DOI: 10.1002/cbic.200800553

Mapping the Limits of Substrate Specificity of the Adenylation Domain of TycA

Benoit R. M. Villiers and Florian Hollfelder*[a]

The catalytic potential of tyrocidine synthetase 1 (TycA) was probed by the kinetic characterization of its adenylation activity. We observed reactions with 30 substrates, thus suggesting some substrate promiscuity. However, although the TycA adenylation (A) domain was able to accommodate alternative substrates, their k_{cat}/K_M values ranged over six orders of magnitude. A comparison of the activities allowed the systematic mapping of the substrate specificity determinants of the TycA A-domain. Hydrophobicity plays a major role in the recognition of substrate analogues but can be combined with shape complementarity, conferring higher activity, and/or steric exclusion, leading to substantial discrimination against larger substrates. A comparison of the k_{cat}/K_M values of the TycA A-domain and phenylalanyl-tRNA synthetase showed that the level of discrimination was comparable in the two enzymes for the adenylation reaction and suggested that TycA was also subjected to high selective pressure. The specificity patterns observed and the quantification of alternative activities provide a basis for exploring possible paths for the future directed evolution of A-domain specificity.

Introduction

Natural products are well-recognized as a central repertoire of molecular scaffolds with diverse functional potential that has yielded biochemical tools, drug leads and useful drugs.[1] Many of these natural products such as nonribosomal peptides or polyketides are synthesized by large enzymes, nonribosomal peptide synthetases (NRPSs) or polyketide synthases (PKSs), which can be seen as molecular "assembly lines".[2,3] The prospect of using these assembly lines for making altered natural products is crucially dependent on the tolerance and limitations of substrate recognition by the domains involved. Protein catalysts are usually described in textbooks as very specific for one unique substrate, selected from the cellular cocktail of hundreds of alternative compounds to ensure metabolic fidelity. If this high degree of substrate specificity were absolute, it would pose substantial limitations for the potential applications of enzymes, because substrate analogues would not be accepted and processed in the same way as the natural product. In recent years, this hardline view of enzyme specificity has been softened by the observation of enzyme promiscuity.[4–6] Enzyme promiscuity is defined as the ability of some enzymes to turn over different substrates, and the growing number of enzymes identified as catalytically promiscuous indicates that promiscuity is not a rare but a rather widespread intrinsic feature of enzymes.[5–8]

It has been postulated that enzyme promiscuity played a role in evolution, where a duplicated gene might have relieved the original gene copy from selective pressure, allowing it to mutate randomly and adapt a new function by enhancing an existing promiscuous activity.[5,8,9] A consequence of this scenario is the definition of evolutionary links between related activities based on enzyme promiscuity.[10] Conceptually, this idea allows for the approximation of the evolutionary distance between two activities, which could ultimately give rise to a systematic understanding of the phylogeny of catalysis. For NRPSs, this phylogeny of catalysis could complement previous sequence alignments of adenylation domain (A-domain) signatures.[11,12] Furthermore, such relationships might not just report on evolutionary history, but might also provide a guide for future directed evolution as how to efficiently traverse diversity space with evolutionary stepping stones and shortcuts to accelerate directed evolution efforts, thereby economising the screening or selection effort. In an extreme view, the relevance of promiscuity for protein engineering has been stated as "one cannot evolve what is not already there",[13] explaining the surge in interest in catalytic promiscuity.

This logic can also be brought to bear on the question of how specificity evolved in NRPSs—and could help to realize the long-term objective of combinatorial biosynthesis to create biosynthetic enzyme assembly lines with new and altered specificities—by pointing to the best strategy for evolving modules with new specificities. Combinatorial biosynthesis attempts to manipulate the enzymes involved in the biosynthesis of such natural products to make analogues chemoenzymatically accessible by reprogrammed biosynthesis. Examples are the NRPSs (such as those of daptomycin,[14] balhimycin[15] and surfactin[16]) and the PKSs.[3,17,18] The modularity of the natural product assembly lines, in principle, offers prospects for their reprogramming to create novel functional natural products with new or altered properties. Module and domain swap-

[a] *B. R. M. Villiers, Dr. F. Hollfelder*
Department of Biochemistry, University of Cambridge
80 Tennis Court Road, Cambridge, CB2 1GA (UK)
Fax: (+44) 1223-766002
E-mail: fh111@cam.ac.uk

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.200800553.

