Supplementary Material I for: Bailey *et al. manuscript 'An analysis of subdomain orientation, conformational change and disorder in relation to crystal packing of aspartic proteinases.'*

Abbreviations

Boc butoxycarbonyl, CBz benzyloxycarbonyl, DCCI dicyclohexylcarbodiimide, DMF N,N-dimethylformamide, DMSO dimethyl sulfoxide, HOBt hydroxybenzotriazole, TEA triethylamine, TFA trifluoroacetic acid, –R– reduced peptide bond.

Inhibitor synthesis

The inhibitors were made using manual solid-phase Boc synthetic techniques and had sequences based on those commonly occurring at chymosin and *Mucor Pusillus* pepsin cleavage sites in the insulin β-chain.

H_2N-His-Pro-His-Leu-Ser-Phe–R–Met-Ala-Tyr-COOH ... (DB1)
H_2N-His-Pro-His-Leu-Ser-Phe–R–Met-Ala-Ile-COOH ... (DB2)
H_2N-His-Pro-His-Leu-Ser-Ile–R–Met-Ala-Ile-COOH ... (DB3)
H_2N-His-Pro-His-Leu-Ser-Phe–R–Met-Ala-His-COOH ... (DB4/5)
H_2N-His-Ser-Leu-Phe-His-Phe–R–Phe-Thr-Pro-COOH ... (DB6)

Synthetic Strategy

The first three amino acids at the C-terminal end of each inhibitor were linked to the resin sequentially and a small amount of the resin was removed for analysis. During this time the compound required to form the reduced bond with the free amino group was synthesised. This compound was always a Boc-protected aldehyde of the next amino acid in the sequence which was coupled to the free amino end of the resin-linked tripeptide by reductive amination. The resin would then be reloaded and the synthesis continued until the remaining five other amino acids were attached. After this each peptide was cleaved from the resin using anhydrous hydrogen fluoride, analysed as before and then purified by HPLC chromatography.

The imidazole ring of histidine was protected with Boc, serine and threonine hydroxyls with benzyl ethers and the tyrosine hydroxyl with the Cbz group.

Synthesis of the Aldehyde

Two Boc-protected aldehydes were synthesised, Boc-Leu-CHO and Boc-Phe-CHO. A summary of the steps which are detailed below is shown in Figure 1.

Preparation of Boc-Phe-COOMe

An ethereal solution of diazomethane was prepared by adding 20g of N-methyl-N-nitrosourea to 60ml of a 40% KOH solution under 240ml of ether. The diazomethane was distilled using dry ice/ethanol as a cooling medium. The resulting solution was poured onto an ethereal solution of Boc-Phe-COOH. The un-recrystallised residue Boc-Phe-COOMe was used in the next stage of the synthesis.

