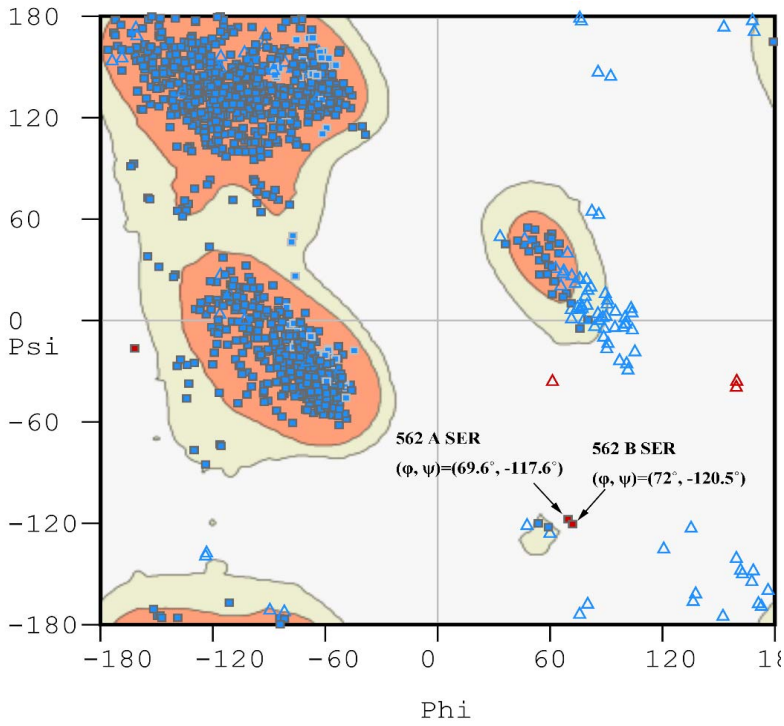


Figure S6 Ramachandran plot diagrams of residues in native *Ma*POP. The main chain conformation of Ser^{562} in both chain A and chain B of the native structure were strained, with (φ, ψ) = (69.6˚, -117.6˚) and (φ, ψ) = (72˚, -120.5˚), respectively.