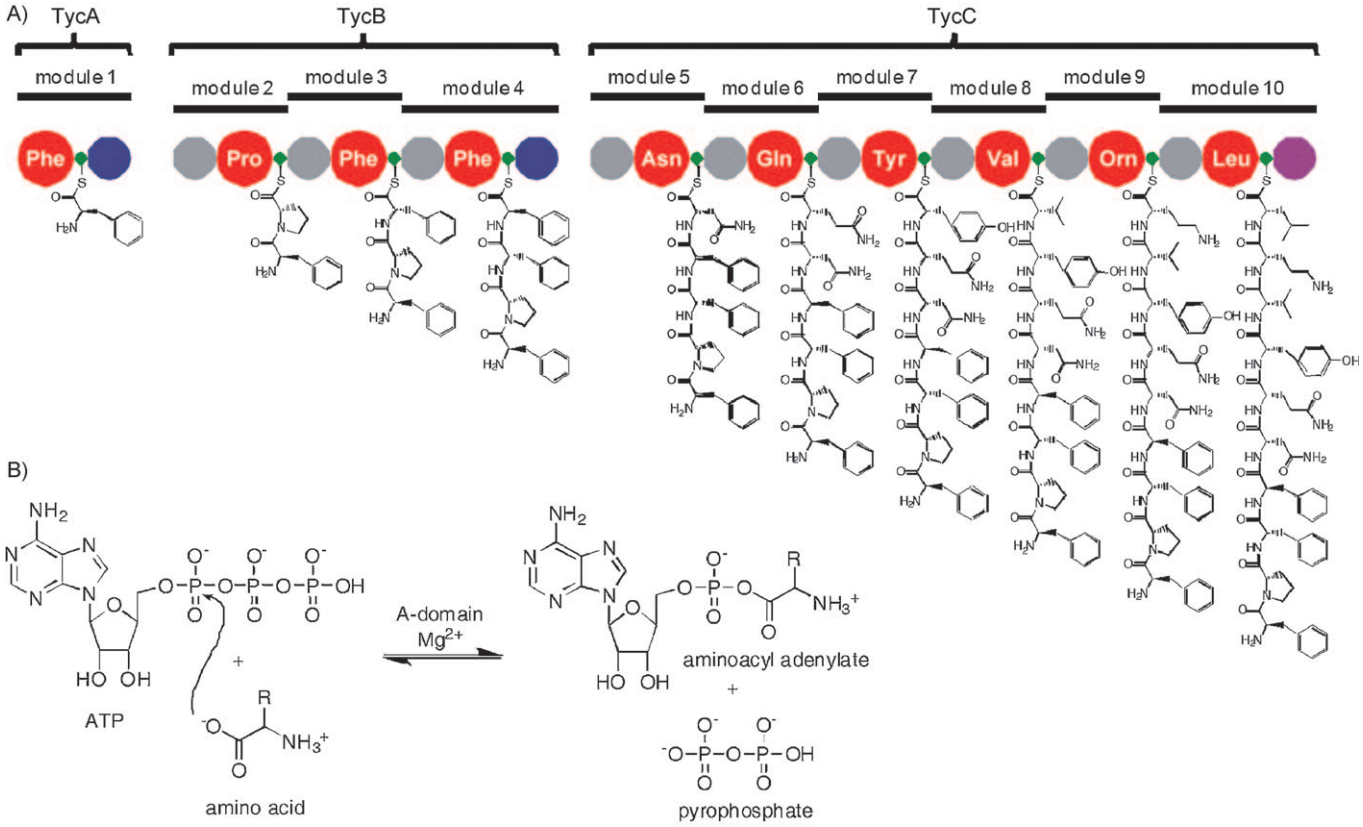


Figure 1. The tyrocidine biosynthetic system and the enzymatic reaction catalyzed by A-domains. A) The enzymatic assembly line of the cyclic decapeptide antibiotic tyrocidine A consists of three polypeptide chains (TycA, TycB and TycC), which are composed of one, three and six modules, respectively, each of which is responsible for the incorporation of one monomeric amino acid into the nascent peptide chain. Each module can be further divided into individual catalytic domains: adenylation domains (A, red), thiolation domains (or peptidyl-carrier protein domains, T, green), condensation domains (C, grey), epimerization domains (E, blue) and thioesterase domains (TE, violet) for peptide cylization. TycA, studied here, is composed of an A-domain, a T-domain and an E-domain and is often referred to as PheATE since its natural substrate is L-Phe. B) A-domains catalyze the reversible formation of a reactive aminoacyl adenylate at the expense of ATP. The ATP/PP_i-exchange reaction is used to test substrate acceptance. By adding radiolabeled pyrophosphate ($[^{32}P]PP_i$), the equilibrium is shifted to the left, leading to the formation of radiolabeled ATP that can be adsorbed onto activated charcoal. The formation of radiolabeled ATP is proportional to the activity of the A-domain.[24]

ping have been frequently investigated in this regard.[17,19] Unfortunately, NRPSs modified in this way are often considerably impaired and exhibit low activity, presumably due to the disruption of crucial interdomain contacts[20] limiting the production of the desired altered natural products.[14,21] Altering the specificity of NRPS domains by the manipulation of individual residues—either by directed evolution[22] or by rational redesign[23]—is attractive because it offers a less invasive approach to the tailoring of NRPS domains for combinatorial biosynthesis than does the transplantation of domains from other enzyme systems.

In this paper, we explore the substrate specificity of the adenylation activity of tyrocidine synthetase 1 (TycA; Figure 1 A), an NRPS module that is involved in the biosynthesis of the cyclic, decapeptide, antibiotic tyrocidine A by activating the first amino acid building block (L-Phe). The multi-modular enzymatic architecture of the tyrocidine biosynthetic system is shown in Figure 1 A. The first specificity-determining step of nonribosomal peptide synthesis is the reversible activation of the amino acid building blocks by the adenylation of their carboxylates under ATP consumption and in the presence of a magnesium ion. This step is carried out by A-domains (Figure 1 B), such as the TycA A-domain studied here. NRPS A-domains belong to a superfamily of adenylate-forming enzymes including the acyl-coenzyme A synthase and luciferase families. They share no sequence or structure homology with class I and class II aminoacyl-tRNA synthetases, which are different adenylate-forming enzymes bearing different folds.[25]

The activation of amino acids by the NRPS A-domains is followed by the formation of a thioester (by the thiolation domain, T-domain) and a peptide bond (by a condensation domain, C-domain). Although a large number of A-domains have been studied, completely nonspecific A-domains have been elusive so far. This makes it important to probe the specificity of existing domains, and TycA, representing a large number of enzymes with homologous A-domains, is used here as a model NRPS A-domain. The crystal structure of the A-domain of gramicidin S synthetase 1 (GrsA),[26] a close homologue of TycA, and sequence alignments have helped to define a specificity-conferring code consisting of ten amino acids lining the binding pocket.[11,12] GrsA is the best-characterized A-domain, but the measurement of the kinetic parameters

defining its substrate specificity has been limited to very few substrates that are chemically and sterically similar to the natural substrate, L-Phe.[27] The physico-chemical and thermodynamic forces driving substrate binding and catalysis in NRPS A-domains have never been investigated comprehensively to quantify and understand substrate specificity in this class of homologous enzymes, which typically show 30–80% sequence identity.[25]