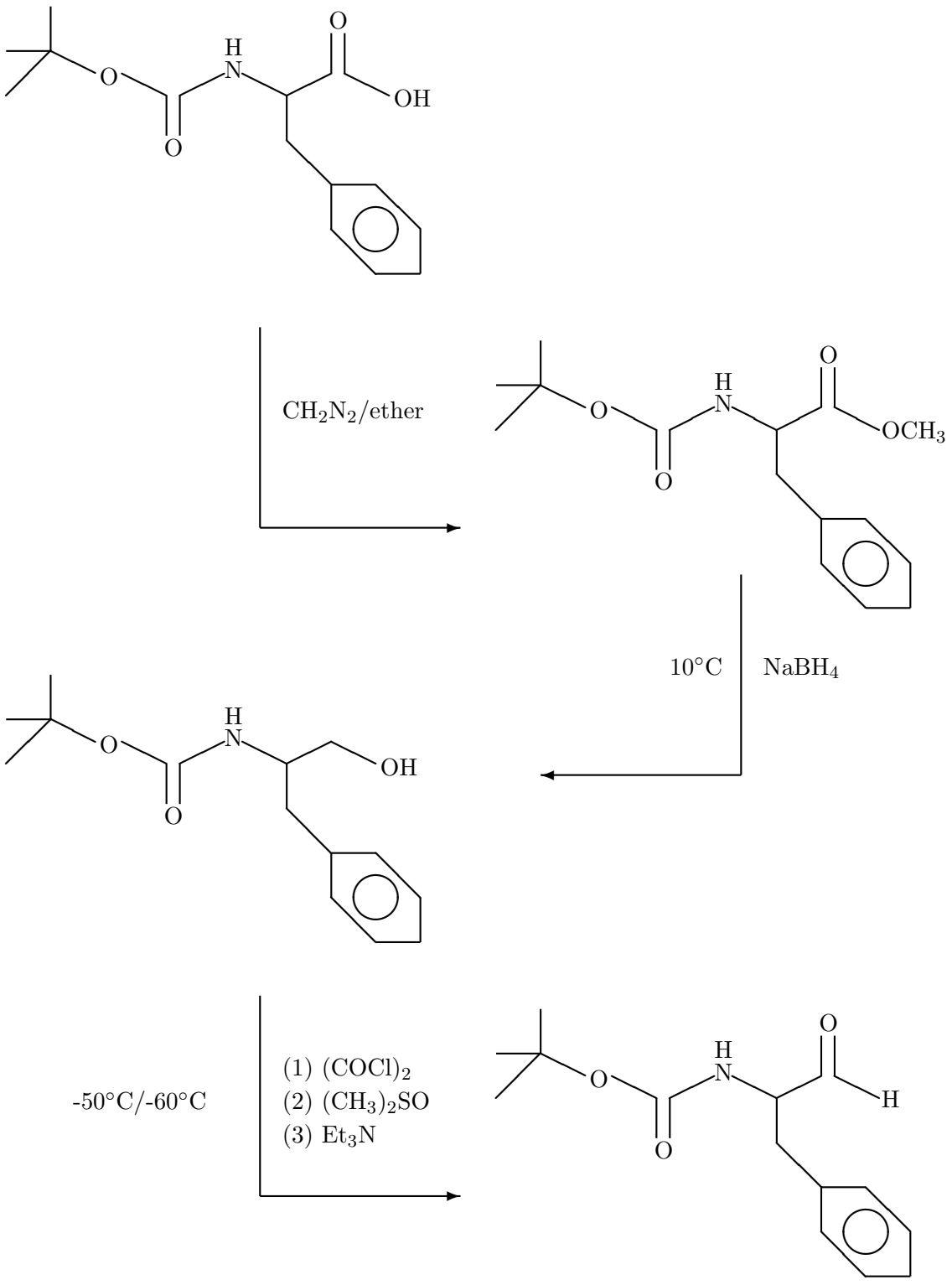
Preparation of Boc-Phe-CH_2-OH

The un-recrystallised methyl ester was dissolved in 350ml of dioxane and cooled to 10°C whereupon 20g of sodium borohydride ($NaBH_4$) was added and the temperature maintained at 10°C for 1 hour whilst at the same time adding a solution of 25ml acetic acid in 100ml dioxane. This mixture was then refluxed for $1\frac{1}{2}$ hours, after which 100ml of methanol was added. The solvent was removed by rotary evaporation and the residue dissolved in 250ml of saturated sodium bicarbonate. This solution was extracted three times with 150ml portions of ethyl acetate and dried over sodium sulphate. The residue remaining after rotary evaporation was redissolved in ether and crystallised from a mixture of butyl methyl ether and hexane to give the alcohol Boc-Phe-CH_2-OH.

Preparation of Boc-Phe-CHO

A four-necked flask was set up with a stirrer, spirit thermometer and two dropping funnels. 1ml of oxalyl chloride ($(COCl)_2$) was dissolved in 25ml of CH_2Cl_2 and cooled using a dry ice/ethanol mix to -50°/-60°. From one funnel, 1.7ml of dimethylsulphoxide ($(CH_3)_2SO$) in 5ml of CH_2Cl_2 was added and after two minutes a solution of 2.51g of Boc-Phe-CH_2-OH in 20ml of CH_2Cl_2 was added over five minutes from the second dropping funnel. After a further 15 minutes, 7ml of TEA was added and the resulting solution stirred for five minutes. The reaction mixture was then allowed to reach room temperature and the water layer extracted with 50ml of CH_2Cl_2. The organic layers were then combined and *quickly* extracted with saturated NaCl and dried over $MgSO_4$. After rotary evaporation a yellow oil was obtained which was immediately stored in the cold to prevent further degradation. This raw aldehyde undoubtedly contained some impurities.

