SUPPLEMENTARY FIGURES

Contents:

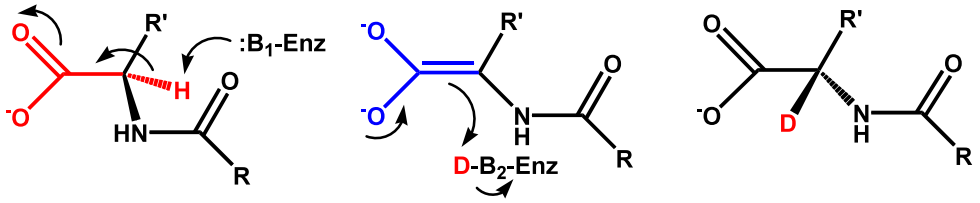


Figure S1. Catalytic mechanism of the dipeptide epimerases. In all cases, the enzyme side chains responsible for protonation and deprotonation are Lys. Here, deuterium is shown being incorporated, as in the mass spectroscopy screening assay. The carboxylate is coordinated to the required magnesium ion.

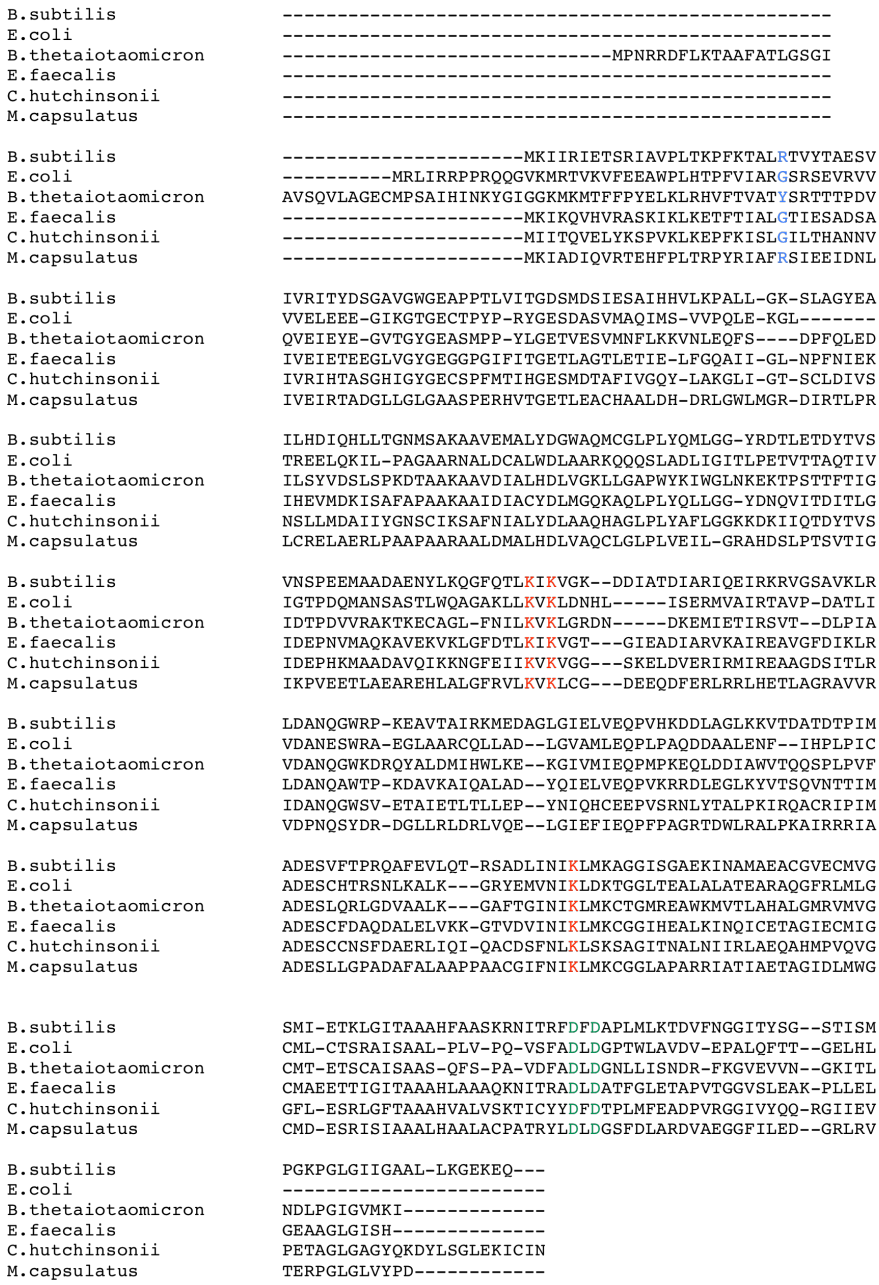


Figure S2. Selected alignments used for generating homology models. Catalytic Lys residues are in red, the DxD motif in green, and the position corresponding to Arg24 in 1TKK (which coordinates the Glu side chain of the Ala-Glu ligand) in red.

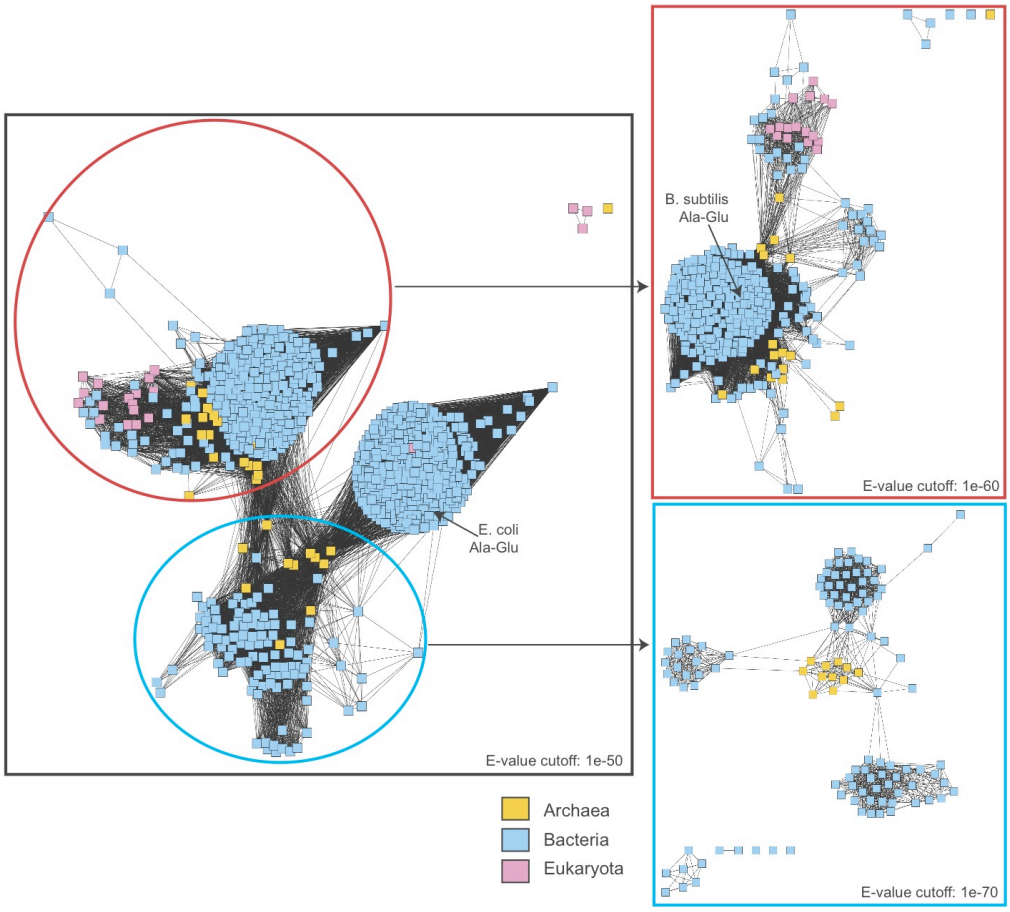


Figure S3. Sequence similarity network showing the distribution of bacterial, archaeal, and eukaryotic sequences. The network is constructed as in Figures 1 and 2.

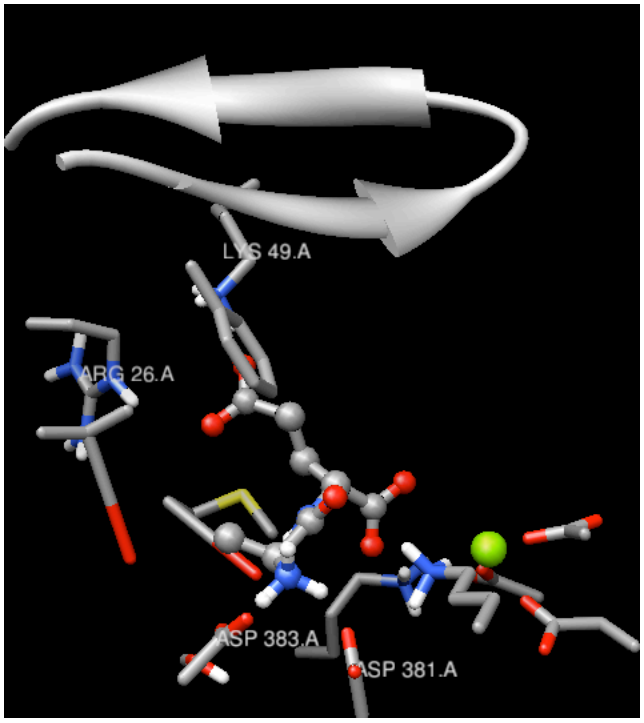


Figure S4. L-Ala-L-Glu modeled in the binding site of a homology model for XP_002649142, a predicted dipeptide epimerase from *Dictyostelium discoideum*. This sequence has not been experimentally characterized.

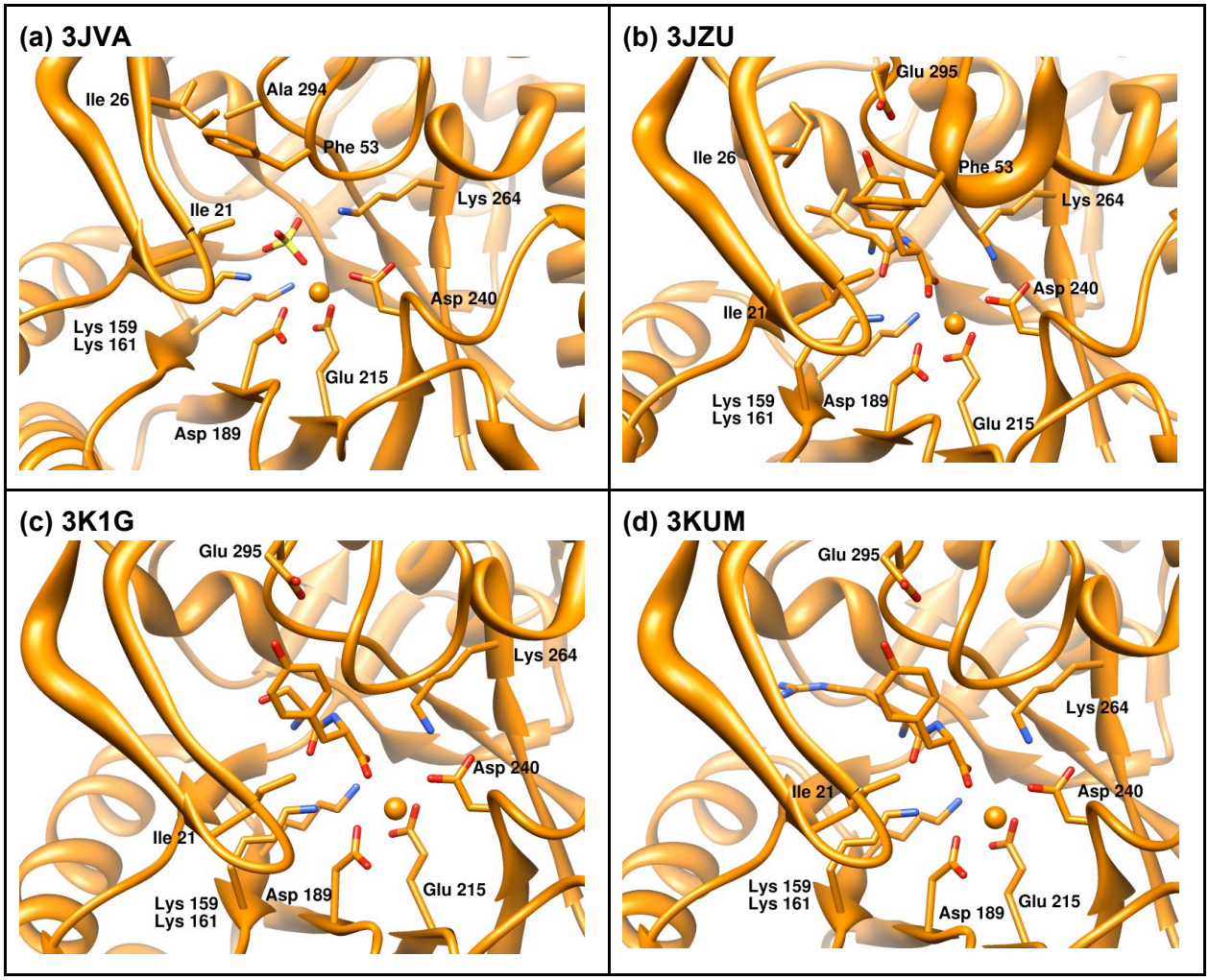


Figure S5. Binding site of the dipeptide epimerase from *E. faecalis*, bound to (a) inorganic sulfate, (b) L-Leu-L-Tyr, (c) L-Ser-L-Tyr, and (d) L-Arg-L-Tyr. The dipeptide and selected active site residues are shown in stick representation.

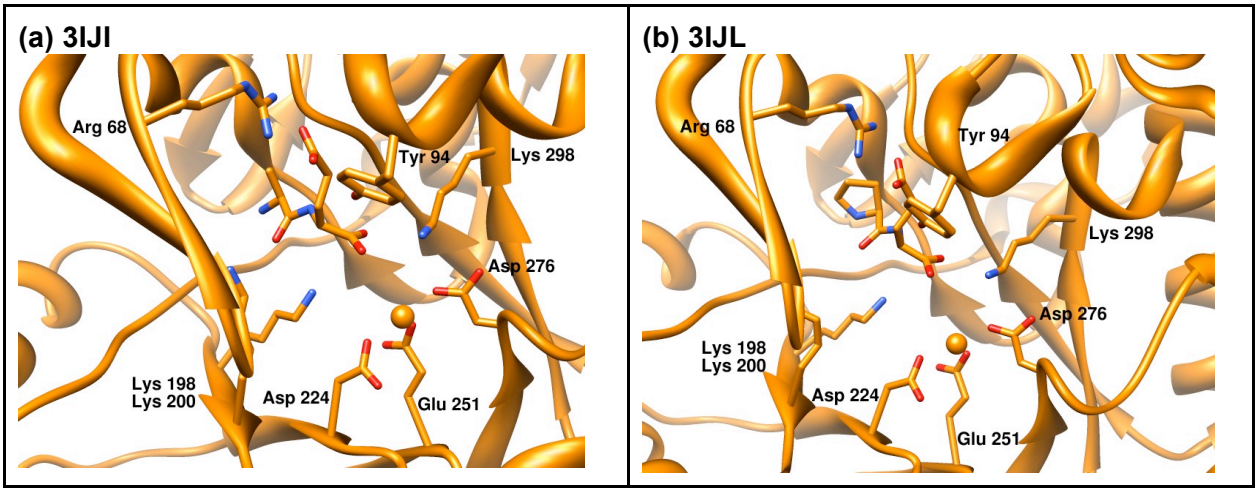


Figure S6. Binding site of the dipeptide epimerase from *B. thetaiotamicron*, bound to (a) L-Ala-D-Glu and (b) L-Pro-D-Glu. Both ligands are bound in a non-catalytically productive conformation, in which the carboxylate group is not in direct contact with the metal ion.

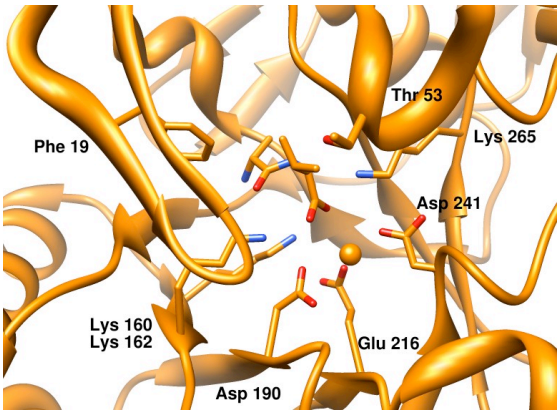


Figure S7. Binding site of the dipeptide epimerase from *C. hutchinsonii*, bound to D-Ala-L-Val (3Q45).

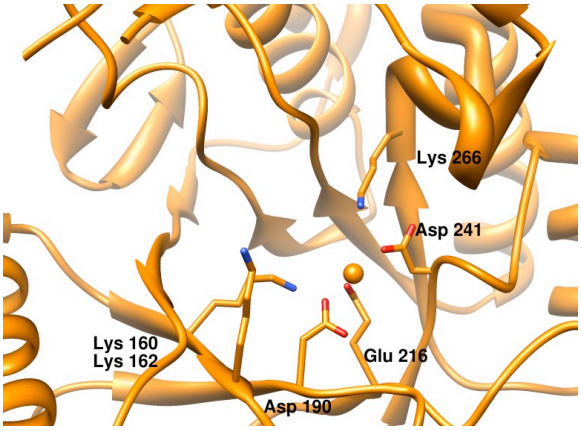


Figure S8. Binding site of the dipeptide epimerase from *M. capsulatus* in the apo state (3RO6).

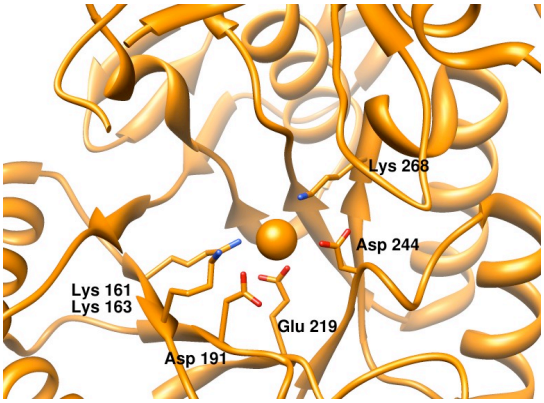


Figure S9. Binding site of the dipeptide epimerase from *H. aurantiacus* in the apo state (3IK4). The metal ion is K^+, not Mg^{2+} as required for catalysis.