We report herein an extensive survey of the substrate specificity of the adenylation activity of TycA using a discontinuous ATP/pyrophosphate(PP_i)-exchange assay shown in Figure 1B. This assay has been used to determine the adenylation rates for aminoacyl-tRNA synthetases[28] and NRPS A-domains[24,29] and relies on the reversibility of the amino acid adenylation. For NRPS A-domains, it has been shown that the results obtained by this assay correlate well with the amino acid loading assays on the T-domain (using radiolabeled amino acid trans-

Table 1. Michaelis–Menten parameters[a] for substrate analogues[b] of TycA in the order of their k_{cat}/K_M values within their category.

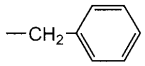
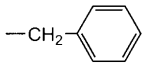
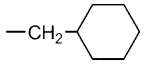
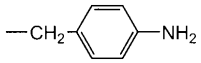
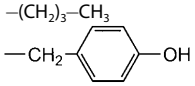
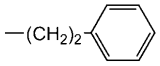
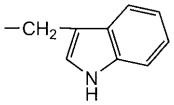
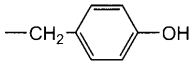
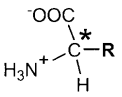
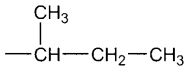
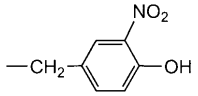
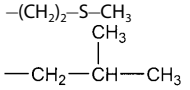
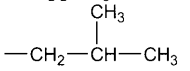
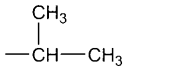
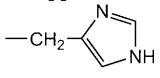
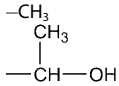
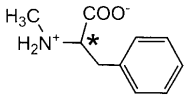
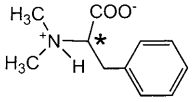
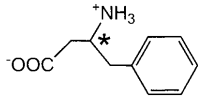
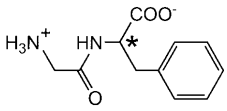
Category	No.	R or full structure	*	k_{cat} [min^{-1}]	K_M [mM]	k_{cat}/K_M [min^{-1} mM^{-1}]	Relative catalytic efficiency
	1		L	118.5 ± 0.7	12.0 ± 0.3 × 10^{-3}	9887 ± 325	1
	2		D	113 ± 2	24 ± 2 × 10^{-3}	4650 ± 383	4.7 × 10^{-1}
	3		L	18.8 ± 0.3	83 ± 6 × 10^{-3}	227 ± 20	2.3 × 10^{-2}
	4		L	105 ± 3	3.2 ± 0.3	33 ± 4	3.3 × 10^{-3}
	5	–$(CH_2)_3$–CH_3	L	92 ± 2	7.4 ± 0.6	12.6 ± 1.4	1.3 × 10^{-3}
	6		L	42 ± 2	3.4 ± 0.2	12.2 ± 1.4	1.2 × 10^{-3}
	7		L	8.3 ± 0.7	1.45 ± 0.29	5.7 ± 1.7	5.8 × 10^{-4}
	8		L	11.3 ± 0.2	2.1 ± 0.2	5.4 ± 0.5	5.5 × 10^{-4}
	9		D	51 ± 13	13.9 ± 4.1	3.7 ± 2.2	3.7 × 10^{-4}
amino acids 	10		L	27.9 ± 0.8	8.3 ± 0.7	3.4 ± 0.4	3.4 × 10^{-4}
	11		L	22.5 ± 1.6	8.1 ± 1.2	2.8 ± 0.6	2.8 × 10^{-4}
	12	–$(CH_2)_2$–S–CH_3	L	23.2 ± 0.5	10.9 ± 0.7	2.1 ± 0.2	2.1 × 10^{-4}
	13	–$(CH_2)_2$–S–CH_3	D	29.4 ± 1.3	21 ± 2	1.37 ± 0.20	1.4 × 10^{-4}
	14		L	13.8 ± 0.3	11.0 ± 0.8	1.26 ± 0.12	1.3 × 10^{-4}
	15	–$(CH_2)_2$–CH_3	L	36.1 ± 0.7	53 ± 3	0.68 ± 0.05	6.9 × 10^{-5}
	16		D	7.2 ± 0.4	26 ± 3	0.28 ± 0.05	2.8 × 10^{-5}
	17		L	13.5 ± 0.2	103 ± 4	0.130 ± 0.008	1.3 × 10^{-5}
	18	–CH_2–CH_3	L	23.9 ± 0.7	506 ± 28	47 ± 4 × 10^{-3}	4.8 × 10^{-6}
	19	–$(CH_2)_2$–SH	L	2.24 ± 0.08	68 ± 6	33 ± 4 × 10^{-3}	3.3 × 10^{-6}
	20		L	0.62 ± 0.03	19.5 ± 2.1	32 ± 5 × 10^{-3}	3.2 × 10^{-6}
	21	–CH_3	L	17.4 ± 1.8	577 ± 82	30 ± 8 × 10^{-3}	3.0 × 10^{-6}
	22		L	6.6 ± 0.8	544 ± 97	12.2 ± 3.7 × 10^{-3}	1.2 × 10^{-6}
	23	–CH_2–OH	L	1.60 ± 0.08	249 ± 43	6.4 ± 1.5 × 10^{-3}	6.5 × 10^{-7}
	24	–CH_2–SH	L	0.89 ± 0.05	158 ± 18	5.6 ± 1.0 × 10^{-3}	5.7 × 10^{-7}