Figure 1: Preparation of Boc-Phe-Aldehyde



Peptide Synthesis Protocol

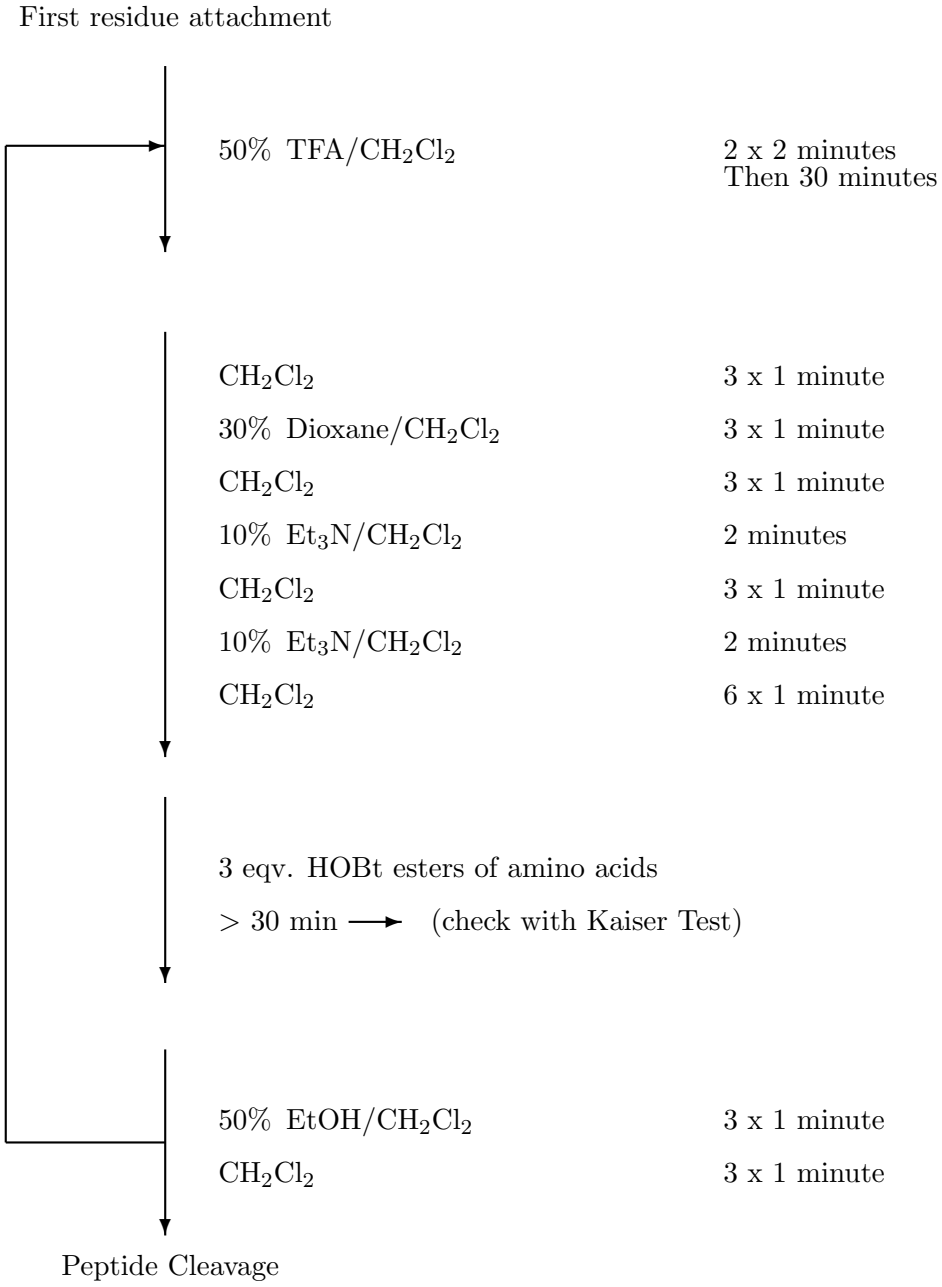
The synthesis of each peptide inhibitor consisted of the following stages.

1. Attachment of first amino acid residue to the resin
2. Attachment of the 3 amino acid residues up to the reduced bond
3. Formation of reduced bond
4. Attachment of remaining amino acid residues
5. Cleavage of peptide inhibitor
6. Purification and characterisation of inhibitor

The sequential attachment of the residues consisted of three stages, deprotection, neutralisation and coupling, as shown in Figure 2.

The attachment of the initial amino acid to the resin was performed using DMF and KI at 60°C. The resin was then loaded into the reaction vessel and the protecting Boc group removed by washing the resin twice with a 50% solution of trifluoroacetic acid (TFA) in CH_2Cl_2 and stirring for 30 minutes. The excess acid was removed with washes of CH_2Cl_2 and a solution of 30% dioxane in CH_2Cl_2. Neutralisation of the product was achieved using TEA as a 10% solution in CH_2Cl_2 to form the tert-ammonium salt, which was washed away prior to attachment of the next amino acid. A DCCI/HOBt mix was used as the coupling agent.