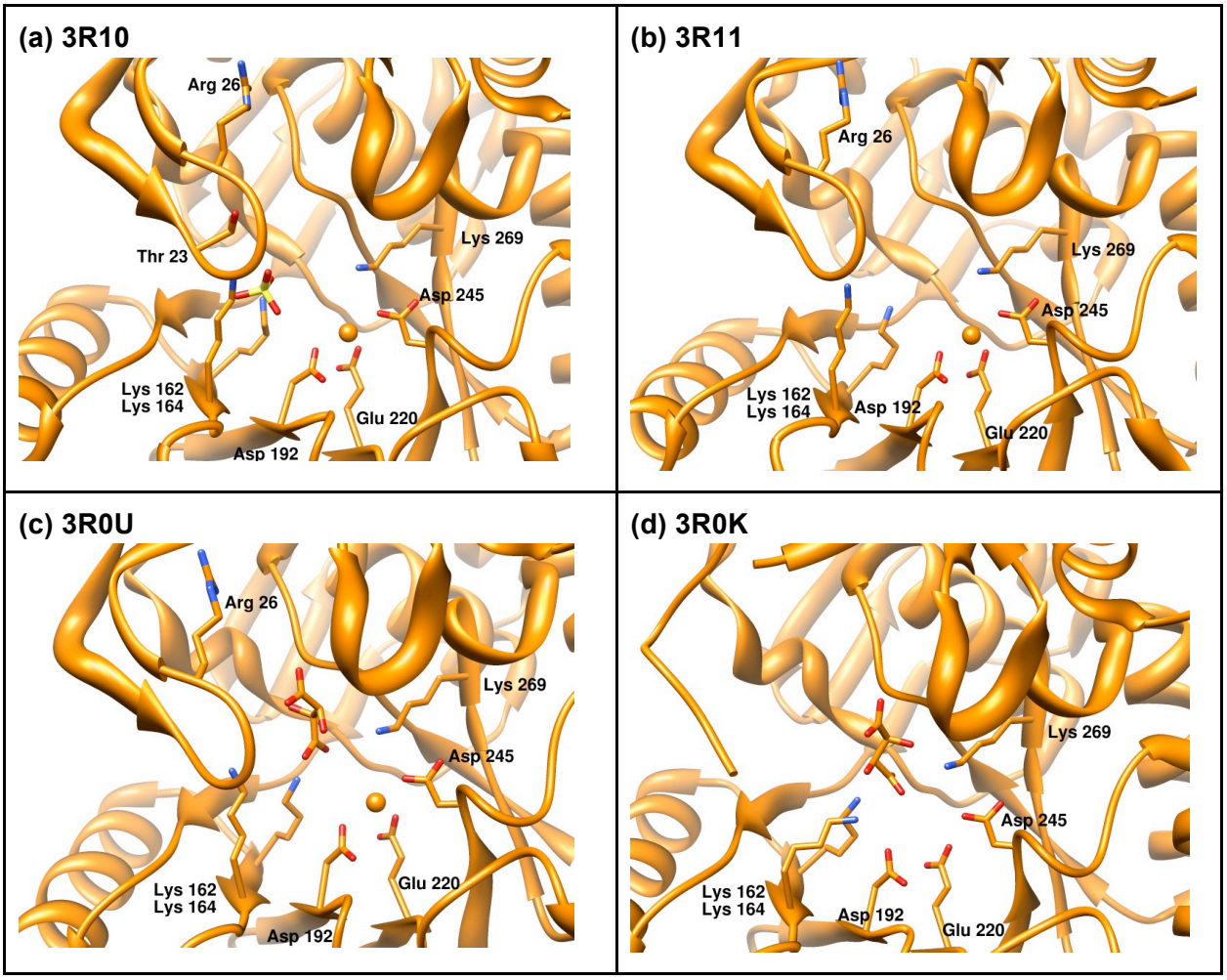
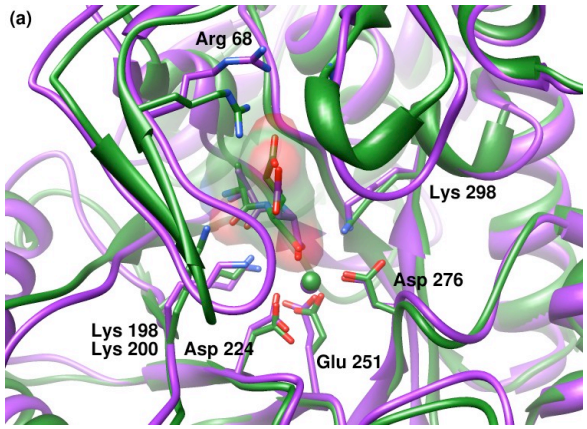
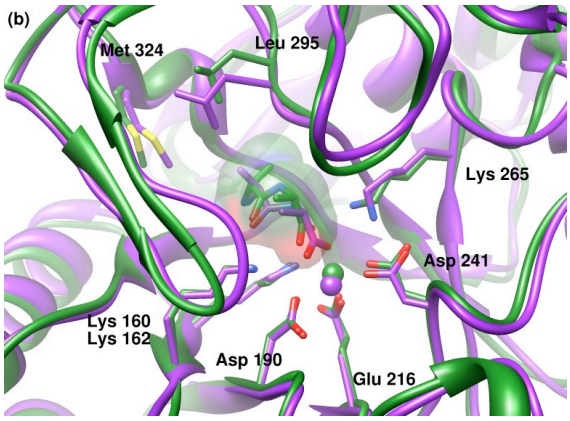


Figure S10. Binding site of the dipeptide epimerase from *F. philomiragia*, (a) with only Mg^{2+} bound, (b) in complex with fumarate, (c) in complex with tartrate, and (d) in complex with tartrate, with no Mg^{2+} bound.

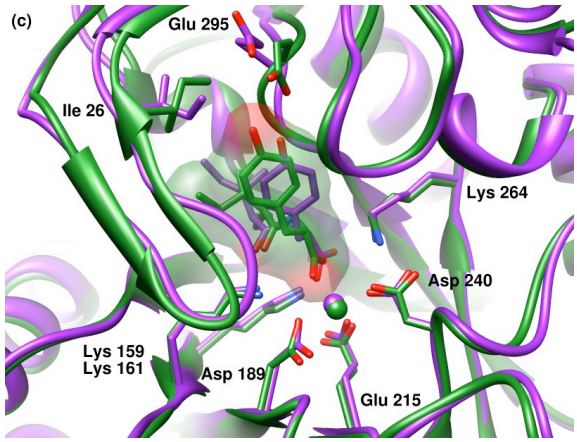
Figure S11. Superposition of representative structures of the dipeptide epimerases with varying specificities, from *B. subtilis* (1TKK, white); *B. thetaiotamicron* (3IJQ, magenta); *E. faecalis* (3JW7, cyan); *C. hutchinsoni* (3Q4D, yellow); *M. capsulatus* (3RIT, red); and *F. philomiragia* (3R1Z, blue). The pairwise Cα RMSD's are 1.1 Å or less.



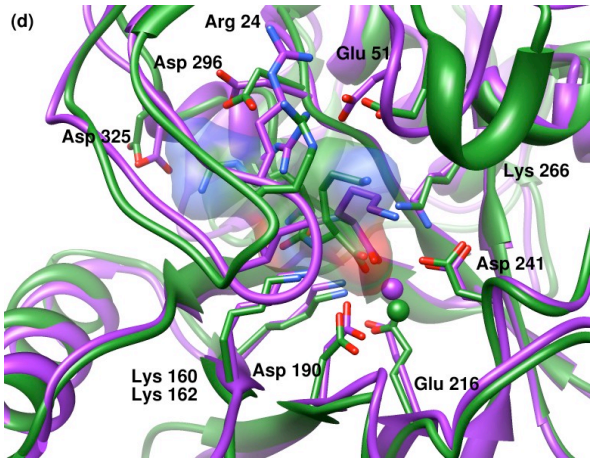
Dipeptide epimerase from *B. thetaiotamicron* in complex with Ala-Glu (3IJI)



Dipeptide epimerase from *C. hutchinsonii* in complex with Ala-Ala (3Q4D)



Dipeptide epimerase from *E. faecalis* in complex with Ile-Tyr (3JW7)



Dipeptide epimerase from *M. capsulatus* in complex with Arg-Lys (3RIT)

Figure S12. Comparisons of computationally predicted (purple) and experimentally determined (green) structures of dipeptide epimerases in complex with dipeptide substrates. The stereochemistries of the computationally predicted ligands are all L amino acids. The stereochemistries assigned to the co-crystallized ligands are detailed in Table 2.

SUPPLEMENTARY TABLES

Table S1: Information about homology models.

Table S2: Dipeptide epimerases characterized *in vitro*.

Table S3. Screening results with the L-Ala-L-Xxx library.

Tables S4-S20: Complete results from the mass spectroscopy based assays.

Tables S21-S25: Complete kinetic results obtained for selected dipeptide epimerases.

Tables S26-S31: Data collection and refinement statistics for all determined crystal structures.

Table S1: Information about homology modeled sequences, organized according to predicted specificity classes: organisms, sequence identity to the *Bacillus subtilis* template structure (1TKK), RMSDs between the model and the crystal structure, in cases where a structure was subsequently obtained.

gi number	Organism	Sequence identity to 1tkk (%)	zdope Model score[(a)]	RMSD (Å) Overall/ Binding site	Ligand RMSD (Å)
Ala-Glu (B.subtilis like)					
48477093	Picrophilus torridus DSM 9790	31.1	-1.09		
29376078	Enterococcus faecalis V583	45.2	-1.23	2.1/1.3	1.3
15615568	Bacillus halodurans C-125	40.9	-1.31		
30020968	Bacillus cereus ATCC 14579	49.4	-1.53		
30262817	Bacillus anthracis str. Ames	49.4	-1.56		
49479091	Bacillus thuringiensis serovar konkukian str. 97-27	50.0	-1.51		
52142662	Bacillus cereus E33L	50.6	-1.51		
42781935	Bacillus cereus ATCC 10987	50.3	-1.55		
47567668	Bacillus cereus G9241	49.7	-1.61		
23100420	Oceanobacillus iheyensis HTE831	61.6	-1.72		
18158850	Bacillus subtilis	100.0	-1.72	0.0/0.5	1.0
52079796	Bacillus licheniformis ATCC 14580 (DSM 13)	69.3	-1.66		
21233235	Xanthomonas campestris pv. campestris str. ATCC 33913	47.1	-1.25		
58583725	Xanthomonas oryzae pv. oryzae KACC10331	46.7	-1.35		
21110264	Xanthomonas axonopodis pv. citri str. 306	47.0	-1.38		
19712800	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	47.4	-1.62		
15893485	Clostridium acetobutylicum ATCC 824	52.0	-1.60		
54297883	Legionella pneumophila str. Paris	39.7	-1.38		
52842173	Legionella pneumophila subsp. pneumophila str. Philadelphia 1	40.0	-1.40		
54294846	Legionella pneumophila str. Lens	40.0	-1.38		
Lys-Xxx					
43283124	Environmental sequence	33.0	-1.25		
44360464	Environmental sequence	29.4	-1.23		
44581982	Environmental sequence	29.4	-1.24		
44237260	Environmental sequence	29.8	-1.20		
44534308	Environmental sequence	30.1	-1.30		
37522668	Gloeobacter violaceus PCC 7421	36.0	-1.08		

53757661	Methylococcus capsulatus str. Bath	31.7	-1.11	1.7/1.8	1.8
Ala-Hyd (group 1)[(b)]					
15642781	Thermotoga maritima MSB8	33.7	-1.47		
55229873	Haloarcula marismortui ATCC 43049	33.9	-1.12		
22971600	Chloroflexus aurantiacus	32.7	-1.29		
55770159	Oryza sativa (japonica cultivar-group)	35.1	-1.21		
18401824	Arabidopsis thaliana (thale cress)	31.6	-1.28		
Ala-Hyd (group 2)[(b)]					
44367065	Environmental sequence	32.3	-1.33		
48856714	Cytophaga hutchinsonii	38.7	-1.21	1.4/0.7	1.0
48864913	Oenococcus oeni PSU-1	29.8	-1.09		
16799274	Listeria innocua Clip11262	40.8	-1.42		
52858062	Lactobacillus gasseri	37.6	-1.65		
Ala-Glu (E.coli like)					
48765618	Rhodospirillum rubrum	27.2	-0.72		
50122039	Pectobacterium atrosepticum SCRI1043	28.9	-1.16		
53690990	Desulfovibrio desulfuricans G20	29.6	-0.77		
22958170	Rhodobacter sphaeroides 2.4.1	29.7	-0.92		
52009879	Silicibacter sp. TM1040	29.9	-0.97		
56680164	Ruegeria pomeroyi DSS-3	29.2	-0.73		
17935543	Rubrobacter xylanophilus DSM 9941Agrobacterium tumefaciens str. C58	29.7	-1.01		
23501898	Brucella suis 1330	30.0	-0.97		
15965587	Sinorhizobium meliloti 1021	31.9	-0.90		
13470960	Mesorhizobium loti MAFF303099	31.6	-1.02		
45917069	Mesorhizobium sp. BNC1	30.3	-0.97		
27381006	Bradyrhizobium japonicum USDA 110	27.8	-0.75		
48783179	Burkholderia fungorum LB400	29.9	-0.68		
45916315	Mesorhizobium sp. BNC1	25.8	-0.84		
16127343	Caulobacter crescentus CB15	28.0	-0.52		1.0
56543698	Zymomonas mobilis subsp. mobilis ZM4	24.5	-0.92		
50119100	Pectobacterium atrosepticum SCRI1043	25.2	-0.74		
16122565	Yersinia pestis CO92	28.4	-0.94		
26108063	Escherichia coli CFT073	28.5	-1.09	2.3/1.7	
6759977	Shigella dysenteriae	29.0	-1.08		
38703962	Escherichia coli O157:H7 str. Sakai	28.7	-1.07		
30062843	Shigella flexneri 2a str. 2457T	29.0	-1.06		1.3
2506874	Escherichia coli K-12	29.0	-1.07		
16760206	Salmonella enterica subsp. enterica serovar Typhi str. CT18	30.9	-1.00		
16765024	Salmonella enterica subsp. enterica	30.3	-0.97		

	serovar Typhimurium str. LT2				
56413397	Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150	30.3	-0.95		
46913516	Photobacterium profundum SS9	27.8	-0.73		
28900184	Vibrio parahaemolyticus RIMD 2210633	29.0	-0.78		
44422107	Environmental sequence	28.8	-1.12		
44343066	Environmental sequence	29.9	-1.06		
44393977	Environmental sequence	27.9	-0.98		
44535747	Environmental sequence	27.3	-1.21		
44273181	Environmental sequence	26.7	-1.21		1.8
44240938	Environmental sequence	26.7	-1.20		
Xxx-Glu					
45513787	Synechococcus elongates PCC 7942	24.2	-1.10		
23124447	Nostoc punctiforme PCC 73102 (Nostoc punctiforme ATCC 29133)	30.9	-1.24		
17231024	Nostoc sp. PCC 7120	29.7	-1.15		
46135511	Anabaena variabilis ATCC 29413 (Anabaena flos-aquae UTEX 1444)	29.4	-1.20		
44544497	Environmental sequence	27.8	-0.92		
44635156	Environmental sequence	27.5	-0.72		2.1
44618589	Environmental sequence	26.0	-0.97		1.0
44418089	Environmental sequence	24.2	-0.84		
32476798	Rhodopirellula baltica SH 1	27.8	-0.87		
29346723	Bacteroides thetaiotaomicron VPI-5482	29.4	-1.02	2.2/2.0	2.1
53714038	Bacteroides fragilis YCH46	30.3	-0.95		
32475496	Rhodopirellula baltica SH 1	25.7	-0.96		
23135920	Cytophaga hutchinsonii	23.1	-0.45		
55232713	Haloarcula marismortui ATCC 43049	31.1	-1.01		
15790741	Halobacterium sp. NRC-1	32.9	-0.81		

(a) zdope score (http://modbase.compbio.ucsf.edu/evaluation/).
(b) The two groups of predicted Ala-Hyd dipeptide epimerases were phylogenetically distinct but each was assigned the same predicted specificity; we have merged them into one specificity group in Figure 1 for simplicity.

Table S2: Dipeptide epimerases characterized *in vitro*.

gi number	Genbank annotation	Organism
53803900	Chloromuconate cycloisomerase	*Methylococcus capsulatus*
29376078	mandelate racemase/muconate lactonizing enzyme family protein	*Enterococcus facaelis*
223936049	mandelate racemase/muconate lactonizing protein	*Pedosphaera parvula Ellin514*
159897643	mandelate racemase/muconate lactonizing protein	*Herpetosiphon aurantiacus ATCC 23779*
29346723	Muconate cycloisomerase	*Bacteroides thetaiotaomicron*
110638536	mandelate racemase/muconate lactonizing family protein	*Cytophaga hutchinsonii ATCC 33406*
224368313	chloromuconate cycloisomerase-like protein	*Desulfobacterium autotrophicum*
225012942	mandelate racemase/muconate lactonizing protein	*Flavobacteria bacterium MS024-2A*
305667445	Chloromuconate cycloisomerase	*Flavobacteriales bacterium HTCC2170*
118497296	Hypothetical protein	*Francisella tularensis subsp. novicida U112*
146299924	mandelate racemase/muconate lactonizing protein	*Flavobacterium johnsoniae UW101*
163735722	mandelate racemase/muconate lactonizing enzyme	*Roseobacter litoralis Och 149*
167627873	Enolase superfamily protein	*Francisella philomiragia subsp. philomiragia ATCC 25017*
224118774	Predicted protein	*Populus trichocarpa*
254457379	mandelate racemase/muconate lactonizing enzyme	*Campylobacterales bacterium*
197118796	L-alanyl-D/L-glutamate epimerase	*Geobacter bemidjiensis Bem*
78777979	mandelate racemase/muconate lactonizing protein	*Sulfurimonas denitrificans DSM 1251*

Table S3. Screening results with the L-Ala-L-Xxx library. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Ala	Ser	Pro	Val	Thr	Leu/ Ile	Asn	Asp	Lys	Gln	Glu	Met	His	Phe	Arg	Tyr	Trp
M. capsulatus[1]													+		+		
E. faecalis	++					+						+	+++	++		+++	+
D. autotrophicum									++						+++		
bacterium Ellin514	+++	+++			+++	+	+	+		+	+	+++	++	+++		+++	++
Campylobacterales										+	++	+					
G. bemidjiensis	++	+	+		+++	++				+	+	+++		+++		+++	
S. denitrificans					+						++	+					
H. aurantiacus	+++	+++		+++	+++	+++						+++		++		+++	+++
Flavobacteria							+	+				++		++		+++	+
Flavobacteriales	+++	+++			+++	++				++	++	+++		+	+	+	+
F. tularensis							+	+			++		+				
F. johnsoniae	+++	+++			+++	+++		++	++	++	++	+++	+++	+++	++	+++	+++
R. litoralis		++			+++	++		++			+	+++	+++	+++	++	+++	+++
F. philomiragi								+			++						
P. trichocarpa	+++	+++			+++	+						+		+		+	+++
B.thetaiotaomicron											+++						
C. hutchinsonii	+++	+++		+++	+++	+++						+++		+++		++	+

[1] Full species names and gi numbers of the sequences are provided in Supplementary Table S2.