Table 1. (Continued)

Category	No.	R or full structure	*	k_{cat} [min^{-1}]	K_M [mM]	k_{cat}/K_M [min^{-1} mM^{-1}]	Relative catalytic efficiency
N-methylated amino acids	25		L	10.7 ± 0.4	4.2 ± 0.4	2.6 ± 0.4	2.6×10^{-4}
	26		L	0.38 ± 0.02	7.7 ± 1.1	$50\pm10\times10^{-3}$	5.1×10^{-6}
β amino acid	27		L	1.00 ± 0.02	0.95 ± 0.07	1.05 ± 0.09	1.1×10^{-4}
dipeptide	28		L	5.0 ± 0.3	7.4 ± 1.0	0.67 ± 0.13	6.8×10^{-5}

Conditions: [PheATE-His] = 0.1 or 1 μM (see Figure 2 or Figure S1), [$MgCl_2$] = 10 mM, [HEPES] = 50 mM, pH 8.0, [EDTA] = 1 mM, [NaCl] = 100 mM, [^{32}P]PP_i: 0.045 μCi per reaction, [PP_i] = 0.1 mM, [ATP] = 2 mM. [a] Typical time courses and Michaelis–Menten plots are shown in Figure 2. Figure S1 (Supporting Information) contains these data for the remaining substrates. For some substrates (for example, **6**, **9**, **16**, **18**, **20**, **21**, **22**, **24**, and **28**) concentrations $>5K_M$ could not be reached, so the statistical error indicated might underestimate the actual error in these data sets. [b] L-Phe (**1**), D-Phe (**2**), β-cyclohexyl-L-alanine (**3**), 4-amino-L-Phe (**4**), L-norleucine (**5**), L-Tyr (**6**), L-homophenylalanine (**7**), L-Trp (**8**), D-Tyr (**9**), L-Ile (**10**), 3-nitro-L-Tyr (**11**), L-Met (**12**), D-Met (**13**), L-Leu (**14**), L-norvaline (**15**), D-Leu (**16**), L-Val (**17**), L-2-aminobutyric acid (**18**), L-homocysteine (**19**), L-His (**20**), L-Ala (**21**), L-Thr (**22**), L-Ser (**23**), L-Cys (**24**), *N*-methyl-L-Phe (**25**), *N*,*N*-dimethyl-L-Phe (**26**), L-β-homophenylalanine (**27**), Gly-L-Phe (**28**).

fer[27] or mass spectrometry[30]). However, the determination of the kinetic parameters for alternative substrates has been limited by the sensitivity of the assay,[27] which, being discontinuous, is also labour-intensive.

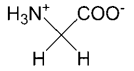
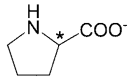
We have previously adapted the classic ATP/PP_i-exchange assay to a 96-well format allowing for the measurement of four plates (or 384 time points) per day, with high sensitivity, surpassing the performance of previous protocols.[31] We now systematically explore the ability of TycA to process substrate analogues, taking advantage of the assay procedure to provide accurate quantitative detail for 30 alternative substrates, which differ by six orders of magnitude in k_{cat}/K_M. The analysis of the trends in substrate recognition is made possible by studying an unprecedentedly large number of alternative substrates and reveals patterns that characterize the factors governing substrate specificity in the TycA A-domain active site. These patterns provide a basis for approximating how remotely these other activities are related functionally and allow an assessment of the difficulty of directed evolution approaches in reprogramming A-domains in combinatorial biosynthesis.

Results

Kinetic parameters for 30 substrates

Tables 1 and 2 list Michaelis–Menten parameters for the 30 substrates studied. In addition to the native substrate, L-Phe (**1**), kinetic parameters were accessible for compounds with both smaller and larger side chains including hydrophobic and polar groups. The substrates were mostly natural or unnatural amino acids but also included *N*-methylated amino acids, a β amino acid, a dipeptide and compounds with alternative stereochemistry at C_α.

Table 2. Kinetic parameters[a] for substrates exhibiting substrate inhibition.

No.	Structure	*	k'_{cat} [min^{-1}]	K'_M [mM]	K_{SI} [mM]
29		n.a.[b]	0.23 ± 0.06	116 ± 50	249 ± 114
30		L	0.71 ± 0.23	1695 ± 681	903 ± 437

Conditions: [PheATE-His] = 1 μM, [$MgCl_2$] = 10 mM, [HEPES] = 50 mM, pH 8.0, [EDTA] = 1 mM, [NaCl] = 100 mM, [^{32}P]PP_i: 0.045 μCi per reaction, [PP_i] = 0.1 mM, [ATP] = 2 mM. [a] The time courses and the substrate inhibition plots are shown, respectively, in Figure S1 BA and BB for Gly (**29**), and Figure S1 BC and BD for L-Pro (**30**) (see the Supporting Information). [b] Not applicable.

Typical Michaelis–Menten plots for the best (L-Phe, **1**) and worst (L-Cys, **24**) measurable substrates are shown in Figure 2 B and D, respectively. The rates shown were derived from time courses for which examples are shown in Figure 2 A (for L-Phe, **1**) and C (for L-Cys, **24**). All time courses and Michaelis–Menten plots for the other substrates listed in Tables 1 and 2 are shown in Figure S1 in the Supporting Information.