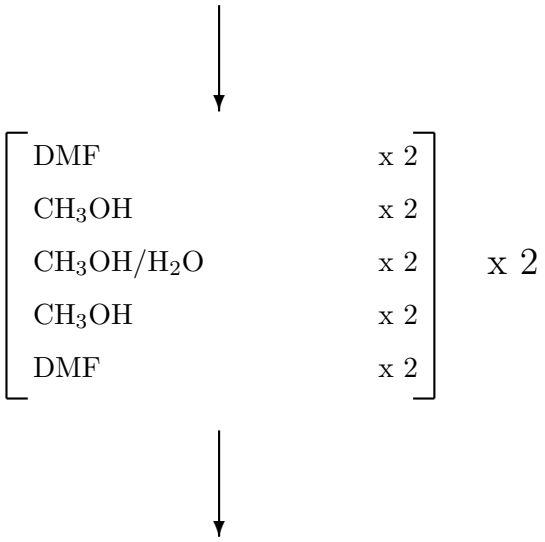
Figure 2: Peptide Synthesis Protocol



Formation of the reduced bond

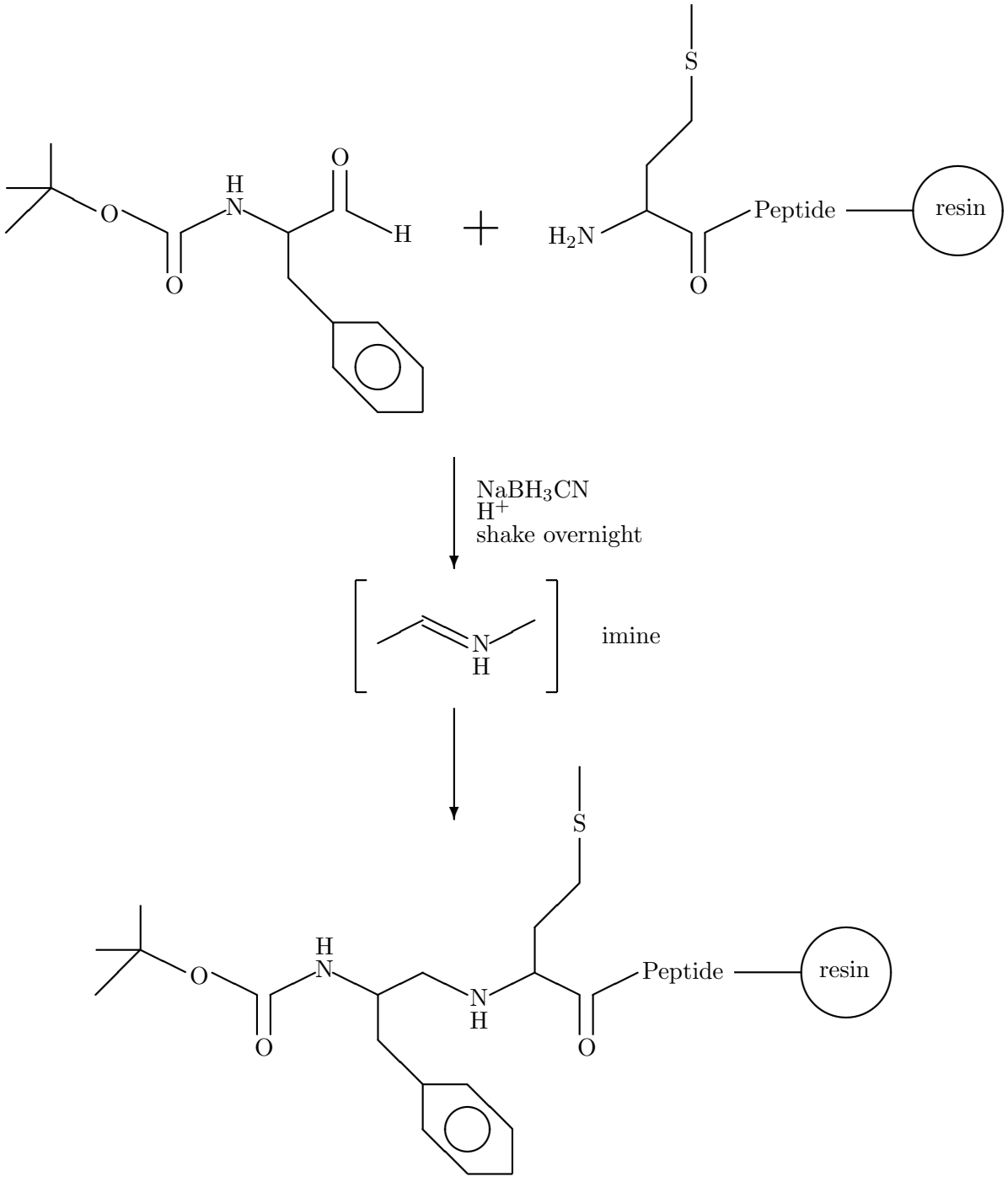
The reduced peptide bond analogue was formed by reductive amination of the Boc-Phe-Aldehyde by the resin linked peptide. The reaction proceeds via an imine intermediate as shown in Figure 3. The reducing agent used, sodium cyanoborohydride, reduces the imine group more rapidly than the carbonyl.

The raw aldehyde was dissolved in CH_2Cl_2 as was the peptide resin, and to this mixture a 3eqv. excess of sodium cyanoborohydride ($NaBH_3CN$) in acetic acid buffered-DMF was added, followed by shaking overnight. The resulting reaction mix was then treated according to the following protocol :

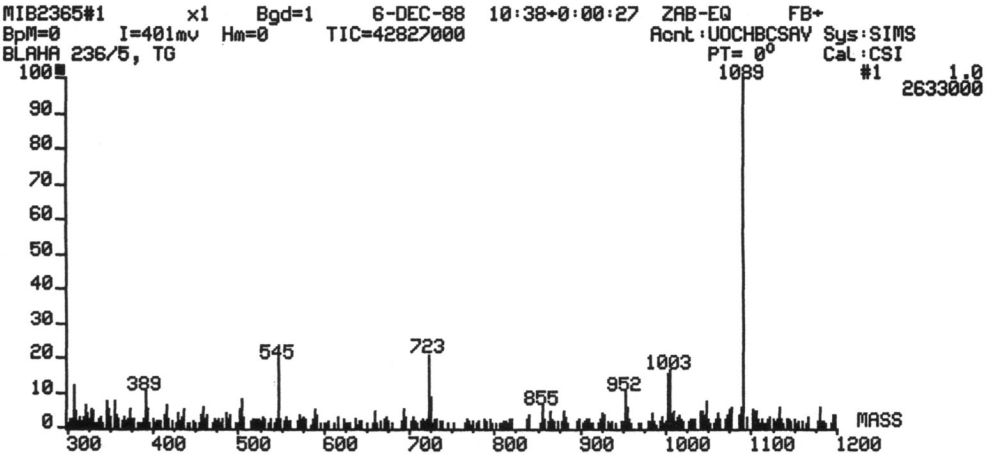


The reaction was repeated with more aldehyde, after which the final six residues were added in the conventional way. Anhydrous hydrogen fluoride was used to cleave the peptide from the resin in anisole for one hour on ice and excess HF was then flushed away with nitrogen. Each peptide was then recovered from the resin and any anisole degradation products by trituration with ether. Peptides were purified by reverse phase HPLC on a Sepharon C18 column and were characterised by liquid secondary ion mass spectrometry (LSIMS) using a ZAB-EQ instrument from V6 Analytical Ltd, Manchester (see following mass-spectra). Formation of the reduced bond is expected to lead to some racemisation at the P1 residue. However when the DB5 and DB6 complexes were solved, the *L* configuration was confirmed.