Table S4. Mass spectroscopy screening data for the dipeptide epimerase from bacterium Ellin514. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Xxx-L-Ala	L-Xxx-L-Ile	L-Xxx-L-Leu	L-Xxx-L-Asp	L-Xxx-L-Glu	L-Xxx-L-Phe	L-Xxx-L-Tyr
Gly							+++			
Ala		+++		++			++	+++	+++	+++
Ser	+++	+++		+					++	+
Pro										
Val				+++			+++	+++	+++	++
Thr	+	+++					+	+	+	+
Leu/Ile		+		++			+	+	+	+
Asn		+								
Asp		+								
Lys										
Gln	+	+								
Glu	+	+								
Met	++	+++						++		
His		++								
Phe	+	+++								+
Arg										
Tyr		+++								+
Trp		++		++					+	+

Table S5. Mass spectroscopy screening data for the dipeptide epimerase from *Sulfurimonas denitrificans* DSM 1251. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly- L-Xxx	L-Ala- L-Xxx		L-Xxx- L-Glu
Gly				+++
Ala				+++
Ser				+++
Pro				+++
Val				+++
Thr		+		+++
Leu/Ile				++
Asn				+
Asp				
Lys				
Gln	+			
Glu	+	++		
Met		+		+++
His				
Phe				+++
Arg				
Tyr				++
Trp				+

Table S6. Mass spectroscopy screening data for the dipeptide epimerase from *Campylobacterales bacterium*. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Xxx-L-Asp	L-Xxx-L-Glu
Gly					
Ala					+++
Ser					
Pro					++
Val					++
Thr					+
Leu/Ile					+
Asn					
Asp					
Lys				+	
Gln		+			
Glu	+	++			
Met		+			+
His					
Phe					+
Arg				+	
Tyr					
Trp					

Table S7. Mass spectroscopy screening data for the dipeptide epimerase from *Herpetosiphon aurantiacus* ATCC 23779. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Xxx-L-Ala	L-Xxx-L-Ile	L-Xxx-L-Leu	L-Xxx-L-Tyr	L-Xxx-L-Phe	L-Xxx-L-Asp	L-Xxx-L-Glu
Gly				+++	+	+				
Ala	+++	+++		++	+	+++	++	+++		
Ser	+++	+++					+			
Pro										
Val	+++	+++					++	++		
Thr	++	+++						+		
Leu/Ile	+++	+++					+	+		
Asn										
Asp										
Lys										
Gln										
Glu										
Met	+++	+++				+	+++	+++		
His										
Phe	+	++		+		++	+++	+++		
Arg										
Tyr	+	+++		+		+	+++	++		
Trp	+++	+++		+++		+	+++	+		

Table S8. Mass spectroscopy screening data for the dipeptide epimerase from *Methylococcus capsulatus.* Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Arg-L-Xxx	L-Lys-L-Xxx	L-Orn-L-Xxx		L-Xxx-L-Arg	L-Xxx-L-Lys	L-Xxx-L-Orn	L-Xxx-L-His
Gly											
Ala				+				+++			+++
Ser				+++	+	+		+			+++
Pro											
Val								+++	++		+++
Thr				++	++			+			
Leu/Ile								+++	++		+++
Asn						+					
Asp											
Lys				+++	++			+++	+++	++	+++
Gln											
Glu											
Met				+	+			++	+		+
His	++	+		+++	+	++					
Phe				+							
Arg	+	+		+++	+++	++		+++	+++	+++	+++
Tyr				++	++	+		+	+		++
Trp								+	+		++

Table S9. Mass spectroscopy screening data for the dipeptide epimerase from *Desulfobacterium autotrophicum*. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly- L-Xxx	L-Ala- L-Xxx		L-Xxx- L-Arg	L-Xxx- L-Lys	L-Xxx- L-Orn
Gly						
Ala						
Ser						
Pro						
Val				+++	+++	+
Thr						
Leu/Ile				+	+	
Asn						
Asp						
Lys	+	++		+++	+++	+++
Gln						
Glu						
Met						
His						
Phe					+	
Arg	++	+++		+++	+++	+++
Tyr					+	
Trp						

Table S10a. Mass spectroscopy screening data for the dipeptide epimerase from *Geobacter bemidjiensis* Bem. Specificity for C-terminal position. Deuterium incorporation was measured at 16 hours after incubation, unless otherwise specified. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	Gly-L-Xxx (1hr)	Gly-L-Xxx (5hr)	L-Ala-L-Xxx	L-Ala-L-Xxx (1hr)	L-Ala-L-Xxx (5hr)		L-Phe-L-Xxx (1hr)	L-Phe-L-Xxx (5hr)	L-Met-L-Xxx (5hr)
Gly										
Ala	+++	+	+	++		+			+	+
Ser	++	+	+	+				+	+++	+
Pro	+			+						
Val	+		+							
Thr	+++	+	++	+++	+	++		+	+++	+
Leu/Ile	++	+	+	++	+	+				
Asn	+									
Asp	+									
Lys										
Gln	+			+						
Glu	+			+						
Met	+++	+++	+++	+++	+++	+++		++	+++	+
His										
Phe	+++	+	+++	+++	+++	+++		+	++	+
Arg										
Tyr				+++		+			++	+
Trp										

Table S10b. Mass spectroscopy screening data for the dipeptide epimerase from *Geobacter bemidjiensis* Bem. Specificity for N-terminal position. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	L-Xxx-L-Ala	L-Xxx-L-Ile	L-Xxx-L-Leu	L-Xxx-L-Tyr	L-Xxx-L-Phe	L-Xxx-L-Asp	L-Xxx-L-Glu
Gly		+	+	+++	+++		+++
Ala	++	+++	+++	+++	++		+++
Ser					++		
Pro							
Val	+				+		
Thr					+		
Leu/Ile	+				+		
Asn							
Asp							
Lys							
Gln							
Glu							
Met	+			+	++		
His							
Phe	+++			+++	+++		++
Arg							
Tyr	+++			++	+++		+
Trp	++			+++	+++		

Table S11. Mass spectroscopy screening data for the dipeptide epimerase from *Francisella philomiragia* subsp. philomiragia ATCC 25017. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Thr-L-Xxx	L-Ile-L-Xxx	L-Met-L-Xxx	L-Phe-L-Xxx		L-Xxx-L-Glu
Gly									
Ala									+++
Ser		++					+		
Pro									+++
Val									+++
Thr		++			+		+		+++
Leu/Ile									+
Asn									
Asp		++			+				
Lys									
Gln									
Glu	+++	+++		++	++	++	+		
Met									+++
His									
Phe									+
Arg									
Tyr									
Trp									

Table S12. Mass spectroscopy screening data for the dipeptide epimerase from *Enterococcus faecalis.* Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Thr-L-Xxx	L-Arg-L-Xxx		L-Xxx-L-Tyr	L-Xxx-L-Phe	L-Xxx-L-His
Gly									
Ala		++					++	+	
Ser							+		
Pro							+		
Val							+++	+++	+
Thr							+		
Leu/Ile		+		+			++	+++	+
Asn								+	
Asp									
Lys							+	+	
Gln								+	
Glu									
Met		+		+	+		++	++	
His		+++			+		+		
Phe		++		+	+		+++	+++	+
Arg							+	++	+
Tyr	++	+++		++	+		++	++	
Trp		+					++	+++	+

Table S13a. Mass spectroscopy screening data for the dipeptide epimerase from *Roseobacter litoralis* Och 149. C-terminal specificity. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Ile-L-Xxx	L-Thr-L-Xxx	L-Met-L-Xxx	L-Phe-L-Xxx
Gly							
Ala				+++			++
Ser		++		+++			+++
Pro							
Val				+		+	+
Thr		+++			+++	+	+++
Leu/Ile		++		++	+	+	
Asn							
Asp		++					
Lys				++	++		+++
Gln							
Glu		+					
Met		+++		+++	+++	++	+++
His		+++		+++	+		+++
Phe	+	+++		+++	+++	+	+++
Arg		++		+	+++		+++
Tyr		+++		+++	+++	++	+++
Trp		+++		+++		+	+++

Table S13b. Mass spectroscopy screening data for the dipeptide epimerase from *Roseobacter litoralis* Och 149. N-terminal specificity. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	L-Xxx-L-Ile	L-Xxx-L-Leu	L-Xxx-L-Phe	L-Xxx-L-Tyr	L-Xxx-L-Arg	L-Xxx-L-Asp
Gly						
Ala			+++	+++	++	
Ser			+++		+	
Pro						
Val	+	+	+++			+
Thr			+++		+	
Leu/Ile	+	++	+++		++	+
Asn						
Asp					+	
Lys						
Gln						
Glu					+	
Met	+	+	+++		+	++
His						
Phe	+		+++		+++	
Arg						
Tyr			+++		+++	
Trp			+++		+++	

Table S14. Mass spectroscopy screening data for the dipeptide epimerase from *Flavobacteria bacterium* MS024-2A. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Phe-L-Xxx	L-Met-L-Xxx	L-Ile-L-Xxx		L-Xxx-L-Phe	L-Xxx-L-Tyr	L-Xxx-L-Asp
Gly										
Ala								+	++	
Ser								+	+	
Pro										
Val								+		
Thr									+	
Leu/Ile								+	++	
Asn		+								
Asp		+								
Lys										
Gln										
Glu										++
Met		++						+	++	
His										
Phe		++								
Arg										
Tyr		+++							+	
Trp		+						+	++	

Table S15. Mass spectroscopy screening data for the dipeptide epimerase from *Populus trichocarpa.* Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Phe-L-Xxx	L-Met-L-Xxx		L-Xxx-L-Ala	L-Xxx-L-Ile	L-Xxx-L-Leu	L-Xxx-L-Tyr	L-Xxx-L-Phe
Gly											
Ala		+++								+	+
Ser		+++									
Pro											
Val							+				
Thr		+++									
Leu/Ile	+	+									
Asn											
Asp											
Lys											
Gln											
Glu											
Met	+++	+								++	+
His											
Phe		+					+			++	+
Arg											
Tyr		+									
Trp		+++					+				

Table S16. Mass spectroscopy screening data for the dipeptide epimerase from *Francisella tularensis* subsp. novicida U112. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Phe-L-Xxx	L-Met-L-Xxx	L-Thr-L-Xxx	L-Ile-L-Xxx		L-Xxx-L-Glu
Gly									
Ala									+++
Ser									+++
Pro									
Val									+++
Thr				+					+++
Leu/Ile									+
Asn		+							
Asp		+							
Lys									
Gln									
Glu	+++	++		++	+	+++	+		
Met									++
His		+							
Phe									+
Arg									
Tyr									
Trp									

Table S17. Mass spectroscopy screening data for the dipeptide epimerase from *Flavobacterium johnsoniae* UW101. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx	L-Thr-L-Xxx		L-Xxx-L-Ala	L-Xxx-L-Ile	L-Xxx-L-Leu	L-Xxx-L-Phe	L-Xxx-L-Arg/Lys	L-Xxx-L-Glu	L-Xxx-L-Asp
Gly											
Ala		+++				+		+++		+++	
Ser		+++									+++
Pro											
Val					+					+	
Thr		+++									
Leu/Ile	+	+++									
Asn											
Asp		++									
Lys		++									
Gln		++									
Glu	+	++									
Met	+	+++									
His		+++									
Phe		+++									
Arg		++									
Tyr		+++									+++
Trp		+++									

Table S18. Mass spectroscopy screening data for the dipeptide epimerase from *Flavobacteriales bacterium* HTCC2170. Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Xxx-L-Ala	L-Xxx-L-Leu	L-Xxx-L-Ile	L-Xxx-L-Phe	L-Xxx-L-Tyr	L-Xxx-L-Asp	L-Xxx-L-Glu
Gly										
Ala		+++					+	+		+++
Ser		+++								
Pro										
Val										
Thr		+++								
Leu/ Ile		++								
Asn										
Asp										
Lys										
Gln		++								
Glu		++								
Met		+++								
His										
Phe		+								
Arg		+								
Tyr		+								
Trp		+								

Table S19. Mass spectroscopy screening data for the dipeptide epimerase from *Bacteroides thetaiotaomicron.* Deuterium incorporation was measured at 16 hours after incubation, unless otherwise specified. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Xxx-L-Glu (16hrs)	L-Xxx-L-Glu (0.5hr)
Gly					
Ala				+++	+++
Ser				+++	
Pro				+++	+++
Val				+++	+++
Thr				+++	
Leu/Ile				+++	+
Asn					
Asp					
Lys				+	
Gln				+	
Glu	+++	+++			
Met				+++	
His				++	
Phe				+++	
Arg					
Tyr				++	
Trp				+++	

Table S20. Mass spectroscopy screening data for the dipeptide epimerase from *Cytophaga hutchinsonii.* Deuterium incorporation was measured at 16 hours after incubation. "+++" indicates >50% incorporation of deuterium, "++"represents 25-50%, and "+" represents <25% incorporation. Blanks indicate no detectable incorporation of deuterium. "na" indicates that the specified dipeptide was not included in the screening.

	Gly-L-Xxx	L-Ala-L-Xxx		L-Xxx-L-Ala	L-Xxx-L-Ile	L-Xxx-L-Leu	L-Xxx-L-Phe	D-Ala-L-Xxx
Gly								
Ala	+++	+++		+++	++		+++	+++
Ser		+++						na
Pro				+++				na
Val	+++	+++						na
Thr	++	+++						na
Leu/Ile	+++	+++		+++			++	++
Asn								
Asp				na				
Lys								
Gln								
Glu				na				
Met	+++	+++		++			++	+++
His								
Phe	+++	+++		+++			+++	++
Arg								
Tyr		++		+++				
Trp	+	+		+				

Table S21. Kinetic constants obtained for epimerization of dipeptide substrates for *Bacteroides thetaiotaomicron*.

	k_{cat} [s^{-1}]	K_M [mM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
L-Ala-L-Glu	**147 ± 10**	**2.0 ± 1**	**7.4×10^4**
L-Ala-D-Glu	**59 ± 7**	**6.0 ± 2**	**9.8×10^3**
L-Val-L-Glu	**96 ± 7**	**1.7 ± 0.2**	**5.6×10^4**
L-Val-D-Glu	**43 ± 6**	**4.4 ± 2**	**9.8×10^3**
L-Ile-L-Glu	**30 ± 2**	**5.9 ± 1**	**5.1×10^3**
L-Leu-L-Glu	**-**	**-**	**3.9×10^3**
L-Pro-L-Glu	**-**	**-**	**1.0×10^3**

Table S22. Kinetic constants obtained for epimerization of dipeptide substrates for *Cytophaga hutchinsonii*.

	k_{cat} [s^{-1}]	K_M [mM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
L-Ala-L-Ala	**37 ± 3**	**5.0 ± 0.7**	**7.5×10^3**
L-Ala-L-Val	**7.5**	**10**	**7.5×10^2**
L-Ala-L-Met	**15**	**2.5**	**6.0×10^3**
L-Ala-L- Leu	**1.4**	**1.4**	**1.0×10^3**
L-Val-L-Ala	**5.5 ± 0.05**	**9.9 ± 0.3**	**5.6×10^2**
L-Phe-L-Ala	**7.0 ± 0.3**	**3.4 ± 0.6**	**2.1×10^3**
L-Leu-L-Ala	**21 ± 0.8**	**22 ± 2**	**9.4×10^2**
L-Tyr-L-Ala	**3.0 ± 0.1**	**3.6 ± 0.6**	**8.3×10^2**
D-Ala-D-Ala	**43 ± 1.3**	**1.9 ± 0.2**	**2.3×10^4**
D-Ala-L-Ala	**58 ± 4**	**1.1 ± 0.3**	**5.3×10^4**
D-Ala-L-Val	**19 ± 0.8**	**0.50 ± 0.1**	**3.7×10^4**
D-Ala-L-Met	**15 ± 0.9**	**2.5 ± 0.4**	**6.0×10^3**
D-Ala-L-Leu	**5.0 ± 0.2**	**1.1 ± 0.1**	**4.5×10^3**
D-Ala-L-Ser	**68**	**6.2**	**1.1×10^4**

Table S23. Kinetic constants obtained for epimerization of dipeptide substrates for *Enterococcus faecalis*.

	k_{cat} [s^{-1}]	K_M [mM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
L-Ile-L-Tyr	**9.2 ± 0.7**	**0.77 ± 0.2**	**1.2×10^4**
L-Val-L-Tyr	**8.7 ± 2**	**0.70 ± 0.4**	**1.4×10^4**
L-Arg-L-Tyr	**15 ± 0.07**	**1.4 ± 0.2**	**1.0×10^4**
L-Leu-L-Tyr	**6.3 ± 0.5**	**0.81 ± 0.2**	**8.1×10^3**
L-Tyr-L-Tyr	**8.9 ± 0.09**	**1.7 ± 0.3**	**5.4×10^3**
L-Phe-L-Phe	**3.9 ± 0.04**	**1.0 ± 0.05**	**3.8×10^3**
L-Ser-L-Tyr	**5.0 ± 0.2**	**1.9 ± 0.5**	**2.8×10^3**
L-Ala-L-Tyr	**8.8 ± 3**	**4.3 ± 0.6**	**2.0×10^3**
L-Ala-L-His	**-**	**-**	**1.4×10^3**
L-Leu-L-Phe	**2.7 ± 0.07**	**3.4 ± 0.2**	**7.7×10^2**
L-Arg-L-Phe	**7.5 ± 0.6**	**9.9 ± 1**	**7.6×10^2**
L-Ala-L-Leu	**-**	**-**	**7.1×10^2**
L-Tyr-L-Lys	**-**	**> 30**	**0.61**
L-Trp-L-Lys	**-**	**> 30**	**0.40**
L-Ile-L-Lys	**-**	**> 10**	**0.17**

Table S24. Kinetic constants obtained for epimerization of dipeptide substrates for *Herpetosiphon aurantiacus*.