Generally, conventional saturation curves were observed, but Gly (**29**) and L-Pro (**30**) exhibited substrate inhibition (Table 2). This means that higher substrate concentrations were inhibitory, giving rise to an uncompetitive inhibition pattern, and that an inhibition constant could be determined (Table 2). These two amino acids are small (Gly, **29**) or relatively compact (L-

Pro, **30**), so that the binding of a second substrate molecule at the active site is conceivable.[32] For some substrates, a full saturation profile could not be obtained due to limiting substrate solubility (see the Supporting Information).

In our discussion, we use the Michaelis–Menten parameters to describe the binding and catalysis of alternative substrates. Second order rate constants (k_{cat}/K_M) were used to describe relative specificity.[33] Given the relatively low turnover rates, substrate association is likely to be fast relative to the catalytic turnover, allowing for an approximation of substrate binding by K_M. Nonproductive activation (that is, the hydrolysis of the miscognate aminoacyl adenylates, or pretransfer editing)[34] has been explicitly ruled out by Luo et al. in the highly homologous GrsA A-domain.[27,35]

The dynamic range covered by these substrates varies 50000-fold in K_M, 300-fold in k_{cat} and 10^6-fold in k_{cat}/K_M. In most cases, the data points adduced for the calculation of initial rates correlate extraordinarily well (correlation coefficients $R > 0.992$), demonstrating that the assay procedure produced high-quality data, even for the substrates whose activity was marginal, allowing for the measurement of k_{cat}/K_M as small as 6×10^{-3} min^{-1} mM^{-1}. Activities against substrates with a second order rate constant (k_{cat}/K_M) approximately tenfold below this value could not be detected, providing a lower limit for possible assignments at 10^{-2}–10^{-3} min^{-1} mM^{-1}.

Five substrates had been tested previously in the same assay with the highly homologous GrsA A-domain, which shares an identical binding pocket (Figure 3)[26] and exhibits 79% amino acid sequence homology (or 64% amino acid

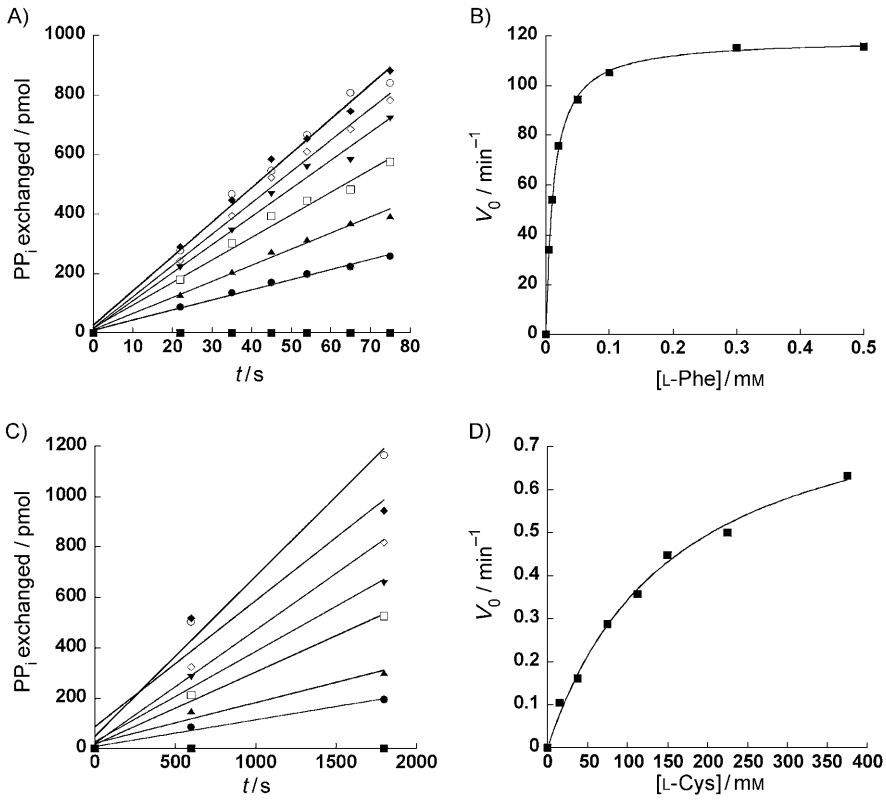


Figure 2. Determination of Michaelis–Menten parameters for substrate analogues of TycA by the ATP/PP$_i$-exchange assay. A) Linear time courses and B) Michaelis–Menten plot for the best substrate, L-Phe (**1**); C) Linear time courses and D) Michaelis–Menten plot for the worst substrate, L-Cys (**24**). Conditions: [PheATE-His] = 0.1 μM for L-Phe and [PheATE-His] = 1 μM for L-Cys, [$MgCl_2$] = 10 mM, [HEPES] = 50 mM, pH 8.0, [EDTA] = 1 mM, [NaCl] = 100 mM, [^{32}P]PP$_i$: 0.045 μCi per reaction, [PP$_i$] = 0.1 mM, [ATP] = 2 mM.