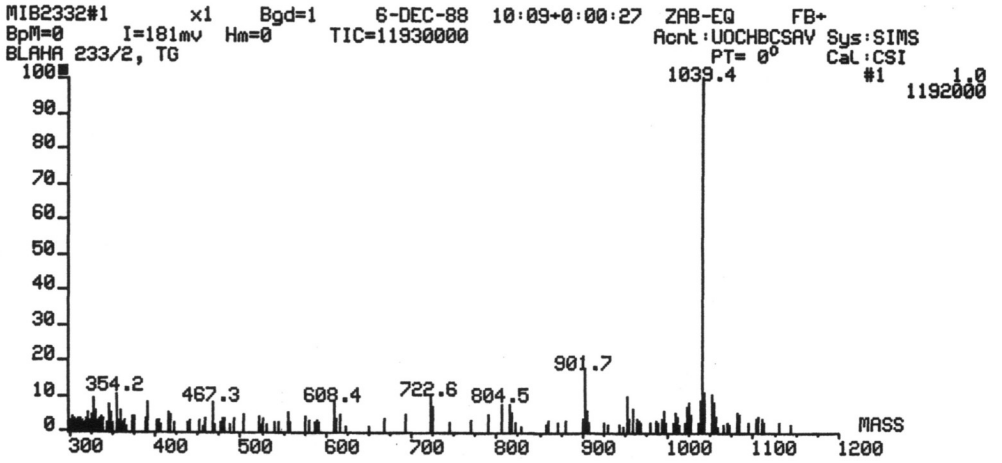
Figure 3: Reductive Amination



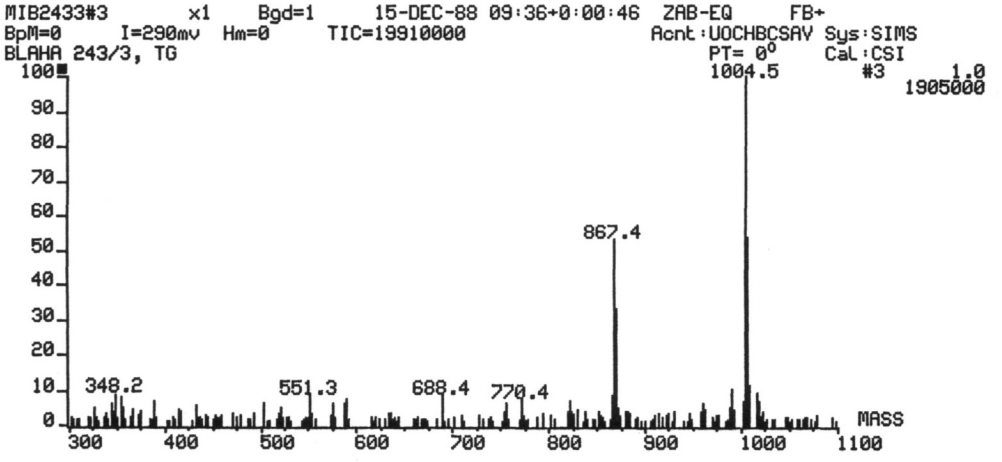
DB1


MIB2365#1 x1 Bgd=1 6-DEC-88 10:38+0:00:27 ZAB-EQ FB+
BpM=0 I=401mv Hm=0 TIC=42827000 Acnt:UOCHBCSAV Sys:SIMS
BLAHA 236/5, TG PT= 0° Cal:CSI
#1 1.0
1089
2633000
545
723
389
855
952
1003
MASS

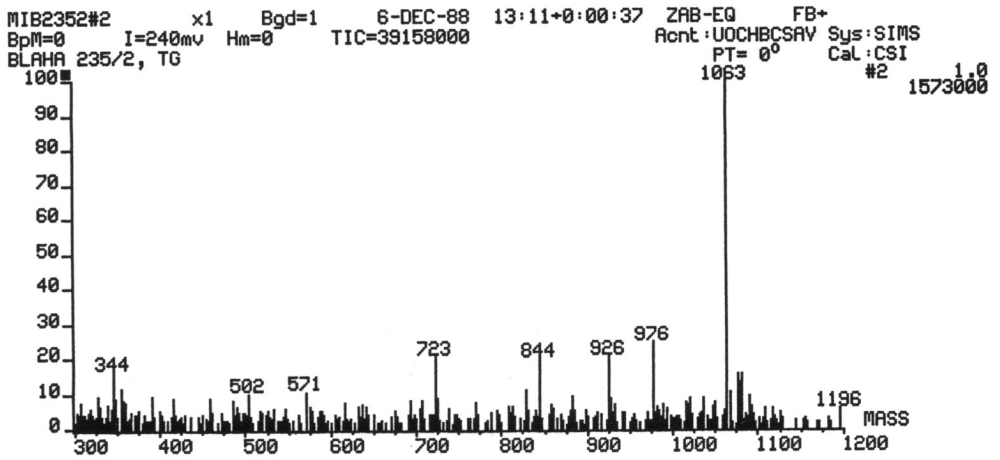
DB2


MIB2332#1 x1 Bgd=1 6-DEC-88 10:09+0:00:27 ZAB-EQ FB+
BpM=0 I=181mv Hm=0 TIC=11930000 Acnt:UOCHBCSAV Sys:SIMS
BLAHA 233/2, TG PT= 0° Cal:CSI
#1 1.0
1039.4
1192000
354.2
467.3
608.4
722.6
804.5
901.7
MASS