	k_{cat} [s^{-1}]	K_M [mM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
L-Phe-L-Tyr	**4.7 ± 1**	**4.8 ± 1**	**980**
L-Tyr-L-Tyr	**0.19 ± 0.06**	**8.0 ± 3**	**23**
L-Ala-L-Tyr	**-**	**-**	**9.9**
L-Ala-L-Phe	**0.025 ± 0.003**	**5.2 ± 0.8**	**4.8**
L-Met-L-Tyr	**-**	**-**	**2.7**

Table S25. Kinetic constants obtained for epimerization of dipeptide substrates for *Methylococcus capsulatus*.

	k_{cat} [s^{-1}]	K_M [mM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
L-Ala-D-Glu	**0**	**-**	**0**
L-Lys-L-Arg	**8.4 ± 1**	**0.44 ± 0.1**	**1.9×10^4**
L-Lys-L-His	**2.4 ± 0.1**	**0.88 ± 0.01**	**2.8×10^3**
L-Lys-L-Lys	**0.029 ± 0.001**	**0.15 ± 0.02**	**200**
L-Lys-L-Asp	**0.73 ± 0.05**	**3.1 ± 1**	**240**
L-Lys-Citrulline	**0.023**	**-**	**-**
L-Arg-L-Arg	**0.72 ± 0.4**	**0.19 ± 0.03**	**3.6×10^3**
L-Arg-L-His	**4.2 ± 0.7**	**5.4 ± 3**	**770**
L-Arg-L-Orn	**0.16 ± 0.02**	**0.21 ± 0.08**	**730**
L-Arg-L-Ser	**0.12 ± 0.05**	**1.5 ± 1**	**78**
L-Arg-L-Lys	**0.086 ± 0.03**	**1.5 ± 0.8**	**58**
L-Orn-L-His	**1.0 ± 0.09**	**2.9 ± 0.6**	**350**
L-Orn-L-Arg	**0.016 ± 0.004**	**1.3 ± 0.7**	**12**
L-Val-L-Arg	**0.22 ± 0.02**	**0.25 ± 0.09**	**850**
L-Leu-L-Arg	**0.0061 ± 0.0004**	**0.094 ± 0.002**	**65**
L-Ile-L-His	**0.024 ± 0.005**	**0.91 ± 0.5**	**27**
Carnosine	**0**	**-**	**0**
Anserine	**0**	**-**	**0**

Table S26. Data collection and refinement statistics for the dipeptide epimerase from *E. faecalis*.

	apo	L-Ile-L-Tyr	L-Leu-L-Tyr	L-Ser-L-Tyr	L-Arg-L-Tyr
Data collection					
Space group	R3	C2	R3	R3	R3
No. of mol. in asym. unit	8	8	8	8	8
Cell dimensions					
a, *b*, *c* (Å)	162.3, 162.3, 319.6	194.9, 187.4, 91.9	163.9, 163.9, 318.7	163.7, 163.7, 318.1	163.3, 163.3, 317.6
ß (°)		90.01			
Resolution (Å)	1.7	1.8	2.0	2.0	1.9
No. of unique reflections	341811	299981	205342	214313	245947
*R*merge	0.094	0.085	0.078	0.069	0.081
Completeness (%)	99.1	98.8	95.3	99.9	98.7
Refinement					
Resolution (Å)	25.0-1.7	25.0-1.8	25.0-2.0	25.0-2.0	25.0-1.9
Rcryst	0.223	0.221	0.247	0.245	0.236
*R*free	0.224	0.254	0.281	0.282	0.270
No. atoms					
Protein	20920	20920	20920	20920	20920
Waters	888	1625	423	485	776
Mg^{2+} ions	8	8	8	8	8
Bound ligand		L-Ile-L-Tyr	L-Leu-L-Tyr	L-Ser-L-Tyr	L-Arg-L-Tyr
Ligand atoms		168	168	152	192
R.m.s deviations					
Bond lengths (Å)	0.005	0.005	0.006	0.006	0.006
Bond angles (°)	1.2	1.1	1.1	1.2	1.2
PDB entry	3JVA	3JW7	3JZU	3K1G	3KUM

Table S27. Data collection and refinement statistics for the dipeptide epimerase from *B. thetaiotamicron*.

	Mg^{2+}•L-Ala-D-Glu nonproductive binding	Mg^{2+}•L-Pro-D-Glu nonproductive binding	Mg^{2+}•L-Ala-D-Glu productive binding
Data collection			
Space group	C2	C2	C2
No. of mol. in asym. unit	2	2	2
Cell dimensions			
a, *b*, *c* (Å)	141.43, 100.08, 60.00	141.90, 100.07, 60.08	141.71, 99.86, 60.85
β (°)	90.25	90.34	90.32
Resolution (Å)	1.6	1.5	2.0
No. of unique reflections	108564	129985	54965
*R*merge	0.069	0.086	0.078
$I / \sigma I$	23.4	17.7	18.9
Completeness (%)	98.8	97.1	96.1
Refinement			
Resolution (Å)	25.0-1.6	25.0-1.5	25.0-2.0
Rcryst	0.189	0.192	0.247
*R*free	0.218	0.205	0.287
R.m.s. deviations			
Bond lengths (Å)	0.005	0.005	0.006
Bond angles (°)	1.3	1.3	1.3
No. atoms			
Protein	5224	5224	5240
Waters	903	648	188
Inhibitor	30	34	30
Mg2+ ions	2	2	2
PDB entry	3IJI	3IJL	3IJQ

Table S28. Data collection and refinement statistics for the dipeptide epimerase from *C. hutchinsonii.*

	Mg^{2+}•D-Ala-L-Ala	Mg^{2+}•D-Ala-L-Val
Data collection		
Space group	P21	P21
No. of mol. in asym. unit	9	9
Cell dimensions		
a, *b*, *c* (Å)	94.56, 158.09, 182.18	94.00, 158.20, 182.74
α, β, γ (°)	90.00, 100.04, 90.00	90.00, 100.28, 90.00
Resolution (Å)	20-3.0 (3.11-3.0)	20-3.0 (3.11-3.0)
No. of unique reflections	105392	104809
*R*merge	0.138 (0.716)	0.106 (0.553)
$I / \sigma I$	11.88 (2.4)	12.53 (2.56)
Completeness (%)	99.8 (100.0)	99.6 (100.0)
Refinement		
Resolution (Å)	20.00-3.0	20.0-3.0
Rcryst	0.230	0.227
*R*free	0.254	0.251
R.m.s. deviations		
Bond lengths (Å)	0.009	0.009
Bond angles (º)	1.009	1.033
No. atoms		
Protein	25434	25434
Waters	131	103
Substrate	99	117
Mg^{2+} ions	9	9
B-factors		
Wilson plot	75.1	80.6
Protein	66.4	70.4
Ligands	73.1	75.4
Solvent	46.3	47.6
PDB entry	3Q4D	3Q45

Table S29. Data collection and refinement statistics for the dipeptide epimerase from *M. capsulatus.*

	Mg^{2+}•L-Arg-D-Lys	Mg^{2+}
Data collection		
Space group	$I2_13$	$P2_1$
No. of mol. in asym. unit	5	6
Cell dimensions		
a, *b*, *c* (Å)	275.32, 274.32, 275.32	79.19 150.67 106.02
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 104.66, 90.00
Resolution (Å)	19.9-2.7 (2.80-2.70)	19.9-2.2 (2.24-2.20)
No. of unique reflections	94445	121430
R_{merge}	0.154 (0.768)	0.144 (0.541)
$I / \sigma I$	12.85 (2.03)	13.03 (3.45)
Completeness (%)	99.8 (100.0)	99.8 (100.0)
Refinement		
Resolution (Å)	19.9-2.7	19.9-2.2
R_{cryst}	0.175	0.174
R_{free}	0.222	0.229
R.m.s. deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	1.339	1.264
No. atoms		
Protein	13605	15936
Waters	555	534
Substrate	105	-
Mg2+ ions	5	6
B-factors		
Wilson plot	52.4	25.1
Protein	42.2	28.8
Ligands	63.3	24.5
Solvent	37.5	28.6
PDB entry	3RIT	3RO6

Table S30. Data collection and refinement statistics for the dipeptide epimerase from *F. philomiragia*

DATASET[a]	Tartrate, No Mg^{2+}	Mg^{2+}	L-Ala, D/L-Glu,
Space Group	$P4_12_12$	$P4_12_12$	P41212
Unit Cell (Å , °)	*a*=*b*=121.0 *c*=148.8	*a*=*b*=120.7 *c*=150.0	a=b=121.2, c=149.0
Resolution (Å)	40-2.0 (2.11-2.0)	40-2.0 (2.11-2.0)	30-1.9 (2.0-1.9)
Completeness (%)	100 (100)	100 (100)	97.9 (92.5)
Redundancy	28.8 (27.9)	27.8 (24.9)	12.5 (12.1)
Mean(I)/sd(I)	27.6 (11.9)	22.2 (8.1)	16.4 (5.8)
Rsym	0.093 (0.312)	0.122 (0.430)	0.097 (0.380)
Wilson B Factor (Å2)	19.4	18.1	18.3
REFINEMENT			
Resolution (Å)	40-2.0 (2.02-2.0)	40-2.0 (2.07-2.0)	40-1.9 (1.97-1.9)
Unique reflections	74276 (2706)	74571 (7301)	84359 (7647)
Rcryst (%)	16.5 (16.1)	16.0 (16.7)	15.7 (18.2)
Rfree (%, 5% of data)	19.7 (19.8)	19.1 (21.2)	18.5 (21.6)
Contents of model			
Residues (1-356)[b]	A2-22, A30-A363, B2-B365	A2-363, B2-365	A2-363, B2-365
Waters	559	578	708
Mg^{2+}	0	1	0
Other	59	77	83
Atoms total	6172	6260	6404
Average B-factor ($Å^2$)			
Protein/Waters/Ligand	27.3 / 35.4 / 55.3	19.0 / 28.9 / -	23.2 / 34.0 / 23.0

RMSD			
Bond lengths (Å) /Angles (°)	0.008 / 1.18	0.008 / 1.20	0.008 / 1.24
MOLPROBITY STATISTICS			
Ramachandran Favored / Outliers (%)	98.3 (0.0)	98.3 (0.0)	98.2 (0.0)
Rotamer Outliers (%)	0.50	0.82	0.65
Clashscore[c]	9.29 (85th pctl.)	7.30 (93rd pctl.)	7.3 (91st pctl.)
Overall score[c]	1.49 (97th pctl.)	1.40 (98th pctl.)	1.4 (97th pctl.)
PDB ID	3r0k	3r10	3r1z

[a] Statistics in parenthesis are for the highest resolution bin

[b] Additional residues past 356 are from C-terminal tag

[c] Scores are ranked according to structures of similar resolution as formulated in MOLPROBITY

Table S31. Data collection and refinement statistics for the dipeptide epimerase from *H. aurantiacus*

Data collection	
Space group	P21
Cell dimensions	
a, *b*, *c* (Å)	84.026, 93.344, 85.914
α, β, γ (°)	90.00, 112.60, 90.00
Resolution (Å)	2.07-50.00(2.07-2.14) *
*R*merge	0.063(0.78)
$I / \sigma I$	7.9(0.9)
Completeness (%)	98.9(91.2)
Refinement	
Resolution	2.1
No. reflections	68809
R_{work} / R_{free}	0.217/0.275
No. atoms	
Protein	10290
Ligand/ion	4
Water	249
B-factors	
Protein atoms	42.09
Solvent atoms	34.76
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.25
Ramachandran Plot Statistics	
Most favored regions	90.20%
Additional allowed regions	9.00%
Generously allowed regions	0.70%
PDB entry	3IK4

*Values in parentheses are for highest-resolution shell.

SUPPLEMENTARY METHODS

Detailed Methods: General

Restriction enzymes, Platinum Pfx DNA polymerase, and T4 DNA ligase were purchased from Invitrogen. Some restriction enzymes were purchased from New England Biolabs Inc. Commercially available dipeptides were purchased from Bachem America, Inc., Sigma/Aldrich, Chem-Impex International, or INDOFINE Chemical Company. Reagents for dipeptide synthesis were purchased from EMD Biosciences or Sigma/Aldrich. Oligonucleotide primers were synthesized by Bio-Synthesis Inc. (Lewisville, TX). Crystallization screens were purchased from Hampton Research. All other reagents were purchased from Sigma/Aldrich or Fisher and were of the highest quality grade available.

All ^{1}H NMR spectra were recorded on a Varian Unity INOVA 500NB MHz spectrometer. Electrospray ionization (ESI) mass spectrometry spectra were acquired on a quadrupole-hexapole-quadrupole (QHQ) mass spectrometer (Quattro II) from Waters using MassLynx software.

Detailed Methods: Solution phase synthesis of dipeptide substrates

Dipeptide Libraries. Syntheses of dipeptide libraries were performed by the solid phase peptide synthesis methods described by Song *et al.* (1)

Synthetic Dipeptides. D-Ala-L-Val, D-Ala-L-Ile, D-Ala-L-Met, D-Ala-L-Leu, L-Ala-D-Glu, L-Val-L-Glu, and L-Val-D-Glu were synthesized by solution phase methods. The synthetic procedures and characterization are described below.

Synthesis of D-Ala-L-Val

A typical reaction mixture for the synthesis of D-Ala-L-Val was set up in 14 mL of $CHCl_3$ and contained 2 mmoles of Boc-D-Ala-OH and H-Val-OMe. For coupling, 4 mmoles of 1-hydroxybeonzotriazole (HOBt) and N,N'-diisopropylcarbodiimide were added and placed on a shaker for overnight incubation at room temperature (RT). The mixture was filtered and rotary evaporated to a syrup and washed 3 times with chilled acetone; the white precipitate was filtered and discarded. Remaining pale yellow solution was applied to a silica column (2 cm x 30 cm) and eluted with 125 mL of acetone. Fractions were spotted on a TLC and visualized for the presence of Boc-D-Ala-L-Val-OMe with ninhydrin spray. Fractions containing the protected dipeptide were rotary evaporated to a syrupy substance. The protective groups were removed in two steps: incubation with 1:1 mix of 2 N NaOH and CH_3OH for 2 hours at RT, rotary-evaporation to syrup, incubation with 2:1 mix of TFA and CH_2Cl_2 for 2 hours at RT, and rotary-evaporated to syrup. Five mL of ddH_2O was added to the syrup and washed 3 times with ethyl ether until the organic layer turned colorless. In some cases when further purification was necessary, the dipeptide concentrate was acidified to pH ~2 and applied to a Dowex 50W-X2 cation exchange column (Sigma) in the H^+ form (1.6 cm x 50 cm) and eluted with 1000 mL with a linear gradient of 0-0.5M HCl. The fractions were spotted on a TLC and visualized with ninhydrin spray (Sigma). The purity of the dipeptide was determined by ^{1}H NMR. ^{1}H NMR (D_2O) for D-Ala-L-Val: δ 0.75 [6H, dd, J = 6.9 and 16, -CH(CH_3)$_2$], 1.35 (3H, d, J = 7.1, -CH_3), 2.0 [1H, m, -CH(CH_3)$_2$], 3.9 (1H, q, J = 7.2, 7.0 and 21.3, -CH(CH_3)-], 3.93 {1H, d, J = 5.6, -CH[CH(CH_3)$_2$]-} (Figure A).

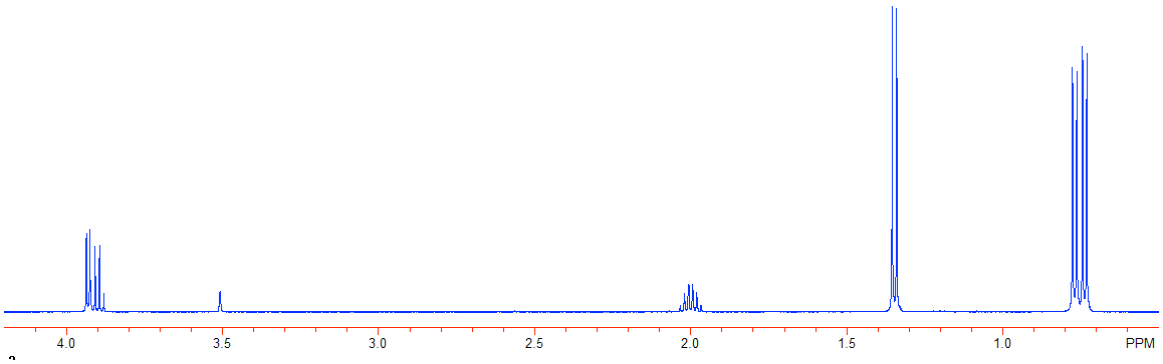


Figure A. *^{1}H NMR spectrum of D-Ala-L-Val at pD ~8.0.*

Synthesis of D-Ala-L-Ile

The synthesis of D-Ala-L-Ile was carried out as described for D-Ala-L-Val. The purity of the dipeptide was determined by ^{1}H NMR. ^{1}H NMR (D_2O) for D-Ala-L-Ile: δ 0.71 (3H, t, *J* = 7.4 and 14.9, -CH_2CH_3), 0.74 [3H, d, *J* = 6.9, -$CH(CH_3)CH_2CH_3$], 1.00 and 1.23 (2H, m, -CH_2CH_3), 1.32 (3H, d, *J* = 7.1, -CH_3), 1.72 [1H, m, -$CH(CH_3)CH_2CH_3$], 3.88 [1H, q, *J* = 7, 7.1 and 21.2, -$CH(CH_3)$-], 3.94 {1H, d, *J* = 5.9, -*CH* [$CH(CH_3)CH_2CH_3$]-} (Figure B).

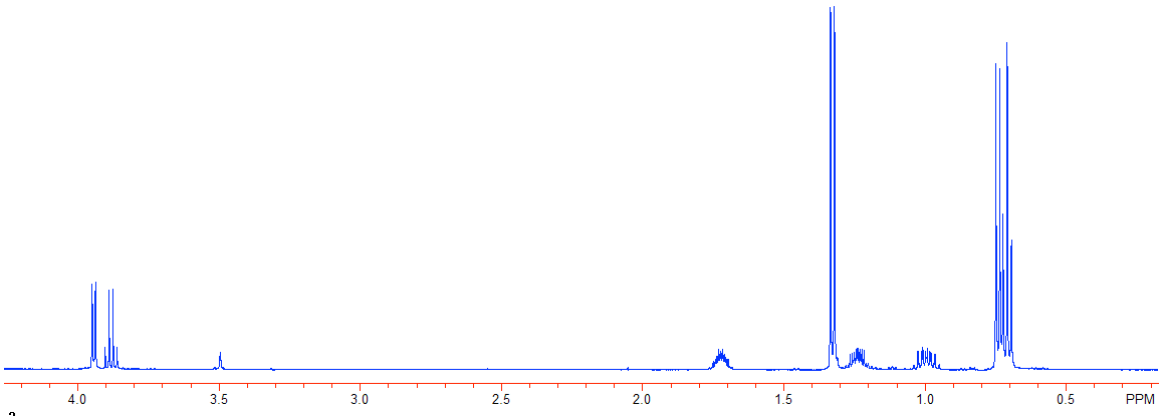


Figure B. *^{1}H NMR spectrum of D-Ala-L-Ile at pD ~8.0.*

Synthesis of D-Ala-L-Met

The synthesis of D-Ala-L-Met was carried out as described above. The purity of the dipeptide was determined by ^{1}H NMR. ^{1}H NMR (D_2O) for D-Ala-L-Met: δ 1.32 (3H, d, *J* = 7.1, -CH_3), 1.80 and 1.96 (2H, m, -C*H*$_2$$CH_2SCH_3$), 1.94 (3H, s, -$SCH_3$), 2.35 and 2.43 (2H, m, -CH_2C*H*$_2$$SCH_3$), 3.85 [1H, q, *J* = 7.1, 7.1 and 21.2, -C*H*(CH_3)-], 4.14 [1H, q, *J* = 4.4, 5 and 13.8, -C*H*($CH_2CH_2SCH_3$)-] (Figure C).