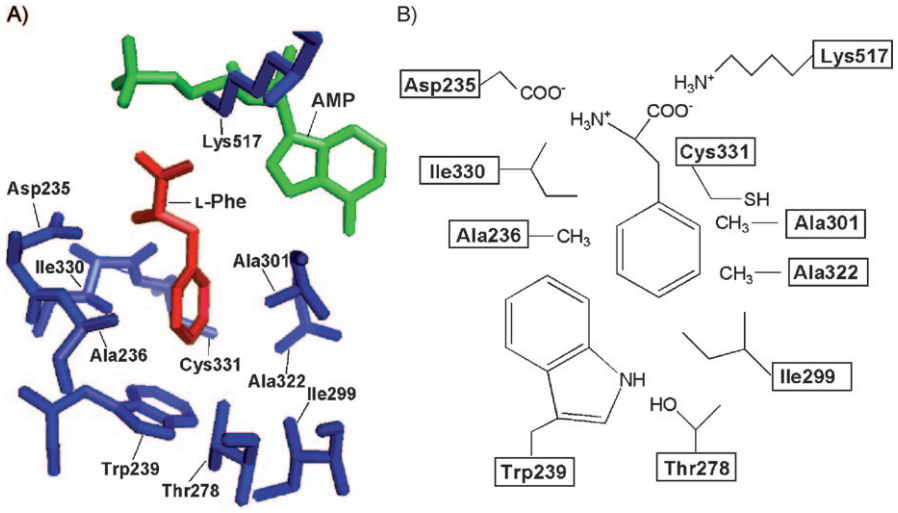


Figure 3. Substrate binding interactions in the GrsA A-domain. The structure of the GrsA A-domain[26] was used as a model of the TycA A-domain based on their high homology. Sequence alignments exhibit 79% amino acid sequence homology (or 64% amino acid identity, as determined by BLAST[36]) and indicate that they have identical residues in the binding pocket. The homology modeling of the TycA A-domain by analogy to the GrsA A-domain with SWISS-MODEL[37] showed a perfect superposition of the binding pocket residues between both structures. A) The 3D structure generated from PDB ID: 1AMU[26] using PyMOL (v0.99). The ten binding-pocket residues[11] are shown in blue, the natural amino acid substrate L-Phe (**1**) in red and the ATP breakdown product AMP in green. B) A 2D representation of the binding pocket with L-Phe (**1**). The labels follow the numbering in GrsA. In TycA Asp235 corresponds to Asp223, Ala236 to Ala224, Trp239 to Trp227, Thr278 to Thr266, Ile299 to Ile287, Ala301 to Ala289, Ala322 to Ala310, Ile330 to Ile318, Cys331 to Cys319 and Lys517 to Lys506.

identity, determined by BLAST[36]) with the TycA A-domain. Our results are broadly consistent with this previous study. As in GrsA,[27] L-Trp (**8**), L-Tyr (**6**) and L-Leu (**14**) were amongst the best proteinogenic substrates (Table 1). However, the extent to which L-Phe (**1**) was favored was much more pronounced for the TycA adenylation activity; k_{cat}/K_M is three orders of magnitude higher for L-Phe (**1**) than it is for the second-best proteinogenic amino acid substrate, L-Tyr (**6**), whereas in GrsA, this discrimination is only 13-fold over L-Trp (**8**), 16-fold over L-Leu (**14**) and 300-fold over L-Tyr (**6**).[27] Given that the active-site residues in TycA and GrsA are identical, the reason for this difference in discrimination is not clear but might be ascribed to distinct complex conformational dynamics. Alternatively, changes in the "second shell" amino acids close to the binding pocket might contribute to this change in specificity.

Diversity of alternative substrates

The collection of substrates studied encompasses the following groups in comparison to the native substrate:

1) L versus D amino acids: The TycA A-domain exhibits only a two- to fourfold stereoselectivity for L over D isomers, as measured by k_{cat}/K_M (Table S1). Similar k_{cat} values together with an increased K_M indicate that binding is slightly compromised but that a similar catalytic conformation can be reached, which leads to a nearly identical turnover number. The selectivity is much less pronounced than that observed for the downstream C-domain donor site, ProC.[38] The tolerance can be explained in terms of functional evolution by the lack of D-Phe (**2**) in cells but remains remarkable, as it indicates that the substrate has some flexibility in binding while still maintaining the proper alignment of the reactive centre. Indeed, the GrsA A-domain structure[26] had already indicated that L- and D-Phe can be accommodated with little or no change in protein conformation. Similarly, the k_{cat}/K_M values of L- and D-Tyr differ only by threefold, and again the difference is mainly due to an increase in K_M (although the parameters in this case are less reliable due to the lack of full saturation, see Figure S1P).
2) α versus β amino acid: The introduction of β homo amino acids into nonribosomal peptides would decrease their susceptibility to degradation by proteolysis and broaden their structural diversity.[39] We detected activity for one such β homo amino acid, L-β-homophenylalanine (**27**), although k_{cat}/K_M is dramatically decreased by 10^4-fold compared to that of L-Phe (**1**). Both kinetic parameters are affected by 100-fold, but relative to their respective dynamic range in this study, k_{cat} is much more affected than K_M. Despite the identity of the side chain (corresponding to a limited change in K_M), there appears to be little flexibility regarding the alignment of the ammonium and carboxylate binding contacts that position the substrate for catalysis between Asp223 and Lys506 (corresponding to Asp235 and Lys517, respectively, in GrsA numbering (Figure 3)).[11,26] Thus, the extension of the amino acid backbone has a much larger effect on k_{cat} than does the substitution of various side chains that differ more strongly in structure from the native substrate; only three substrates—**20**, **24** and **26** in Table 1—show lower k_{cat} values.
3) Modification at the amino terminus: To further probe the alignment of the substrate between Asp235 and Lys517, N-terminally modified substrates were tested. The introduction of *N*-methylation at the amino terminus increases the steric demands for the interaction between the substrate ammonium group and Asp235. *N*-methyl-L-Phe (**25**) and *N*,*N*-dimethyl-L-Phe (**26**) were accepted as substrates and showed decreases in catalytic efficiency of three and five orders of magnitude, respectively (Table 1). These results demonstrate that even electronically conservative and merely steric changes at the ammonium group are penalized and severely impair catalysis. This observation is also confirmed by the kinetic parameters obtained with the dipeptide Gly-L-Phe (**28**) (Table 1), where the amine function is sterically and electronically altered by acylation. The acylation of the amino group of L-Phe (**1**) causes a decrease of four orders of magnitude in catalytic efficiency. This 10^4-fold discrimination between the k_{cat}/K_M values for L-Phe (**1**) versus the dipeptide demonstrates stringent control of the length of the substrate (and subsequently of the final peptide product).
4) Amino acids that were not accepted as substrates: Finally, no activity was detectable for some amino acids bearing polar side chains (that is, the amide-containing L-Gln (**35**) and L-Asn (**36**)), and for all those with charged side chains (L-Glu (**37**), L-Asp (**38**), L-Arg (**39**) and L-Lys and its derivatives (**40**–**43**), Table 3), although the latter were reported to bind to GrsA.[27] Table 3 shows all the potential substrates tested for which no turnover could be detected, that is, k_{cat}/K_M was below the limit of 10^{-3} min^{-1} mM^{-1}, even at substrate concentrations approaching the solubility limit.