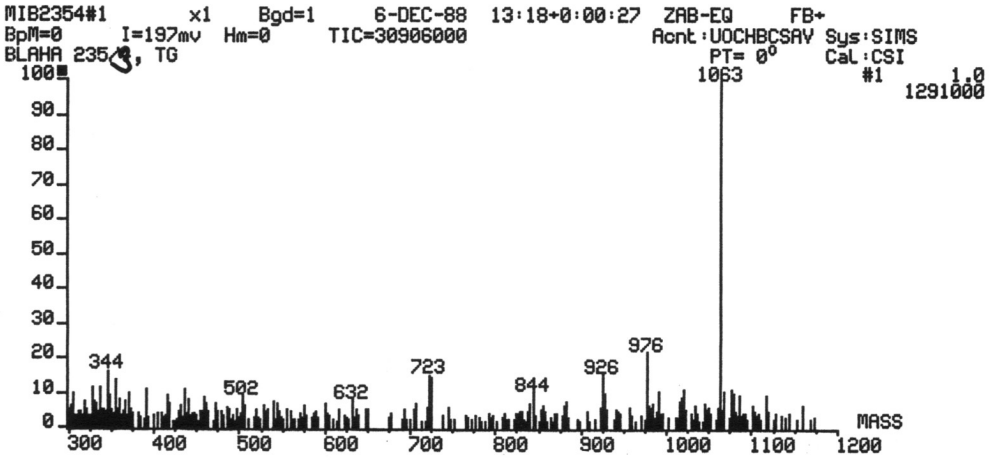
DB3


MIB2433#3 x1 Bgd=1 15-DEC-88 09:36+0:00:46 ZAB-EQ FB+
BpM=0 I=290mv Hm=0 TIC=19910000 Acnt:UOCHBCSAV Sys:SIMS
BLAHA 243/3, TG PT= 0° Cal:CSI
#3 1.0
1004.5
1905000
348.2
551.3
688.4
770.4
867.4
MASS

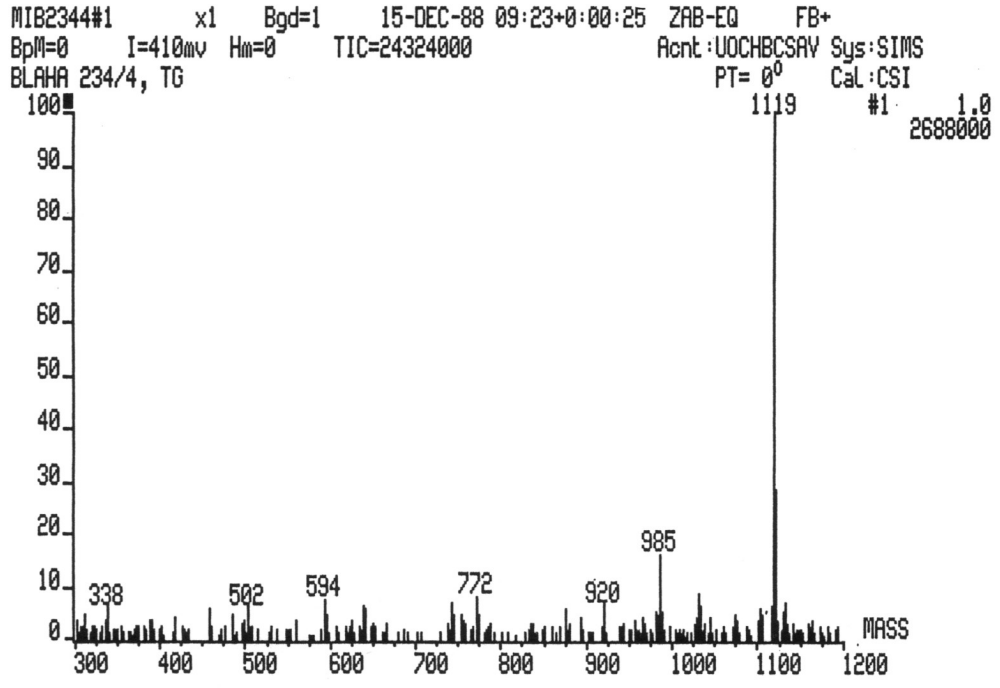
DB4


MIB2352#2 x1 Bgd=1 6-DEC-88 13:11+0:00:37 ZAB-EQ FB+
BpM=0 I=240mv Hm=0 TIC=39158000 Acnt:UOCHBCSAV Sys:SIMS
BLAHA 235/2, TG PT= 0° Cal:CSI
#2 1.0
1573000
1063
344
502
571
723
844
926
976
1196
MASS