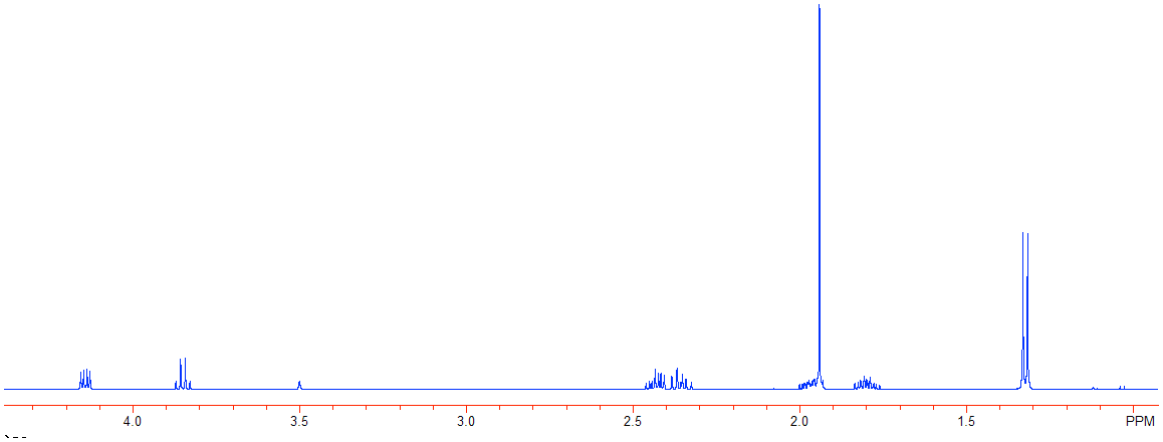


Figure C. *^{1}H NMR spectrum of D-Ala-L-Met at pD ~8.0.*

Synthesis of D-Ala-L-Leu

Synthesis of D-Ala-L-Leu was carried out as described earlier. The purity of the dipeptide was determined by ^{1}H NMR. ^{1}H NMR (D_2O) for D-Ala-L-Leu: δ 0.75 [6H, dd, *J* = 6.3 and 20.2, -CH(*CH*$_3$)$_2$], 1.29 (3H, d, *J* = 7.1, -*CH*$_3$), 1.46 [3H, m, -*CH*$_2$*CH*(CH_3)$_2$], 3.78 [1H, q, *J* = 7.1, 7.1 and 21.2, -*CH*(CH_3)-], 4.03 {1H, t, *J* = 5.7, 8.7 and 14.5, -*CH*[$CH_2CH(CH_3)_2$]-} (Figure D).

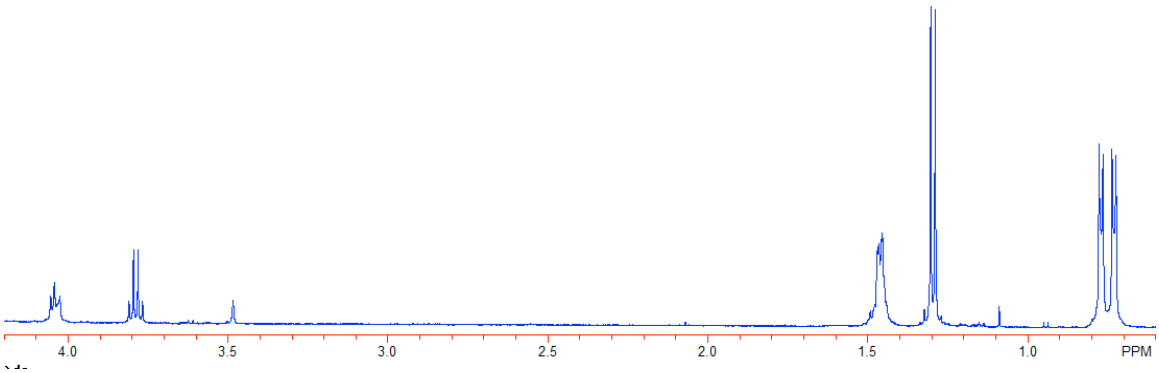


Figure D. *^{1}H NMR spectrum of D-Ala-L-Leu at pD ~7.0.*

Synthesis of L-Ala-D-Glu

The synthesis of L-Ala-D-Glu was carried out as described above, with the exception of the last purification step. The pH of the dipeptide was adjusted to ~12 with NH_4OH and then applied to a Dowex AG1-X8 anion exchange column in the acetate form. The column was subsequently washed with ddH_2O. The dipeptide substrate was typically eluted with a 1000 mL with a gradient of 0-1M acetic acid. The same purification procedure was used for all glutamate containing dipeptides.. The purity of the dipeptide was determined by ^{1}H NMR. ^{1}H NMR (D_2O) for D-Ala-L-Glu: δ 1.33 (3H, d, J = 7, -CH_3), 1.72 and 1.92 (2H, m, -$CH_2CH_2CO_2H$), 2.06 (2H, t, J = 8.1 and 16.1, -$CH_2CH_2CO_2H$), 3.84 [1H, q, J = 7.1, 7.1 and 21.3, -$CH(CH_3)$-], 3.96 [1H, q, J = 4.2, 5.1 and 13.7, -$CH(CH_2CH_2CO_2H)$-] (Figure E).

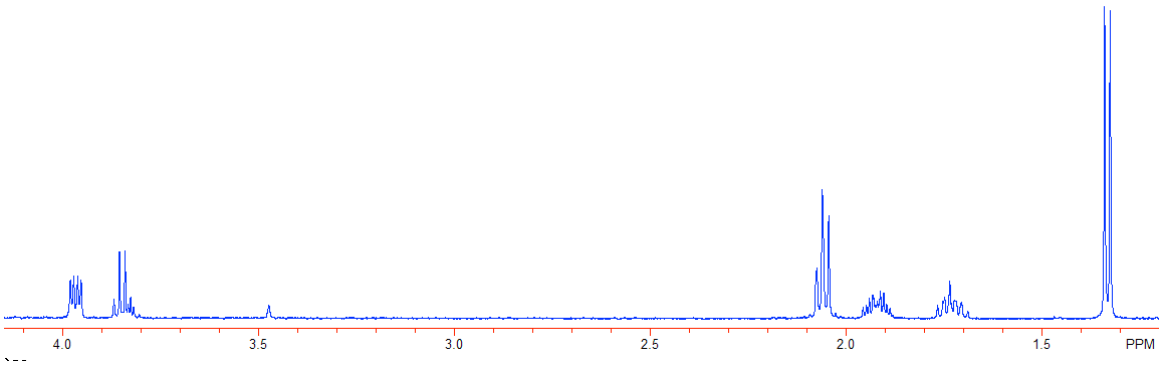


Figure E. *^{1}H NMR of L-Ala-D-Glu at pD ~8.0.*

Synthesis of L-Val-L-Glu

The synthesis of L-Val-L-Glu was carried out as described for L-Ala-D-Glu. The purity of the dipeptide was determined by ^{1}H NMR. ^{1}H NMR (D_2O) for L-Val-L-Glu: δ 0.79 [6H, dd, *J* = 6.9 and 18.5, -CH(C*H*$_3$)$_2$], 1.83 [3H, m, -*CH*(CH$_3$)$_2$, -*CH*$_2$CH$_2$CO$_2$H], 2.08 [2H, t, *J* = 8.4 and 17, -CH$_2$*CH*$_2$CO$_2$H] , 3.08 {1H, d, *J* = 6.1, -*CH*[CH(CH$_3$)$_2$]-}, 4.2 [1H, q, *J* = 4.6, 4 and 13.4, -*CH*(CH$_2$CH$_2$CO$_2$H)-] (Figure F).

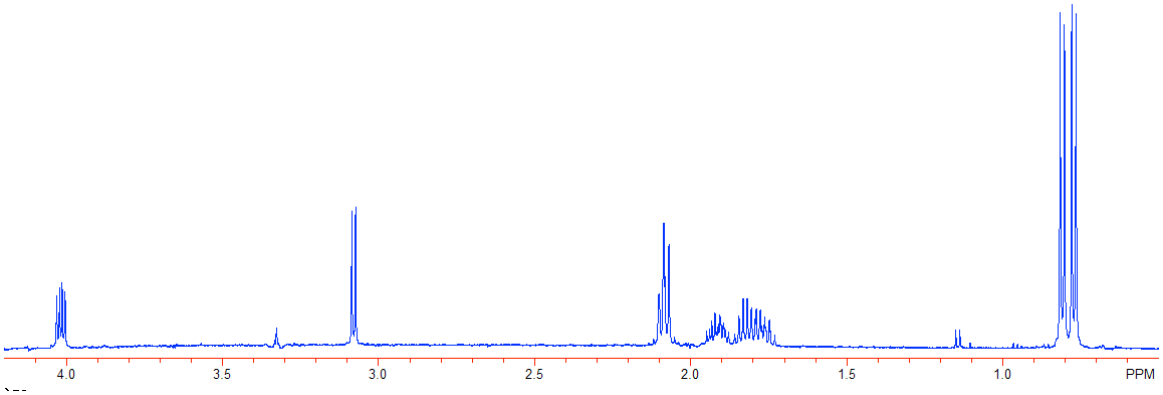


Figure F. *^{1}H NMR spectrum of L-Val-L-Glu at pD ~8.0.*

Synthesis of L-Val-D-Glu

The synthesis of L-Val-D-Glu was carried out as described for L-Ala-D-Glu. The purity of the dipeptide was determined by ^{1}H NMR. ^{1}H NMR (D_2O) for D-Val-D-Glu: δ 0.85 [6H, dd, *J* = 2 and 6.8, -CH(CH_3)$_2$], 1.74 and 1.94 [3H, m, -*CH*(CH_3)$_2$, -*CH*$_2$$CH_2CO_2H$], 2.08 [2H, t, *J* = 8 and 17.1, -CH_2*CH*$_2$$CO_2H$] , 3.47 {1H, d, *J* = 6.6, -*CH*[CH(CH_3)$_2$]-}, 3.98 [1H, q, *J* = 4.3, 5.1 and 13.7, -*CH*($CH_2CH_2CO_2H$)-] (Figure G).

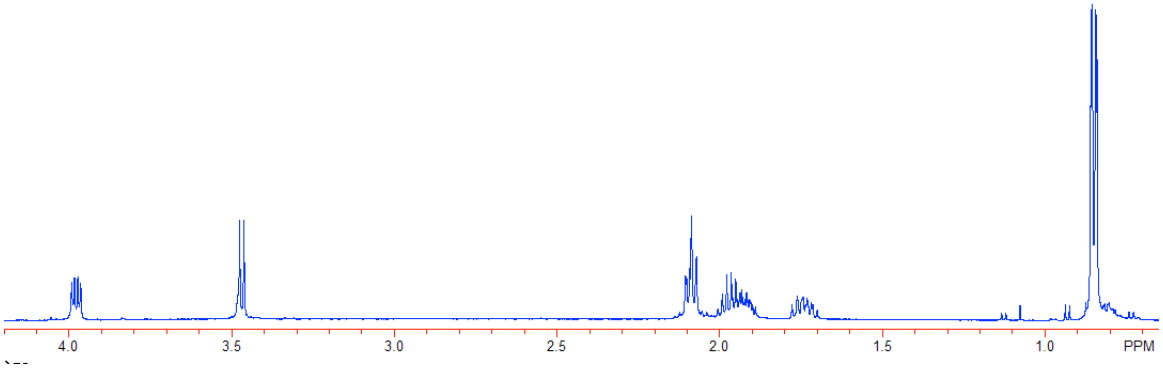


Figure G. *^{1}H NMR spectrum of L-Val-D-Glu at pD ~7.0.*

L-Arg-L-Lys synthesis

C-terminal amino acid with a protecting group t-butyl ester (OtBu), 1 mmol of H-Lys(Boc)-OtBu·HCl, was dissolved in 10 mL of dry DMF solvent. Then, 270 mg of HOBT (2 molar equivalent), and 722 uL of DIC (4 molar equivalent) were added to the solution, followed by addition of N-terminal amino acid protected by Boc, 1 mmol of Boc-Arg(Boc)$_2$-OH. After combining all, the coupling reaction proceeded with shaking at room temperature for 16 hr and the dipeptide production was monitored by TLC. The reaction mixture was filtered to remove the precipitated residues and concentrated to a syrup using a high-vac rotary evaporator over a 30 °C bath. The syrup was washed with acetone solution three times. The crude dipeptide product was purified by silica gel chromatography to remove the unreacted starting materials. The purified product from silica gel chromatography was concentrated to a crude dipeptide product

with protecting groups which was dissolved in 9 mL of TFA/CH_2Cl_2 (1:2 mixture) to remove protecting groups, OtBu and Boc. The reaction proceeded with stirring at room temperature for 2 hr. Then the solvent was removed using a high-vac rotary evaporator over a 40 °C bath. The product was dissolved in 10 mL ddH_2O and impurities were removed by 10 mL ethyl ether extraction which was repeated three times. The product was then purified over Dowex AG-50 cation exchange resin, eluting with 0 – 3 N HCl. Fractions containing pure dipeptide were identified by TLC, combined, and concentrated using a high-vac rotary evaporator over a 40 °C bath. 5 mL ddH_2O was added to the concentrated dipeptide product, and the pH was adjusted to 6 – 8. The water was removed to obtain the final dried product by lyophilization. The product yield was 114 mg (28% yield) and characterized by ^{1}H NMR (500 MHz, D_2O): 4.06 ppm (1H, m, $HO_2CC\underline{H}NH$), 3.94 ppm (1H, m, $NH_2C\underline{H}COO$), 3.12 ppm (2H, t, J = 7.5 Hz, $CH_2CH_2C\underline{H}_2NH$), 2.89 ppm (2H, t, J = 7.9 Hz, $CH_2CH_2CH_2C\underline{H}_2NH_2$), 1.86 – 1.30 ppm (4H, m, $C\underline{H}_2C\underline{H}_2CH_2NH$), and 1.86 – 1.30 ppm (6H, m, $C\underline{H}_2C\underline{H}_2C\underline{H}_2CH_2NH_2$).

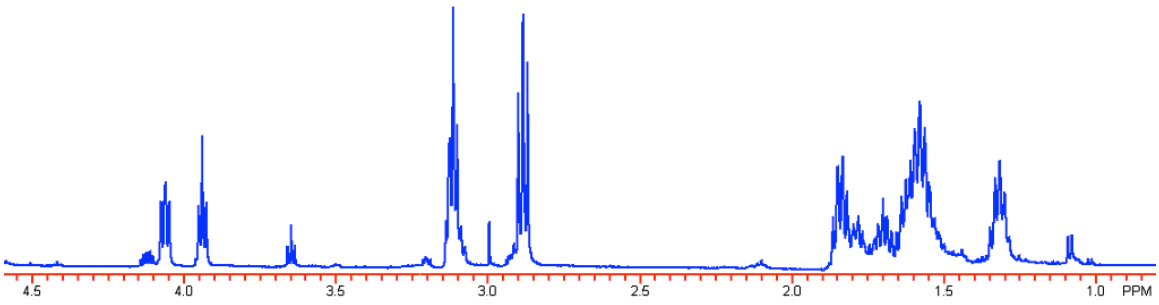


Figure H. ^{1}H NMR spectrum of L-Arg-L-Lys.

L-Arg-L-Orn synthesis

L-Arg-L-Orn substrate synthesis was performed following the protocol described in the previous section except the next two points. First, protected C-terminal amino acid with methyl ester (OMe), 1 mmol of H-Orn(Boc)-OMe·HCl, was added to the reaction instead of H-

Lys(Boc)-OtBu·HCl. Second, an extra deprotection reaction was performed to remove OMe before the deprotection reaction by TFA/CH_2Cl_2. After silica gel chromatography separation, 10 mL of MeOH/NaOH (2 N) (1:1 mixture) was added to the crude product and stirred at room temperature for 2 hr. The product yield was 119 mg (41% yield) and characterized by ^{1}H NMR (500 MHz, D_2O): 4.09 ppm (1H, m, $HO_2CC\underline{H}NH$), 3.94 ppm (1H, m, $NH_2C\underline{H}COO$), 3.11 ppm (2H, t, J = 7.4 Hz, $CH_2CH_2C\underline{H}_2NH$), 2.91 ppm (2H, t, J = 7.2 Hz, $CH_2CH_2C\underline{H}_2NH_2$), 1.86 – 1.51 ppm (4H, m, $C\underline{H}_2C\underline{H}_2CH_2NH$), and 1.86 – 1.51 ppm (4H, m, $C\underline{H}_2C\underline{H}_2CH_2NH_2$).

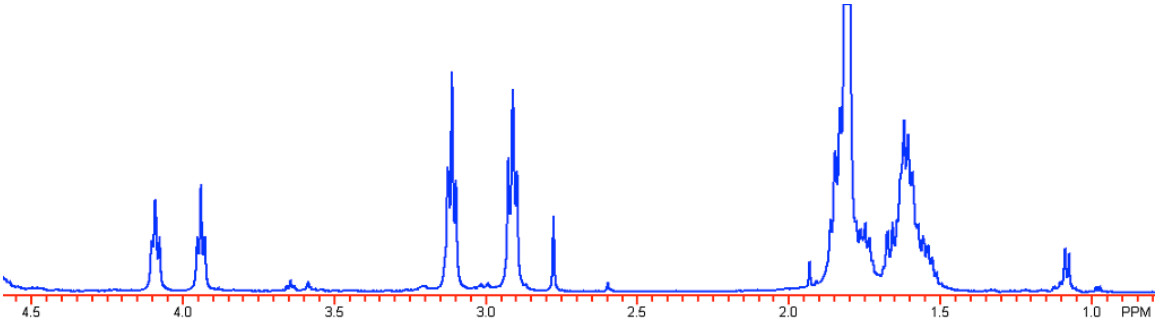


Figure I. ^{1}H NMR spectrum of L-Arg-L-Orn.