Discussion

The large number of substrates studied allows an attempt to identify the driving forces and the specificity determinants involved in amino acid recognition by the TycA A-domain. To this end, comparisons of k_{cat}/K_M values were used to evaluate the energy barrier starting from free enzyme and substrate, whereas comparisons of k_{cat} would refer to different starting complexes ($E{\cdot}S_1$ vs. $E{\cdot}S_2$). The different starting points explain why the range of k_{cat}/K_M values observed spans three orders of magnitude more than do the k_{cat} values. In many previous studies of A-domains, k_{cat}/K_M data are not directly available, making a direct comparison often impossible. A distinction between these parameters is important, as the quantitative assessment of promiscuity crucially hinges on which comparison is used.

The pattern that emerges from studies of alternative substrates indicates that this A-domain is relatively intolerant to tampering with the molecular recognition of the substrate ammonium and carboxylate groups by Asp235 and Lys517, respectively. Increasing the distance between these groups (in β

amino acids) or derivatising the amino group (by *N*-methylation or -acylation in dipeptides) leads to a substantial loss of catalysis. In contrast, the binding pocket can accommodate side chains of different sizes and electronic properties, although all charged and some polar substrate side chains show no activity. In this study, different series of substrates were tested, providing the basis for the identification of general trends contributing to substrate recognition. For example, in the four following series comparing closely isosteric side chains, the catalytic efficiency (k_{cat}/K_M) decreases with polarity (or increases with hydrophobicity):

$CH(CH_3)_2$ (**17**) > $CH(CH_3)OH$ (**22**)

CH_2CH_3 (**18**) > CH_2OH (**23**) ≈ CH_2SH (**24**) > $CH_2NH_3^+$ (**43**)

$(CH_2)_2CH_3$ (**15**) > $(CH_2)_2SH$ (**19**) > $(CH_2)_2NH_3^+$ (**42**)

$(CH_2)_3CH_3$ (**5**) > $(CH_2)_2SCH_3$ (**12**) > $(CH_2)_3NH_3^+$ (**41**).

Comparisons of substrates with different side chain lengths also show that catalytic efficiency (k_{cat}/K_M) increases with increasing side-chain length:

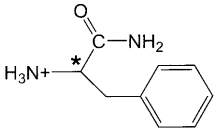
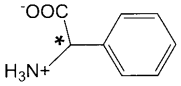
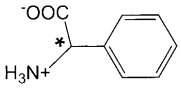
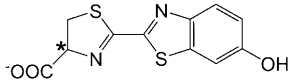
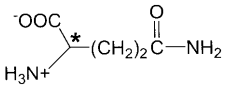
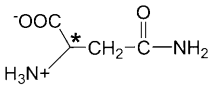
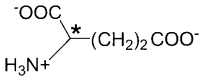
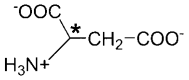
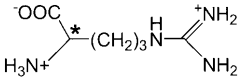
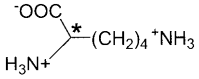
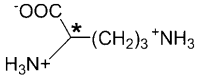
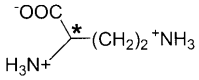
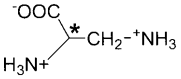
$(CH_2)_3CH_3$ (**5**) > $(CH_2)_2CH_3$ (**15**) > CH_2CH_3 (**18**) > CH_3 (**21**)

$CH(CH_3)CH_2CH_3$ (**10**) > $CH(CH_3)_2$ (**17**)

$(CH_2)_2SCH_3$ (**12**) > $(CH_2)_2SH$ (**19**) > CH_2SH (**24**).

For these compounds, in which the side chains are smaller than that of the natural substrate L-Phe (**1**), the discrimination between substrates must be due to well-defined interactions with the binding pocket. The various contributions to substrate discrimination are discussed below.

Table 3. Substrates[a] exhibiting no measurable activity under the conditions listed.

No.	Structure	*	[TycA] [μM]	t_R[b] [min]	[Substrate] tested [mM] Min	Max
31		L	0.3	60	2	50
32		L	0.1	20	0.96	24
33		D	0.1	20	0.96	24
34		D	1	60	2	50
35		L	1	120	9.2	230
36		L	1	120	6.2	155
37		L	1	120	0.2	5
38		L	0.1	30	72.8	1820
39		L	1	60	48	1200
40		L	0.1	30	50	1250
41		L	1	60	50	1250
42		L	1	60	50	1250
43		L	1	60	8.8	220

Conditions: $[MgCl_2]$ = 10 mM, [HEPES] = 50 mM, pH 8.0, [EDTA] = 1 mM, [NaCl] = 100 mM, $[^{32}P]PP_i$: 0.045 μCi per reaction, $[PP_i]$ = 0.1 mM, [ATP] = 2 mM. Note: the pH was unchanged after the addition of the substrate to the reaction mixture. [a] L-Phe amide (**31**), L-phenylglycine (**32**), D-phenylglycine (**33**), D-luciferin (**34**), L-Gln (**35**), L-Asn (**36**), L-Glu (**37**), L-Asp (**38**), L-Arg (**39**), L-Lys (**40**), L-ornithine (**41**), L-2,4-diaminobutyric acid (**42**), L-2,3-diaminopropionic acid (**43**). [b] Reaction time.