DB5


MIB2354#1 x1 Bgd=1 6-DEC-88 13:18+0:00:27 ZAB-EQ FB+
BpM=0 I=197mv Hm=0 TIC=30906000 Acnt:UOCHBCSAV Sys:SIMS
BLAHA 235/3, TG PT= 0° Cal:CSI
#1 1.0
1291000
1063
344
502
632
723
844
926
976
MASS

DB6


MIB2344#1 x1 Bgd=1 15-DEC-88 09:23+0:00:25 ZAB-EQ FB+
BpM=0 I=410mv Hm=0 TIC=24324000 Acnt:UOCHBCSAV Sys:SIMS
BLAHA 234/4, TG PT= 0° Cal:CSI
#1 1.0
2688000
1119
338
502
594
772
920
985
MASS

Supplementary Material II for: Bailey *et al. manuscript 'An analysis of subdomain orientation, conformational change and disorder in relation to crystal packing of aspartic proteinases.'*

Table 5: The contacts made by the sulphate groups in type IV native endothiapepsin. Note that the residues are numbered according to the scheme of Blundell *et al.*, (1990) in which insertions relative to porcine pepsin are indicated by suffixes by A, B, C, etc.

Residue	Sulphate Atom	VDW contacts	Hydrogen Bonds
Intramolecular Contacts			
SO_4 1 (19vdw 4hb)			
Ser 108	O1	1	0
W205	O1	1	0
Ser 108	O2	2	0
W204	O2	1	1
W205	O2	1	1
Ser 108	O3	3	0
Ser 109	O3	2	0
Ser 108	O4	3	0
Ser 109	O4	4	2
W74	O4	1	0
SO_4 2 (15vdw 7hb)			
Ser 240	O1	4	1
Ser 241	O1	1	1
Ser 240	O2	1	1
Ser 242	O2	4	2
Lys 238A	O4	2	1
Ser 240	O4	3	1
SO_4 3 (19vdw 5hb)			
Tyr 175	O1	3	1
Gly 177	O1	1	0
Ser 178	O1	2	0
W238	O1	1	0
Ser 178	O2	3	0
Ile 179	O2	5	1
Tyr 175	O3	1	1
W211	O3	1	1
W238	O3	1	1
Ser 178	O4	1	0
Intermolecular Contacts			
SO_4 1 $\cdots$ $(-x, y+1/2, 1-z)$			
Ser 236	O2	1	1
Ser 236	O3	1	1
SO_4 1 $\cdots$ $(x, y, 1+z)$			
Thr 318	O1	1	0
Thr 318	O3	2	0
Thr 319	O1	1	1
Thr 319	O4	2	1
SO_4 2 $\cdots$ $(x, y-1/2, 1-z)$			
Ser 132	O1	1	1
Ser 132	O3	2	1
Pro 133	O3	1	0
SO_4 3 $\cdots$ $(1-x, y-1/2, -z)$			
Lys 64	O1	1	1

Table 6: The lattice contacts in type IV endothiapepsin crystals. Note that the residues are numbered according to the scheme of Blundell *et al.*, (1990) in which insertions relative to porcine pepsin are indicated by suffixes by A, B, C, etc.

Contactants	H-Bonding	Residue B_{iso}	Type IV B_{iso}
Molecule at $(1+x, y, z)$			
Ser 1 ··· Thr 226		10.8 (2.6) ··· 15.2 (2.9)	12.6 (3.8) ··· 7.3 (2.2)
Gln 19 ··· Ala 295		7.0 (1.5) ··· 17.2 (3.7)	10.9 (9.6) ··· 12.5 (3.7)
Ala 24 ··· Gly 296		11.1 (2.1) ··· 21.1 (5.3)	10.6 (3.6) ··· 17.4 (4.0)
Molecule at $(x, y, 1+z)$			
Ala 47 ··· Ala 144	O → O		
Ser 48 ··· Lys 143			
Glu 49 ··· Thr 318	$\mathbf{O^{\epsilon 2}} \rightarrow \mathbf{O^{\gamma 1}}$		
Asp 51 ··· Asp 147	$O^{\delta 2} \rightarrow O$		
Asp 51 ··· Ser 148	$O^{\delta 1}$ or $O^{\delta 2} \rightarrow O^{\gamma}$		
Asp 51 ··· Pro 149			
Gly 52 ··· Asp 147	**N → O**		
Ser 109 ··· Ala 317			
Ser 109 ··· Thr 318	$\mathbf{O^{\gamma}} \rightarrow \mathbf{N}$	11.6 (6.3) ··· 19.0 (3.8)	22.0 (7.0) ··· 32.8 (9.4)
Ser 109 ··· Thr 319	$\mathbf{O^{\gamma}} \rightarrow \mathbf{N}$	11.6 (6.3) ··· 13.9 (5.3)	22.0 (7.0) ··· 19.9 (6.2)
Glu 113 ··· Asn 315	$\mathbf{O^{\epsilon 1}} \rightarrow \mathbf{N^{\delta 2}}$	20.6 (5.4) ··· 11.4 (4.0)	27.5 (7.0) ··· 16.5 (5.6)
Glu 113 ··· Ala 317		20.6 (5.4) ··· 14.9 (5.5)	27.5 (7.0) ··· 17.8 (6.6)
Ser 279 ··· Ser 178	$\mathbf{O^{\gamma}} \rightarrow \mathbf{O^{\gamma}}$		
Thr 280 ··· Gly 177			
Thr 280 ··· Ser 178	$\mathbf{N} \rightarrow \mathbf{O^{\gamma}}$		
Molecule at $(-x, y+1/2, -z)$			
Phe 125 ··· Ser 201			
Ser 126 ··· Ser 201	**N → O**		
Ser 126 ··· Thr 203			
Thr 127 ··· Ser 201	**N → O**		
Thr 127 ··· Gly 202	**N → O**		
Pro 133 ··· Ser 204A			
Thr 134 ··· Phe 203A			
Thr 134 ··· Ser 204A	$\mathbf{O^{\gamma 1}} \rightarrow \mathbf{N}$		
Gln 134A ··· Thr 203			
Gln 134A ··· Phe 203A	**N → O** & $N^{\epsilon 2}$ or $O^{\epsilon 1} \rightarrow N$		
Gln 135 ··· Thr 203			
Lys 136 ··· Thr 203			
Thr 185 ··· Ser 201			