L-Arg-L-Ser synthesis

L-Arg-L-Ser substrate synthesis was performed following the protocol described in the section of L-Arg-L-Lys synthesis except that protected C-terminal amino acid, 0.5 mmol of H-Ser(tBu)-OtBu·HCl with *tert*-butyl (tBu) protecting group, was added to the reaction instead of H-Lys(Boc)-OtBu·HCl and that the amount of chemical used was reduced to half. The product yield was 111 mg (85% yield) and characterized by ^{1}H NMR (500 MHz, D_2O): 4.22 ppm (1H, m, $HO_2CC\underline{H}NH$), 3.97 ppm (1H, m, $NH_2C\underline{H}COO$), 3.76 ppm (2H, m, $CHC\underline{H}_2OH$), 3.13 ppm (2H, t, J = 6.8 Hz, $CH_2CH_2C\underline{H}_2NH$), 1.86 ppm (2H, m, $C\underline{H}_2CH_2CH_2NH_2$), 1.60 ppm (2H, m, $CH_2C\underline{H}_2CH_2NH_2$).

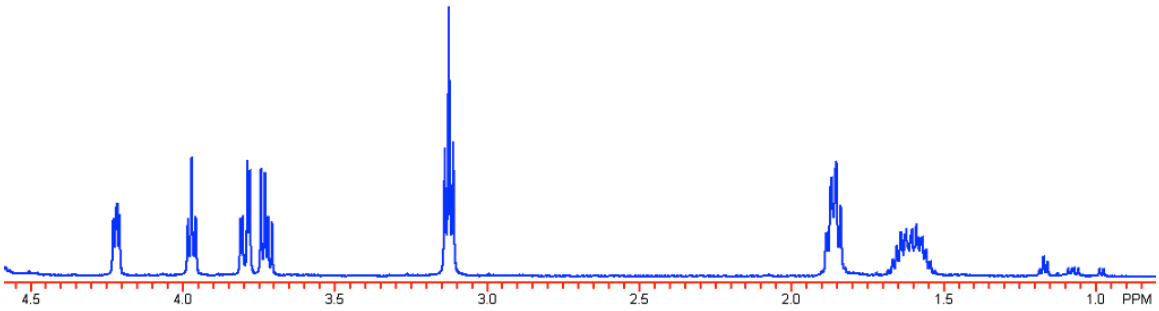


Figure J. ^{1}H NMR spectrum of L-Arg-L-Ser.

L-Orn-L-Arg synthesis

L-Orn-L-Arg substrate synthesis was performed following the protocol described in the section of L-Arg-L-Lys synthesis except the two starting materials; Boc-Orn(Boc)-OH for the N-terminal amino acid and H-Arg-(OtBu)·2HCl for the C-terminal amino acid. Two deprotection steps, first with MeOH/NaOH and second with TFA/CH_2Cl_2, were performed during purification. The product yield was 151 mg (38% yield) and characterized by ^{1}H NMR (500 MHz, D_2O): 4.08 ppm (1H, m, $HO_2CC\underline{H}NH$), 3.95 ppm (1H, m, $NH_2C\underline{H}COO$), 3.13 ppm (2H, t, J = 6.9 Hz, $CH_2CH_2C\underline{H}_2NH$), 2.93 ppm (2H, t, J = 7.4 Hz, $CH_2CH_2C\underline{H}_2NH_2$), 1.87 – 1.51 ppm (4H, m, $C\underline{H}_2C\underline{H}_2CH_2NH_2$), and 1.87 – 1.51 ppm (4H, m, $C\underline{H}_2C\underline{H}_2CH_2NH$).

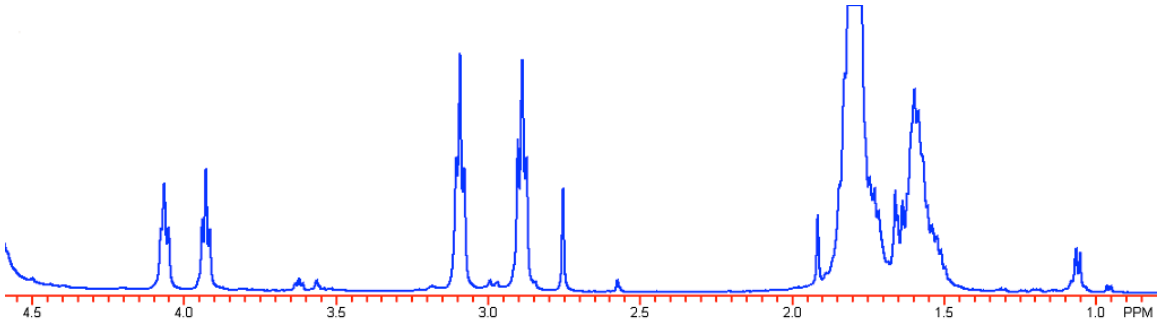


Figure K. ^{1}H NMR spectrum of L-Orn-L-Arg.

L-Val-L-Arg synthesis

L-Val-L-Arg substrate synthesis was performed following the protocol described in the section of L-Arg-L-Lys synthesis except the two starting materials; Boc-Val-OH for the N-terminal acid and H-Arg(OtBu)·2HCl for the C-terminal amino acid. The product yield was 24.0 mg (8.8% yeild) and characterized by ^{1}H NMR (500MHz, $^{2}H_2O$); 4.22 ppm (1H, m, $HO_2CC\underline{H}NH$), 3.48 ppm (1H, m, $NH_2C\underline{H}COO$), 3.01 ppm (2H, m, $CH_2CH_2C\underline{H}_2NH$), 2.15 ppm (1H, m, $C\underline{H}CH_3CH_3$), 1.18 ppm (4H, m, $C\underline{H}_2C\underline{H}_2CH_2NH_2$), and 0.918 ppm (6H, m, $CHC\underline{H}_3C\underline{H}_3$).

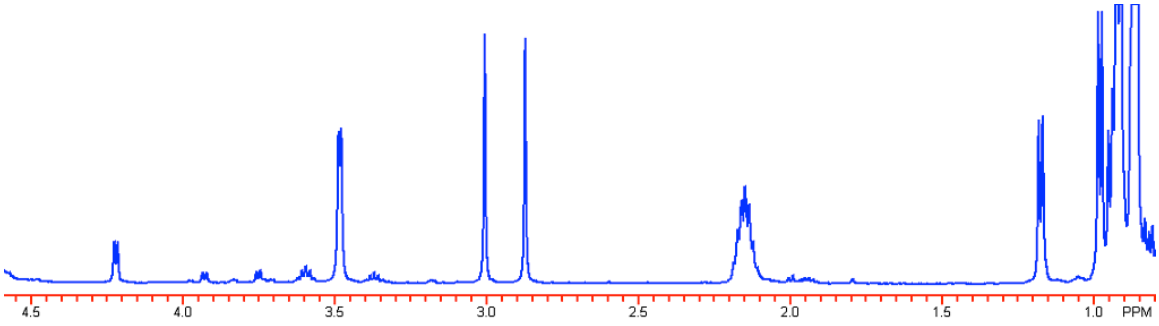


Figure L. ^{1}H NMR spectrum of L-Val-L-Arg.

Detailed Methods: Cloning, Expression and Purification

Cloning, Expression, and Purification of the Dipeptide Epimerase from Bacteroides thetaiotaomicron VPI-5482 ATCC 29148 (GI: 29346723). Alignment of the sequence of BT1313 with those of other dipeptide epimerases suggested that the N-terminal Met for BT1313 is Met 46 in the GenBank sequence. Therefore, the gene BT1313 was PCR-amplified from genomic DNA using primers that amplified the coding sequence for residues 46 383 in the GenBank sequence: 5'_ CATATTAATAAATATGGTATA-GGGGGACATATGAAAATGACTTTTTTCCC-3' and 5' cattctcgaggatcagattttcatgacaccaatccgggtaagtc-3', containing 5' *Nde*I and 3' *Xho*I restriction sites

(underlined), respectively. The PCR reaction mixture (100 µL) contained 0.5 µL of 100 ng/mL template DNA, 2 µL of 50 mM $MgSO_4$, 0.5 µL of 2.5 units/µL platinum *Pfx* DNA polymerase, 10 µL of 10× *Pfx* amplification buffer, 10 µL of 10× enhancer buffer, 2 µL of 20 mM dNTP mix, 2 µL of each 20 µM forward and reverse primer, and 71 µL of ddH_2O. The PCR was performed with the following program: 94 °C for 3 min followed by 40 cycles of 94 °C for 1 min, 65 °C for 1.25 min, 68 °C for 3 min, and a final extension at 68 °C for 10 min. The product was purified using a Qiagen gel extraction kit (Qiagen, Inc). The product was then digested with *Nde*I and *Xho*I for 2 hr at 37 °C and gel purified. The digested product was ligated overnight with T4 DNA ligase into the pET 17b expression vector (Novagen) that had been digested with *Nde*I and *Xho*I. The sequence of the gene was verified at the University of Illinois Core Sequencing Facility.

The gene of interest was expressed using the pTara expression system in *E. coli* MG1655 in which the chromosomal gene encoding L-Ala-D/L-Glu epimerase had been disrupted. For a typical protein purification batch, the cells were grown in 4 L of LB medium in the presence of 34 µg/mL of chloramphenicol and 100 µg/mL of ampicillin at 25°C for 3 days and harvested by centrifugation. The cells were suspended in 20 mM Tris-Cl, pH 8.0, containing 5 mM $MgCl_2$ and lysed on ice using Fisher Scientific 550 Sonic Dismembrator. The insoluble cell debris was removed by centrifugation. The supernatant was applied to a DEAE-Sepharose FF column (2.5 cm × 50 cm, GE Healthcare) and eluted with a linear gradient (1600 mL) of 0 to 1 M NaCl buffered with 20 mM TrisCl, pH 8.0, and 5 mM $MgCl_2$. Fractions containing the protein were applied to Phenyl Sepharose 6FF (low sub) column (1.6 x 20 cm, GE Healthcare) in 0.5 M $(NH_4)_2SO_4$ and 20 mM Tris-Cl, pH 7.9, for further purification. The column was washed with 100 mL of 0.5 M $(NH_4)_2SO_4$ in 20 mM Tris-Cl, pH 7.9 and eluted with a linear gradient of 0.5 to

0 M $(NH_4)_2SO_4$ (600 mL) containing 20 mM Tris -Cl, pH 7.9, and 5 mM $MgCl_2$. The protein was then dialyzed against 50 mM Tris-Cl, pH 7.9, containing 100 mM NaCl and 5 mM $MgCl_2$ for storage. The protein was subjected to ESI-MS to confirm the correct molecular weight (expected mass of 37,493.4 Da, observed mass of 37,495.0 Da). The protein was determined to be a monmer by HPLC size exclusion chromatography.

Cloning, Expression, and Purification of the Dipeptide Epimerase from Cytophaga hutchinsonnii ATCC 33406 (GI:110638536). The gene encoding the putative dipeptide epimerase was PCR amplfied from genomic DNA using the primers 5'-gtctttttttcgtataggtgtattaatatgattataacacagg-3' and 5'-cacaaactcgaggcgctcagtttat-gcatatttttttc-3', containing 5' *Ase*I and 3' *Xho*I restriction sites (underlined), respectively. The procedure described in the previous section was used for PCR amplificiation, purification, digestion (with *Ase*I and XhoI), and ligation of the PCR product into the pET-17b expression vector.

For a typical protein purification, the cells in BL21 were grown in 4L of LB medium at 25 °C for 3 days in the presence of 100 μg/μL of ampicillin. The cells were harvested by centrifugation and resuspended in 60 mL of 20 mM Na^+-HEPES, pH 8.35, containing 5 mM $MgCl_2$. The suspension was then lysed on ice using a Fisher Scientific 550 Sonic Dismembrator; the cell lysate was centrifuged to remove insoluble debris. The supernatant was applied to a DEAE-Sepharose FF column and eluted with a linear gradient (1600 mL) of 0 to 0.5 M NaCl buffered with 20 mM Na^+-HEPES, pH 8.35, containing 5 mM $MgCl_2$. For crystallization purposes, further purification of the protein was halted at this point. Fractions containing the protein, as determined by SDS-PAGE were pooled and glycerol was added to a final concentration of 10%. For enzymatic assays, fractions containing the protein were applied to

Phenyl Sepharose 6FF column (1.6 x 20 cm, GE Healthcare) in 0.5 M $(NH_4)_2SO_4$ and 20 mM Na^+-HEPES, pH 7.9, for further purification. The column was washed with 100 mL of 0.5 M $(NH_4)_2SO_4$ in 20 mM Na^+-HEPES, pH 7.9 and eluted with a linear gradient of 0.5 to 0 M $(NH_4)_2SO_4$ (600 mL) containing 20 mM Na^+-HEPES, pH 7.9, and 5 mM $MgCl_2$. The protein was dialyzed against 50 mM Na^+-HEPES, pH 7.9, containing 100 mM NaCl and 5 mM $MgCl_2$ for storage. The protein was subjected to ESI-MS to confirm the correct molecular weight (expected mass of 40,310.6, observed mass of 40,313.0). The protein was determined to be a monomer by HPLC size exclusion chromatography.

Cloning, Expression and Purification of the Dipeptide Epimerase from Enterococcus faecalis (GI:29376078). The gene encoding the putative dipeptide epimerase was cloned from *E. faecalisi* genomic DNA using *Taq* DNA polymerase and primers that specified an *Nde* I recognition site at the 5' end end and a *Bam* HI recognition site at the 3' end of the gene. Following purification by gel electrophoresis, the amplified PCR product was restricted using *Nde* I and *Bam* HI restriction enzymes per the manufacturer's protocols. The gene was then ligated into pET17b using T4 DNA ligase and transformed in *E. coli* XL1Blue cells for plasmid amplification and maintenance.

For large scale purification, 2L of LB containing 100 μg/mL ampicillin was inoculated with 5 mL of a starter culture of BL21 cells transformed with the pET17b derived plasmid. The culture was grown at 37°C without induction by IPTG for 24 hours and then harvested by centrifugation. The cell pellet was resuspended in 60 mL of cold low salt buffer (10 mM Tris - Cl, pH 7.9, containing 5 mM $MgCl_2$) and lysed by sonication. The lysate was cleared by centrifugation, and the supernatant was loaded onto a DEAE Sepharose Fast Flow column. The

protein was eluted with a linear gradient (1600 mL) of 0 to 1 M NaCl buffered with 10 mM Tris HCl, pH 7.9, containing 5 mM $MgCl_2$. The fractions containing the purest protein as assessed by SDS PAGE were pooled. The protein containing solution was dialyzed three times against low salt buffer at 4°C and then loaded onto Q Sepharose HP column (1.7 x 7 cm, GE Healthcare). The protein was eluted with a linear gradient (250 mL) of 0 to 1 M NaCl in 10 mM Tris-HCl, pH 7.9, containing 5 mM $MgCl_2$. Fractions containing pure protein were dialyzed three times at 4°C against 20 mM Tris-HCl, pH 7.9, containing 100 mM NaCl and 5 mM $MgCl_2$. The solution was concentrated to ~10 mg/mL for assays and ~45 mg/mL for crystallization.

Cloning, Expression, and Purification of the Dipeptide Epimerase from Methylococcus capsulatus str. Bath (GI: 53757661). The gene encoding the dipeptide epimerase (GI: 53757661) was amplified by PCR from *Methylococcus capsulatus str. Bath* genomic DNA (kindly provided by Prof Live J. Bruseth at University of Bergen, Norway) using the following primers: 5' GGAACGGGAGGAAGTCATATGAAGATCGCCGACATC 3' and 5' GGAATTGCATCGGGGATCCTCTAATCCGGGTATACG 3' containing a 5' *Nde*I site and a 3' *Bam*HI site (underlined), respectively. PCR reactions in 50 µL total volume contained 100 ng template, 1 mM $MgSO_4$, 2.5 U of platinum *Pfx* DNA polymerase, 2X *Pfx* amplification buffer, 2X enhancer buffer, 0.4 mM of each dNTP, and 0.4 µM of each forward and reverse primer. The PCR reaction was performed with the following parameters: 94° C for 5 min followed by 40 cycles of 94° C for 1 min, 60° C for 1.25 min, and 68° C for 3 min; the final extension time was 10 minutes at 68° C. Following purification by gel extraction, the amplified PCR product was restricted using *Nde*I and *Bam*HI restriction enzymes per the manufacturer's protocols. The gene was then ligated into the non-tagged expression vector pET17b using T4 DNA ligase and transformed in *E. coli* XL1Blue cells for plasmid maintenance.

The cloned dipeptide epimerase from *M. capsulatus* was expressed in *E. coli* BL21 (DE3) cells for protein purification. In a typical protein preparation, 2 L of LB media was shaken at 37°C without induction and harvested after 48 hours by centrifugation. The cells were resuspended in 80 mL of buffer containing 20 mM Tris-HCl, pH 8.0, and 5 mM $MgCl_2$. The suspension was lysed by sonication and debris was cleared by centrifugation. The supernatant was applied to a DEAE Sepharose FF column and eluted with a linear gradient (1600 mL) of 0 to 1 M NaCl buffered with 20 mM Tris HCl, pH 8.0, containing 5 mM $MgCl_2$. Fractions containing the protein of interest were pooled and applied to a Phenyl Sepharose 6 Fast Flow column in 0.6 M $(NH_4)_2SO_4$ for further purification. The protein was eluted with a linear gradient (800 mL) of 0.6 to 0 M $(NH_4)_2SO_4$ and the purest fractions were pooled and dialyzed against 20 mM Tris-HCl, pH 8.0 and 5 mM $MgCl_2$. For crystallization purposes, the protein was subjected to an additional size exclusion chromatography step to enhance homogeneity. Five mL of the enzyme was applied onto a HiLoad 16/60 Superdex 75 pg column (16 mm X 60 cm, GE Healthcare) and eluted with 50 mM K^+-HEPES, pH 8.0, containing 5 mM $MgCl_2$, and 100 mM KCl. The protein was concentrated to 10 - 15 mg/mL using a Millipore Amicon apparatus fitted with a 10,000 NMWL ultrafiltration membrane and stored at 4°C.

Cloning, Expression, and Purification of the Dipeptide Epimerase from Francisella philomiragia ATCC 25017 (GI:167627873, NYSGRC-200555).