Table 6: continued.

Contactants	H-Bonding	Residue B_{iso}	Type I B_{iso}
Molecule at $(1-x, y+1/2, -z)$			
Lys 64 ··· Gly 177	**$N^{\zeta} \rightarrow O$**		
Molecule at $(-x, y+1/2, 1-z)$			
Thr 70 ··· Lys 238A	**$O^{\gamma} \rightarrow N^{\zeta}$**		
Ser 72 ··· Gly 237	**$O^{\gamma} \rightarrow O$**		
Ser 74 ··· Pro 249			
Ser 80 ··· Ser 236	**$O^{\gamma} \rightarrow O^{\gamma}$**		
Ser 80 ··· Gly 237	$O \rightarrow O^{\gamma}$		
Ser 80A ··· Ser 236			
Ser 81 ··· Ser 236	$N \rightarrow O$		
Ser 81 ··· Gly 237			
Ser 81 ··· Ala 238	**$O^{\gamma} \rightarrow O$**		
Lys 106 ··· Trp 232	**$N^{\zeta} \rightarrow O$**		
Lys 106 ··· Ala 233			
Lys 106 ··· Val 235	**$N^{\zeta} \rightarrow O$**		
Lys 106 ··· Ser 236			
Lys 106 ··· Ala 238	**$N^{\zeta} \rightarrow O$**		
Val 107 ··· Ser 236			
Ser 108 ··· Ser 236	**$N \rightarrow O^{\gamma}$**		
Ser 132 ··· Lys 238A			
Molecule at $(1-x, y+1/2, 1-z)$			
Ser 59 ··· His 158			
Thr 62 ··· Pro 277			
Thr 62 ··· Thr 280			
Thr 62 ··· Gly 281	**$O \rightarrow N$**		
Thr 62 ··· Ser 282	$O^{\gamma 1} \rightarrow O$ or $C^{\beta} \rightarrow O$		
Thr 63 ··· Thr 280			
Thr 63 ··· Gly 281			
Lys 64 ··· Thr 280			

Table 7: The lattice contacts in type I endothiapepsin crystals.

Residues	H-Bonding	Residue B_{iso}	Type I B_{iso}
Molecule at $(x, y, 1+z)$			
Tyr 175 ··· Gly 281			
Ser 178 ··· Gly 281	$O^{\gamma} \rightarrow O$		
Ile 179 ··· Thr 280			
Ile 179 ··· Gly 281	**N → O**		
Ile 179 ··· Ser 282	O → N		
Tyr 181 ··· Ser 279			
Tyr 181 ··· Thr 280	**OH → O**		
Molecule at $(1+x, y, 1+z)$			
Leu 66 ··· Ser 251			
Ser 67 ··· Cys 250			
Ser 67 ··· Ser 251	**O →** $\mathbf{O^{\gamma}}$		
Ser 67 ··· Ser 282A			
Gly 68 ··· Ser 282A			
Pro 133 ··· Pro 249			
Thr 134 ··· Ser 251	$O^{\gamma 1} \rightarrow O^{\gamma}$		
Thr 134 ··· Ala 252	$\mathbf{O^{\gamma 1}}$ **→ N**		
Molecule at $(1-x, y+1/2, -z)$			
Thr 3 ··· Ala 225			
Pro 17 ··· Ser 229			
Gln 19 ··· Ser 229	$\mathbf{O^{\epsilon 1}} \rightarrow \mathbf{O^{\gamma}}$		
Gln 19 ··· Ala 233			
Ala 24 ··· Gln 234	**O →** $\mathbf{N^{\epsilon 1}}$		
Thr 26 ··· Ala 230			
Thr 26 ··· Ala 233			
Asn 28 ··· Lys 204	$O^{\delta 1} \rightarrow N^{\epsilon}$		
Asn 28 ··· Thr 226			
Thr 54 ··· Lys 204			
Pro 162 ··· Ala 295			