The gene for EFI target 200555 was produced using codon optimized whole gene synthesis and cloned into a modified pET30 vector utilizing ligation independent cloning and primers 5’ TTAAGAAGGAGATATACCATGGTGTCAAAAATCATCGACATCAAGACCC 3’

5’ gattggaagtagaggttctctgcCAGATTAAAACCAAACCCCTTCAG 3’ as PCR primers. The modified pet30 vector produces protein with a C-terminal TEV cleavable duel hexahis-STREP tag (sequence CTERM-AENLYFQSHHHHHHWSHPQFEK). Protein was expressed in BL21(DE3)T1R-RIL in a LEX48 (Harbinger Biotechnology) fermentor. 10 ml of an overnight culture was added to a 1L bottle containing M9 SeMET High-Yeild Growth Media (Shanghai Medicilon Inc.) with 100 μg/L SeMET and 100 μg/ml Kanamyacin. Cells were grown at 38.5 °C to an OD600 of 1.2. The temperature of the culture was reduced to 20 °C whereupon IPTG was added to 1mM. Cells were grown for 12-16 hours after which they were pelleted by centrifugation at 4000g for 15 minutes.

Pellets were resuspended in 25 mls of buffer A (20mM HEPES pH 7.6, 300 mM NaCl, 10% glyerol, 0.1% Sodium Azide) containing 1 mM PMSF and 1 ug/ml DNase and lysed by sonication. Cells were pelleted at 6000 g for 45 minutes. The supernatent was applied to a Strep-Tactin (IBA) column, washed with 5 column volumes of buffer A, and eluted with buffer B (20 mM HEPES, pH = 7.6, 300 mM NaCl, 2.5 mM Desthiobiotin, 10% Glycerol, 0.1% Sodium Azide) directly onto a 1ml HighTrap Fast Flow Ni-NTA column (GE-Healthcare). The bound protein was washed with 10 column volumes of buffer C (20 mM HEPES, pH 7.5, 500 mM NaCl, 20 mM Imidazole, 10% Glycerol, 0.1% Sodium Azide) prior to being eluted from the column with buffer D (20 mM HEPES, pH 7.5, 500 mM NaCl, 500 mM Imidazole, 10% Glycerol, 0.1% Sodium Azide), with the peak fraction captured and loaded onto a HighFlow Superdex S200 16 60 Prep Grade column (GE-Helathcare) equilbrated with buffer E (20 mM HEPES, pH = 7.5, 150 mM NaCl, 10% Glycerol, 0.1% Sodium Azide, 0.5mM TCEP). Peak fractions eluting at the correct volume for the expected mass were pooled and concentrated to 10-15 mg using centrifugal concentrators.

Detailed Methods: Screening by Mass Spectroscopy

Typical assay conditions for the incubation of dipeptide epimerases with dipeptide libraries are as follows: Screens were carried out in D_2O (50-250 μL) containing 20 mM NH_4HCO_3 (pD 7.9) and between 100 μM and 1 mM of each dipeptide. A portion of the solution was set aside as a negative control; the dipeptide epimerase was added to the remaining solution at a concentration of 1 μM. The reactions were incubated at 37 ºC and quenched at various time intervals (from 30 min to 16 hr), quenched with 2 μL 5M NH_4OH, and dried by Speedvac. The sample was resuspended in 25 μL of ddH_2O and diluted 1:1 with methanol. Analysis for incorporation of solvent deuterium was carried out by ESI mass spectrometry where a M+1 mass shift indicates deuterium incorporation. The analysis was carried out in positive mode for all libraries with a net neutral or positive charge; the negative mode was used for libraries with a net negative charge. In some cases, the screening was repeated with lower concentrations of enzyme and with shorter incubation times to better assess substrate specificity.

Detailed Methods: Kinetics

Kinetic parameters were determined via polarimetry. The difference in optical rotation between the initial reading before enzyme was added and after attainment of equilibrium was used to obtain the value for the change in molar ellipticity. The assays were carried out in a 100 mm pathlength cuvette and a Jasco P-1010 polarimeter and a Hg 405 nm filter. Assays were performed at room temperature (~28°C) in a total volume of ~1.4 mL. Typical buffer conditions for the assay were the following: 20 mM Tris-HCl (pH 8.0) and 10 mM $MgCl_2$. Values for k_{cat} and K_M were determined by fitting initial velocities to the Michaelis-Menten equation using

EnzFitter (Biosoft, Cambridge, UK), and subsequent rate constants were divided by two to account for reversibility as the K_{eq} values were ~1. A minimum of two independent sets of kinetic analysis were used to determine standard deviations.

Detailed Methods: Computation

Network Analysis. BLAST e-values for sequences in the dipeptide epimerase family were obtained from the Structure Function Linkage Database (SFLD).(2) SFLD BLAST searches are performed by comparing each sequence in a superfamily against each other. For efficiency, searches are performed by BLASTing bundles of 100 query sequences against all other superfamily sequences. Results are post processed to obtain the equivalent blast2seq e-value (independent of database size) based on bit score. Cytoscape networks (3) were created from these BLAST results at several different e-value cutoffs. Tools used for visualization of protein networks were created by the UCSF Resource for Biocomputing, Visualization, and Informatics and are available from the Resource (http://www.rbvi.ucsf.edu). Each node in the network represents a single sequence in the dipeptide epimerase family and each edge represents the pairwise connection between two sequences with the most significant BLAST e-value (better than the cut-off) connecting the two sequences. The nodes were arranged using the yFiles organic layout provided with Cytoscape version 2.7. Using this layout, lengths of edges are not meaningful except that sequences in tightly clustered groups are relatively more similar to each other than sequences with few connections. Annotation information, including species, lineage, and available x-ray structures (obtained from the SFLD) and substrate specificity (obtained from this study) was associated with each node, as applicable. Connections between nodes are only

shown if the e-value of the best BLAST hit between two sequences is at least as good as the specified e-value cutoff.

Homology Modeling and in silico Ligand Docking. The computational procedures were described in detail in an earlier publication.(4) Briefly, homology models of 66 sequences that clustered with experimentally characterized AEEs from *B. subtilis* and *E. coli* have been built using Protein Local Optimization program (PLOP) (commercially available as Prime through Schrödinger LLC). At the time of our modeling there were only 3 AEE structures were available. We have used the structure from *B. subtilis* (1TKK) as the template. This structure was co-crystallized with the dipeptide substrate Ala-Glu and the metal ion Mg^{2+}; we retained both of them during modeling. The complete list of homology models is provided in Table S1, including sequence identity to the template, which was ~30% in most cases. We also provide the zdope score for the models (http://modbase.compbio.ucsf.edu/evaluation/), a quality metric for homology models, where a score lower than -1.0 is considered 'good', while a score between 0 and -1 is 'reasonable' (5). The models with the worst zdope scores are, unsurprisingly, those with the lowest sequence identity to the template, ~25%. In cases where crystal structures have been determined, we also report RMSD's for the model overall and for the binding site specifically; it should be kept in mind that the crystal structures were determined subsequently. In cases where co-crystal structures were determined, we provide the RMSD for the best substrate, i.e., those shown in Figure 4.

After the models were built, we have docked an *in silico* library of all 400 L/L-amino acid dipeptides against the active site of all the enzymes that we have modeled using the software Glide-V4.0108 (Schrödinger LLC). The dipeptide library was prepared using LigPrep-v2.0106 (Schrödinger LLC).

Molecular Graphics Images. The images in Figures 3 and 4 were constructed using Chimera.(6)

Assignment of Predicted Specificity Groups. The predicted specificity groups shown in Figure 1 were assigned by the Jacobson lab in late 2005/early 2006, before experimental results were available, based on the homology models and the docking results. The sequences were first divided into groups according to a phylogenetic tree constructed for the sequences available at the time (see Ref. (1) for the phylogenetic tree with representative sequences shown). The sequence networks shown in Figure 1 and 2 were constructed later using all sequences available in 2011; many of the sequences that were ultimately obtained for experimental characterization were sequences that had not yet been deposited in 2005. However, it can be seen in Figure 1 that the groupings assigned in 2005 track reasonably well with the clustering observed with the clustering obtained with >700 sequences.

The assigned specificities were based primarily on a simple 'consensus' analysis of the docking results, as well as visual inspection of the homology models and selected binding modes. The consensus analysis was based on the hypothesis that 'averaging' the docking results over all of the enzymes in a putative specificity group would lead to more accurate results. That is, docking is notoriously sensitive to even small errors in the precise positioning of side chains, and in this case these errors could in principle be even more pervasive due to the automated nature of the predictions. Thus, while the results from any one homology model might be suspect, the averaged results might more accurately reflect the true specificities.

The consensus results were obtained by analyzing the amino acids found in the N-terminal and C-terminal positions in the top 5% of the hits lists (top 20 hits out of the 400 L/L dipeptides), averaged over the putative specificity clades. The results for the groups containing

the characterized AEE's in *B. subtilis* and *E. coli* did show the expected preference for small amino acids in the N-terminal position, and negatively charged in the C-terminal position. Note that the results for the *E. coli* clade were based on models with ~30% sequence identity (to 1TKK). A third group showed a similar pattern, but inspection of several of the models revealed a larger N-terminal pocket that could accommodate amino acids larger than Ala, and as a result, this group was labeled as Xxx-Glu. The docking appears to have not captured this, at least for most of the members of the clade, due to keeping the binding pocket rigid, a constraint that we have relaxed in more recent work.

In the N-terminal position, a small group (6 sequences, of which 4 were 'environmental' sequences) showed a striking preference for positively charged amino acids. The C-terminal position was less consistent, and so the prediction was left as "Lys-Xxx". Overall, however, the results did not show the correct preference for positively charged amino acids in the C-terminal position, as revealed by subsequent experiments, although dipeptides such as Arg-Tyr did show some turnover in the mass spectroscopy assay for the *M. capsulatus* enzyme.

Finally, another fairly small group showed, in the consensus results, divergent specificity primarily at the C-terminal position, with hydrophobic amino acids preferred on average, although not exclusively. Inspection of several of the homology models, however, confirmed hydrophobic binding pockets. Although Ala-Hyd dipeptides were substrates for enzymes in this group that were characterized experimentally, larger hydrophobic amino acids were frequently tolerated or preferred in the N-terminal position, with the exception of the epimerases from *C. hutchinsonii* and from *T. maritima*.

Table i: Prevalence of amino acids in the N-terminal position; top 5% of docking hits, averaged over the putative specificity groups.

Group (final predicted specificity)	Small (Ala, Ser, Thr, Cys, Gly)	Negative (Glu, Asp)	Positive (Lys, Arg)	Hydrophobic (Met, Leu, Val, Ile, Phe, Trp, Tyr)
Ala-Glu (B. subtilis like)	89%	1%	3%	4%
Ala-Glu (E. coli like)	63%	9%	5%	14%
Xxx-Glu	72%	8%	6%	5%
Lys-Xxx	15%	3%	73%	6%
Ala-Hyd	78%	8%	7%	4%

Table ii: Prevalence of amino acids in the C-terminal position; top 5% of docking hits, averaged over the putative specificity groups.

Group (final predicted specificity)	Small (Ala, Ser, Thr, Cys, Gly)	Negative (Glu, Asp)	Positive (Lys, Arg)	Hydrophobic (Met, Leu, Val, Ile, Phe, Trp, Tyr)
Ala-Glu (B. subtilis like)	5%	60%	6%	19%
Ala-Glu (E. coli like)	5%	89%	0%	0%
Xxx-Glu	6%	48%	9%	19%
Lys-Xxx	6%	19%	14%	54%
Ala-Hyd	13%	26%	13%	36%

Detailed Methods: Crystallography

Crystallization and Data Collection for the diepeptide epimerase from B. thetaiotamicron (BT1313). Attempts to grow crystals of unliganded BT1313 were unsuccessful. Three different crystal forms of dipeptide liganded BT1313 were grown by the hanging drop method at room temperature, two monoclinic forms complexed with Mg^{2+} and L-Ala-D-Glu and a monoclinic form complexed with Mg^{2+} and L-Pro-D-Glu (Table S27).

For the first crystal form of BT1313•Mg^{2+}•L-Ala-D-Glu, the protein solution contained BT1313 (33 mg/mL) in 50 mM Tris-HCl (pH 7.9), 100 mM NaCl, 5 mM $MgCl_2$, and 50 mM L-Ala-D-Glu; the precipitant contained 25% PEG 3350 and 0.1 M Hepes (pH 7.5). Crystals appeared in 1-2 days and exhibited diffraction consistent with the space group C2, with two molecules per asymmetric unit. Prior to data collection, the crystals were transferred to a cryoprotectant solution composed of 50% PEG 3350, 0.1 M Hepes (pH 7.5), and 5 mM $MgCl_2$; the ligand was not included in cryoprotectant solution.

For BT1313•Mg^{2+}•L-Pro-D-Glu, the protein solution contained BT1313 (27.8 mg/mL) in 20 mM Tris-HCL (pH 7.9), 100 mM NaCl, 5 mM $MgCl_2$, and 50 mM L-Pro-L-Glu; the precipitant contained 17% PEG 10000, 0.1 M Bis-Tris (pH 5.5), and 0.1 M ammonium acetate. Crystals appeared in 4 days and exhibited a diffraction pattern consistent with space group C2, with two molecules per asymmetric unit. Prior to data collection the crystals were transferred to cryoprotectant solution composed of 30% PEG 400, 0.1 M Bis-Tris (pH 5.5), 0.1 M ammonium acetate, and 5 mM $MgCl_2$; the ligand was not included in cryoprotectant solution.

For the second crystal form of BT1313•Mg^{2+}•L-Ala-D-Glu, the protein solution contained BT1313 (33 mg/mL) in 50 mM Tris-HCl (pH 7.9), 100 mM NaCl, 5 mM $MgCl_2$, and 50 mM L Ala-D-Glu; the precipitant contained 17% PEG 3350, 0.1 M Bis-Tris (pH 5.5), and 0.1 M ammonium sulfate. Crystals appeared in 7-8 days and exhibited a diffraction pattern consistent with space group C2, with two molecules per asymmetric unit. Prior to data collection the crystals were transferred to a cryoprotectant solution composed of their mother liquid and 20% glycerol; the ligand was included in cryoprotectant solution for this crystal form.

After 30 seconds incubation in the cryoprotectant solutions, the crystals of all three complexes were flash-cooled in a nitrogen stream. The diffraction data sets were recorded at the

NSLS X4A beamline (Brookhaven National Laboratory) using an ADSC CCD detector. Diffraction intensities were integrated and scaled with programs DENZO and SCALEPACK.(7) The data collection statistics are given in Table S27.

Structure Determination and Model Refinement for BT1313. The structures of the complexes (Table S27) were solved by molecular replacement with fully automated molecular replacement pipeline BALBES(8), using only input diffraction and sequence data. The polypeptide of the L-Ala D/L-Glu epimerase from *Bacillus subtilis* complexed with L-Ala-L Glu (PDB ID 1TKK) was used by BALBES as the template in all three structure determinations. Partially refined structures of all three complexes were the output from BALBES without any manual intervention. Several subsequent iterative cycles of refinement were performed for each crystal form including: manual model rebuilding with TOM, refinement with CNS (9), automatic model rebuilding with ARP (10), and solvent building with the CCP4 suite. Final refinement statistics are provided in the Table S27.

In the final models for one form of the L-Ala-D-Glu and the complex with L-Ala-D-Pro, the N and C terminal residues are disordered; in final model for the second form the L-Ala-D Glu complex, only the N terminal residue is disordered. In each complex, Ala64 in one polypeptide lies in the disallowed region of Ramachandran plot; this residue is located at the tip of the 20s loop. In each active site, well-defined electron density is observed for the dipeptide ligand and the Mg^{2+} ion.

In the first crystal form of BT1313•Mg^{2+}•L-Ala-D-Glu (Table S27, column 1), the carboxylate group of the L-Ala-D-Glu ligand is located in the second coordination sphere of the Mg^{2+}, hydrogen bonded to waters that are coordinated to the Mg^{2+}. In addition, the putative acid/base catalysts, the Nξ atoms of Lys200 and Lys298, are too distant from the α-carbon of the

Glu moiety of the ligand to catalyze the 1,1 proton transfer reaction, with the distances 4.7 and 4.9 Å, respectively. Therefore, this structure is necessarily that of a nonproductive complex.

The structure of the BT1313•Mg^{2+}•L-Pro-D-Glu (Table S27, column 2) complex is also that of a nonproductive complex: the carboxylate group of the ligand is in the second coordination sphere of the Mg^{2+}, with the consequence that the Nξ atoms of Lys200 and Lys298 are too distant from the α-carbon of the Glu moiety of the ligand to catalyze the 1,1 proton transfer reaction.

In the second crystal form of BT1313•Mg^{2+}•L-Ala-D-Glu (Table S27, column 3), the two polypeptides in the asymmetric unit have different coordination geometries for the dipeptide ligand. In polypeptide A, the geometry is similar to that observed in the first crystal form. However, in the polypeptide B, the carboxylate group of the L-Ala-D-Glu ligand is a bidentate ligand of the Mg^{2+}. The Nξ atoms of Lys200 and Lys298 are each located ~4 Å from the α-carbon of the Glu moiety of the ligand and can approach the α-carbon atom as to catalyze the 1,1 proton transfer reaction.

Crystallization and Data Collection for the dipeptide epimerase from C. hutchinsonii (CHU2140). Initial attempts to obtain diffracting crystals of CHU2140 were unsuccessful both in the presence and absence of dipepdide ligands. One month after the initial setup of the Hampton Research Index Screen, birefringent precipitate was discovered in conditions #42 and #70. Optimization of #42 was carried out by varying the chain-length of the PEG precipitant, pH of the crystallization condition, and the temperature of incubation environment. After optimization, two forms of fused rod-shaped monoclinic crystals of CHU2140 were obtained, complexed with

Mg^{2+} and D-Ala-L-Ala, and with Mg^{2+} and D-Ala-L-Val. The dipeptide-liganded crystals were grown by the hanging drop vaporization method at 4°C.

For the crystals of CHU2140•Mg^{2+}•D-Ala-L-Ala, the protein solution contained CHU2140 (40 mg/mL) in 20 mM Hepes (pH 8.35), 100 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, and 20 mM D-Ala-L-Ala; the precipitant contained 10% PEG 4000 and 0.1 M Na-citrate (pH 4.0). Crystals appeared in 7-10 days and exhibited diffraction consistent with the space group $P2_1$, with nine molecules per asymmetric unit. Prior to data collection, the crystals were transferred to a cryoprotectant solution composed of 10% PEG 4000, 25% ethylene glycol and 0.1 M Na-citrate (pH 4.0); the ligand was not included in crypoprotectant solution.

For the crystals of CHU2140•Mg^{2+}•D-Ala-L-Val, the protein solution contained CHU2140 (40 mg/mL) in 20 mM Hepes (pH 8.35), 100 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, and 20 mM D Ala-L-Val; the precipitant contained 12% PEG 10000 and 0.1 M Na-citrate (pH 4.0). Crystals appeared in 7-10 days and exhibited diffraction consistent with the space group $P2_1$, with nine molecules per asymmetric unit. Prior to data collection, the crystals were transferred to a cryoprotectant solution composed of 12% PEG 10000, 30% ethylene glycol and 0.1 M Na-citrate (pH 4.0); the ligand was not included in crypoprotectant solution.

After 30 seconds of incubation in the cryoprotectant solution, the crystals were flash-cooled in liquid nitrogen. The diffraction data set for CHU2140•Mg^{2+}•D-Ala-L-Ala crystal was recorded at LS-CAT (Sector 21 ID-F, Advanced Photon Source, Argonne, IL) using a MAR 225 CCD detector. The diffraction data set for CHU2140•Mg^{2+}•D-Ala-L-Val crystal was recorded at LS-CAT (Sector 21 ID-G, Advanced Photon Source, Argonne, IL) using a MAR 300 CCD detector. Diffraction intensities were integrated and scaled with XDS (11). The data collection statistics are given in Table S28.

Structure Determination and Model Refinement for CHU2140. The structures of CHU2140 in complex with dipeptide substrates, and Mg^{2+} were solved by molecular replacement with PHASER (12), using input diffraction, sequence data and the atomic coordinates of the L-Ala-D/L-Glu epimerase from *Bacillus subtilis* (PDB ID 1TKK) modified with CHAINSAW (13) to trim non-identical residues to Ala. The partial solution from PHASER was then subjected to automated model building with *AutoBuild* in PHENIX (14), followed by multiple iterative cycles of manual and automated refinement with COOT (15) and REFMAC5 (16). Final refinement statistics are provided in Table S28. The asymmetric unit of both liganded structures of CHU2140 consists of nine monomers, some of which are less ordered than others. The capping domain loops involved in substrate recognition are well ordered in all monomers. Electron density for the dipeptide ligand and Mg^{2+} ion is observed in all nine polypeptides.

The structure of CHU2140•Mg^{2+}•D-Ala-L-Ala has well defined active sites in all nine monomers. However, the substrate molecule has well-defined density and acceptable B-factors only in polypeptide A. The Nξ atoms of putative acid-base catalysts Lys162 and Lys265 are within ~4 Å of the α-carbon of L-Ala substrate moiety, positioned in proper distance to facilitate 1,1-proton transfer (Figure 4).

The structure of CHU2140•Mg^{2+}•D-Ala-L-Val has similarly ordered active site architectures in all nine monomers. Much like with the previous complex, only polypeptide B contains well-ordered density for the dipeptide substrate with B-factors in the acceptable range for the fitted ligand. The Nξ atoms of putative acid-base catalysts are within ~4 Å of the α-carbon of L-Val moiety, conferring this structure too as a catalytically productive complex.

The active site for both crystal forms contains a dipeptide substrate, which is essentially in rapid equilibrium. However, due to resolution constraints, it is unfeasible to give an unbiased

assessment as to which stereoisomer is predominantly bound in the active sites. Therefore, ligand-fitting into the electron density was performed with D-Ala-L-Xxx dipeptide for both crystal forms.

Crystallization and data collection for the dipeptide epimerase from E. faecalis (EF1511).

The unliganded crystal form was grown by the hanging drop method at room temperature for the dipeptide epimerase from *Enterococcus faecalis*. The protein solution contained the protein (44 mg/mL) in 20 mM Tris (pH 7.9), 100 mM NaCl, and 10 mM $MgCl_2$. The precipitant contained 2.0 M ammonium sulfate, 0.1 M Hepes (pH 7.5), and 10 mM $MgCl_2$. Crystals appeared in 3 days and exhibited diffraction consistent with space group R3, with eight molecules of the protein per asymmetric unit.

Four different crystal forms were grown by the hanging drop method at room temperature for the dipeptide epimerase from *E. faecalis* liganded with dipeptides: (1) The complex EF1511·L-Ile-L-Tyr, (2) The complex EF1511·L-Leu-L-Tyr, (3) The complex EF1511·L-Ser-L-Tyr, and (4) The complex EF1511·L-Arg-L-Tyr. The protein solutions for all four co-cocrystallizations contained the protein (53 mg/mL), 20 mM Tris (pH 7.9), 100 mM NaCl, 40 mM $MgCl_2$, and the corresponding dipeptide (40mM).

(1) For EF1511·L-Ile-L-Tyr complex, the precipitant contained 1.4 M Sodium citrate, and 0.1 M Hepes (pH 7.5). Crystals appeared in 5 days and exhibited diffraction consistent with space group C2, with eight molecules of the complex per asymmetric unit.

(2) For complexes EF1511·L-Leu-L-Tyr and EF1511·L-Arg-L-Tyr, the precipitant contained 1.8 M ammonium sulfate, and 0.1 M Hepes (pH 7.5). In both cases the crystals

appeared in 7 days and exhibited diffraction consistent with space group R3, with eight molecules of the corresponding complex per asymmetric unit.

(3) For EF1511·L-Ser-L-Tyr complex, the precipitant contained 1.6 M sodium citrate, and 0.1 M Hepes (pH 7.5). Crystals appeared in 4 days and exhibited diffraction consistent with space group R3 with eight molecules of the complex per asymmetric unit.

Prior to data collection, the crystals of all 5 dipeptide epimerase EF1511 forms (Table S26) were transferred to cryoprotectant solutions composed of their mother liquids and 20% glycerol and flash-cooled in a nitrogen stream. All 5 X-ray diffraction data sets were collected at the NSLS X4A beamline (Brookhaven National Laboratory) on an ADSC CCD detector. Diffraction intensities were integrated and scaled with programs DENZO and SCALEPACK (7). The data collection statistics are given in Table S26.

Structure determination and model refinement for the dipeptide epimerase from E. faecalis (EF1511)

All 5 EF1511 structures (Table S26) were solved by molecular replacement with fully automated molecular replacement pipeline BALBES (8), using only input diffraction and sequence data. The protein part of the dipeptide epimerase from *Bacillus subtilis* liganded with L-Ala-L-Glu (PDB ID 1TKK) was used by BALBES as a search template. Partially refined structures of all 5 EF1511 crystal forms were the outputs from BALBES. Subsequently, several iterative cycles of refinement were performed for each crystal form including: the model rebuilding with COOT (15), refinement with CNS (9), and automatic model rebuilding with ARP (10).

Final crystallographic refinement statistics for all determined EF1511 structures are provided in the Table S26.

Crystallization and data collection for the dipeptide epimerase from M. capsulatus (MCA1834).

Initial crystallization conditions were determined using the sparse matrix method. All crystals were grown by the hanging drop vaporization method at 4° C. One week after the setup of the Hampton Research Index Screen, diffracting crystals were obtained in condition #77 in the presence of 20x molar excess of L-Arg-L-Lys. Optimization of #77 was carried out as described for CHU2140. After optimization, an additional form of fused plate-like crystals were obtained using similar conditions but in the presence of Mg^{2+}.

For the crystals of MCA1834•Mg^{2+}•L-Arg-D-Lys, the protein solution contained MCA1834 (12 mg/mL) in 50 mM Hepes (pH 8.0), 100 mM KCl, 5 mM $MgCl_2$, and 20 mM L-Arg-L-Lys; the precipitant contained 25% PEG 3500, 200 mM Li_2SO_4, and 100 mM Tris-HCl (pH 8.5). Crystals appeared in 5 days and exhibited diffraction consistent with the space group $I2_13$, with five molecules per asymmetric unit. Prior to data collection, the crystals were transferred to a cryoprotectant solution composed of 30% PEG 3350, 200 mM M Li_2SO_4, and 100 mM Tris-HCl (pH 8.5) and vitrified by direct immersion into liquid nitrogen. The ligand was not included in crypoprotectant solution.

For the crystals of MCA1834•Mg^{2+}, the protein solution contained MCA1834 (10 mg/mL) in 50 mM Hepes (pH 8.0), 100 mM KCl, 5 mM $MgCl_2$, and 5 mM L-Arg-L-Lys; the precipitant contained 10% PEG 4000, 200 mM $MgSO_4$ and 100 mM MES (pH 5.5). Crystals appeared in two weeks and exhibited diffraction consistent with the space group $P2_1$, with six molecules per asymmetric unit. Prior to data collection, the crystals were transferred to a

cryoprotectant solution composed of 12.5% PEG 4000, 25% glycerol, 200 mM $MgSO_4$, and 100 mM MES (pH 5.5) and vitrified as above.

All data were collected on crystals that were maintained at 100K. The diffraction data set for MCA1834•Mg^{2+}•L-Arg-D-Lys crystal was recorded at LS-CAT (Sector 21 ID-G, Advanced Photon Source, Argonne, IL) using a MAR 300 CCD detector. The diffraction data set for MCA1834•Mg^{2+} crystal was recorded at LS-CAT (Sector 21 ID-D, Advanced Photon Source, Argonne, IL) using a MAR 300 CCD detector. Diffraction intensities were integrated and scaled with XDS package (11). The data collection statistics are given in Table S29.

Structure determination and model refinement for MCA1834.

The structures of MCA1834 in complex with L-Arg-D-Lys, and Mg^{2+} were solved by molecular replacement as described for CHU2140. Final refinement was carried out using PHENIX (14). Refinement statistics are provided in Table S29. Of the five molecules in the crystallographic asymmetric unit, all polypeptides contain good electron density for the L-Arg portion of the dipeptide ligand, only polypeptides B, C and E contain well-ordered density for the entire dipeptide substrate. Electron density at the active sites were consistent with the epimerized L-Arg-D-Lys isoform of the substrate. The magnesium ion sits low in the metal binding pocket and has unusually long distances to the carboxylate group of the substrate as well as its respective binding residues, perhaps caused by an elongated substrate molecule. There also seems to be an alternative quaternary Glu242 metal binding ligand that is positioned in an appropriate distance to the metal ion upon substrate binding. The Nξ atoms of Lys162 and Lys266 are positioned both within ~4 Å of the α-carbon of the D-Lys portion of the substrate to facilitate acid-base catalysis.

The structure of MCA1834•Mg^{2+} was solved in a similar manner to that of the dipeptide complexed form of the enzyme. All six monomers of the asymmetric unit are almost identical but are missing the 20s loop portion that is involved in substrate binding. The magnesium ion coordination to the metal binding ligands for this crystal form is more consistent with that observed for other dipeptide epimerase group members. The active site for the enzyme is filled with water molecules, rather than the substrate. Another observable change in addition to the metal-binding site can be seen for the Asp319 residue of the Asp-X-Asp motif. Compared to the dipeptide complexed structure, Asp319 is now rotated ~80˚ about the β-carbon of that residue to make a tight contact with a cluster of water molecules away from the site that would otherwise be occupied by the a-ammonium group of the dipeptide substrate.

Although multiple attempts were made to obtain crystals in the presence of ligands that are catalytically more efficient in enzyme assays, no diffraction quality crystals were produced.

Crystallization and structure determination of Francisella philomiragia ATCC 25017 (GI:167627873, NYSGRC-200555).

The dipeptide epimerase from *F. philomiragia* (Fphi1647) was crystallized by sitting drop vapour diffusion in 96-well *Intelliplates* (Art Robbins Instruments). Protein (2µl, 10 mg/mL, 20 mM HEPES pH 7.5, 150 mM NaCl, 10% Glycerol, 0.1% Sodium Azide, 0.5 mM TCEP) was combined with reservoir solution (2 µl, 2M $(NH_4)_2SO_4$, 100 mM NaCitrate, 200 mM KNaTartrate pH 5.6) and equilbrated against 70 µl of the same reservoir at 20 degrees C. Crystals grew as tetragonal bipyramids (0.2 x 0.2 x 0.2mm) over a period of 2-6 days and belong to space group $P4_12_12$ with cell dimensions of a=b=121, c=149.

For the initial structure determination crystals were soaked in 2.4 M $(NH_4)_2SO_4$, 100 mM NaCitrate, 200 mM KNaTartrate pH 5.6, 20% glycerol prior to vitrification by immersion in liquid nitrogen. The L-Ala-D/L-Glu complex was determined from crystals that were soaked for 30 minutes in 2.4 M $(NH_4)_2SO_4$, 100 mM MES pH 6.0, 20% glycerol, 10 mM L-Ala-D/L-Glu, 50 mM MgSO4. All data were collected at the 31-ID beamline (LRL-CAT), Advanced Photon SOURCE, integrated with MOSFLM (17) and scaled with SCALA (18). Refinement statistics are reported in Table S30.

The structure of Fphi1647 was determined by selenomethionine single anomalous dispersion collected at the selenomethionine anomalous peak. The determination of selenomethionine positions, calculation of solvent-flattened phases, and the autofitting of the majority of the structure were performed in PHENIX (14). Iterative cycles of manual fitting followed by refinement against the data utilized COOT (15) and PHENIX, respectively. Waters were added to Fo-Fc peaks greater than 3.5 sigma that were within hydrogen bonding distance to a potential partner. Translation-libration-screw refinement (TLS (19)) was utilized during the final rounds with individual bodies determined by the TLS-server. The determination of the L or D form of glutamic acid in the dipeptide complex utilized careful analysis of 2Fo-Fc and Fo-Fc maps generated by refined structures with and without the forms of the dipeptide.

There is a dimer in the asymmetric unit that is consistent with the molecular dimer suggested by gel filtration chromatography (data not shown), and is similar to one of the dimer interfaces seen in the *B. subtilus* octamer. The initial Fphi1647 structure bound a tartrate molecule in subunit A and a sulfate in subunit B (3R0K) but was devoid of Mg. Density for the 20's loop in subunit A was not of sufficient quality to build a molecular model, but was

suggestive of an open conformation, while the 20's loop in subunit B was visible and closed over the active site.

Removal of dicarboxylates from the soaking solution and inclusion of 10 mM L-Ala-L-Glu and 50 mM Mg resulted in a dipeptide bound structure (3R1Z), though no Mg was found. The two active sites are unique in their interactions with the bound ligands. In subunit A the density was consistent with a bound L-Ala-L-Glu, while in subunit B the density was consistent with L-Ala-D-Glu. In subunit A, L-Ala-L-Glu is bound with the anomeric carbon 3.5 Å from Lys164 and 4.3 Å from Lys269 the presumed active site acid and/base. In contrast, in subunit B, L-Ala-D-Glu is bound in a similar fashion but is 3.7 Å from Lys164 and 3.4 Å from Lys269. In both subunits the amine of L-Ala is coordinated by the side chains of Asp321, Asp323 and Ser137, while the amide of the dipeptide is coordinated by the backbone carbonyl of Cys297. In addition the side chain carboxylate makes a salt bridge with Arg26 of the 20s loop resulting in a closed loop structure in both subunits. Utilizing the Mg-bound structure as a model, the binding conformation of the dipeptide in subunit A would be monodentate with a ligand distance of 3 Å and in subunit B would be bidentate with ligand distances of 2.5 and 2.8 Å. The differential binding of ligands to the individual subunits is most likely due to the dissimilar crystal contacts of the subunits restricting the flexibility of the protein.

Crystallization and structure determination of the dipeptide epimerase from Herpetosiphon aurantiacus

The target was expressed as a full-length protein supplemented with the C-terminal His-TAG (EGHHHHHH). Following the standard expression and purification protocols, the initial solution was concentrated to about 37 mg/mL of protein by using 30K Amicon centrifugal

microtubes. The protein was crystallized by the sitting drop vapor diffusion method. In brief, 1 microliter of a protein solution was mixed with an equal volume of a mother liquor which consists of 150 mM KBr and 30% PEG2000 MME as precipitants and equilibrated at room temperature against the same mother liquor in clear tape sealed 96-well crystallization plates. The large (around 1 mm in diameter) single diamond-shaped crystal was observed in one of the drops after two months of incubation. The crystal was collected by the Hampton Research cryogenic loop, quickly transferred in liquid nitrogen, and stored frozen in liquid nitrogen until X-ray analysis and data collection. The X-ray diffraction data from the frozen crystal were collected on the Beamline X29A (NSLS, Brookhaven National Laboratory) at wavelength of 1.08 Å. The analysis of diffraction images has revealed the polycrystalline nature of this crystal. However, since it was quite sizable, we have examined diffraction patterns coming from different parts of the crystal and finally found out that its triangular tip represents a single crystal suitable for data collection. The X-ray diffraction data were collected to resolution of 2.07 Å, then they were processed and scaled using HKL2000 program package (7). The crystal structure was solved by molecular replacement using the coordinates for 1JPM (PDB) as a search model. The structure was refined by REFMAC 5.03 (CCP4 Program package (16)) and the model was fixed manually using COOT (15) visualization and refinement software. The data collection and refinement statistics for this structure are shown in the Table S31.

References

1. Song L*, et al* (2007) Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat Chem Biol* 3: 486-491.

2. Pegg SC, *et al* (2006) Leveraging enzyme structure-function relationships for functional inference and experimental design: The structure-function linkage database. *Biochemistry* 45: 2545-2555.

3. Cline MS, *et al* (2007) Integration of biological networks and gene expression data using cytoscape. *Nat Protoc* 2: 2366-2382.

4. Kalyanaraman C, *et al* (2008) Discovery of a dipeptide epimerase enzymatic function guided by homology modeling and virtual screening. *Structure* 16: 1668-1677.

5. Shen MY & Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15: 2507-2524.

6. Pettersen EF, *et al* (2004) UCSF chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.

7. Otwinowski Z & Minor W (1997) in *Methods in Enzymology,* eds Carter CWJ, Sweet RM, Abelson JN & Simon MI, Academic Press (New York), pp 307-326.

8. Long F, Vagin AA, Young P & Murshudov GN (2008) BALBES: A molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr* 64: 125-132.

9. Brunger AT, *et al* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54: 905-921.

10. Lamzin VS & Wilson KS (1993) Automated refinement of protein models. *Acta Crystallogr D Biol Crystallogr* 49: 129-147.

11. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystal* 26: 795-800.

12. McCoy AJ, *et al* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658-674.

13. Stein N (2008) CHAINSAW: A program for mutating pdb files used as templates in molecular replacement. *J Appl Crystal* 41: 641-643.

14. Adams PD*, et al* (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58: 1948-1954.

15. Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of coot. *Acta Crystallogr D Biol Crystallogr* 66: 486-501.

16. Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240-255.

17. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* 26:

18. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62: 72-82.

19. Winn MD, Isupov MN & Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* 57: 122-13