1) Electronic properties and hydrophobic effects: The lack of observed catalysis for charged substrates (Table 3, **37**–**43**) confirms the major role of the electronic character of the side chain and could indicate that interactions of these substrates with the A-domain are not strong enough to overcome the hydrogen-bonding of the substrate with solvent water as a prerequisite for binding in a hydrophobic pocket.[40] The energy of desolvation of a charged group could be so large that it precludes binding in a hydrophobic pocket. The native substrate (L-Phe (**1**)) in the GrsA A-domain crystal structure appears completely desolvated.[26] Alternatively, the protein could rearrange so that the solvated substrate could fit in the binding pocket,[41] but this possibility is unlikely given how hydrophobic the TycA A-domain is, and further ruled out by considerations discussed below.

Hydrophobic effects can profoundly influence the binding of ligands or substrates. Often, these binding events—based on the differential thermodynamics of solvation and desolvation of the substrate and the corresponding binding pocket, and van der Waals interactions—are

accompanied by induced-fit conformational rearrangements,[42] making them difficult to interpret or predict.[40] The approach described herein provides an empirical method that might be useful for the understanding of molecular recognition patterns based on hydrophobic interactions in A-domains but more generally in other enzymes whose substrate specificity is yet to be explored. Defining the TycA A-domain binding pocket in terms of hydrophobicity characterizes its molecular recognition profile and complements previous approaches, in which a steric fit or a specific hydrogen-bonding pattern has been used to explain specificity patterns.[11,12,26]

In quantitative structure-activity relationships,[43] the Hansch log *P* value (based on partitioning coefficients between water and n-octanol) is a useful hydrophobicity descriptor for correlating functional parameters ranging from toxicity[44] to inhibition[45] and catalysis.[46,47] To assess the hydrophobicity of substrate side chains, we calculated their log *P* values.[48] Side chains (rather than the whole substrate) were chosen to focus on the interactions between the variable part of the substrate structure and the binding pocket. Figure 4 correlates the hydrophobicities of the substrate L-amino acid side chains with the catalytic efficiency (Table 1). The entire data set shows a broad trend, namely that the hydrophobic substrates have higher second-order rate constants (k_{cat}/K_M). However, the correlation also exhibits considerable scatter, which can be partly resolved if subsets of the data are used to derive a more precise correlation. Linearity ($R = 0.97$) is observed for the log(k_{cat}/K_M) values plotted against the log *P* values calculated for aliphatic side chains (measured for eight substrates with side chains containing only C and H atoms). The linear correlation with a slope of 2.3 over four orders of magnitude of k_{cat}/K_M values suggests that at least for this subset of eight substrates, hydrophobicity is a major determinant of substrate discrimination. The accuracy of the correlation is remarkable, as the size of the side chain varies between one and seven carbon atoms for residues as different as L-Ala (**21**) and β-cyclohexyl-L-alanine (**3**). These substrates can be presumed to be relatively flexible, possibly allowing an induced-fit scenario, in which both the enzyme and substrate align to maximize van der Waals interactions (dispersion forces), making hydrophobic interactions a major driving force for the transfer of a hydrophobic substrate into the hydrophobic region of the protein. Hydrophobic substrates will partition into a hydrophobic pocket, because the exposure of a hydrophobic compound in an aqueous solvent results in the reorganization of the solvent hydrogen-bonding network and leads to a decrease in the entropy of the system (or an increase in the local order of water molecules). The value of the slope of 2.3 observed in Figure 4 indicates that the TycA A-domain binding pocket appears 2.3-fold more hydrophobic than *n*-octanol, that is, that the decrease in the activation energy is 2.3 times greater than the free energy of transfer of the alkyl groups from water to *n*-octanol. A twofold factor can be ascribed to a corresponding entropy increase in the solvent as two disfavoured hydrophobic-water interfaces—one around the substrate and one inside the enzyme binding pocket—are removed upon the formation of the enzyme-substrate complex.[41] A similar, linear correlation with a slope of 2.2 between amino acid side chain hydrophobicity (log *P*) and log(k_{cat}/K_M) has been observed for the chymotrypsin-catalyzed hydrolysis of a series of esters (*N*-acetyl-L-amino acid methyl esters R-CH(NHAc)CO_2CH_3, where R is an unbranched alkyl chain).[47] In both cases, the favourable entropy of desolvation leads to transfer of a hydrophobic substrate into the hydrophobic pocket of the catalyst.

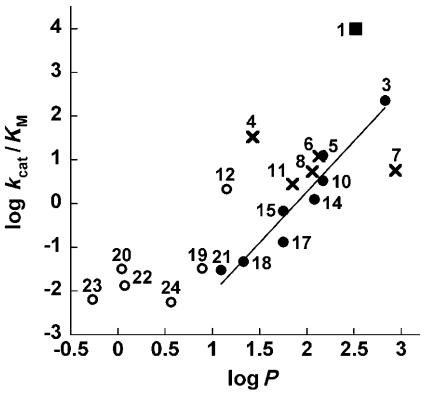


Figure 4. Hydrophobicity plays a key role in the substrate specificity of the TycA A-domain. A plot of log(k_{cat}/K_M), derived from the ATP/PP_i-exchange assays versus Hansch log *P* (calculated by using ChemBioDraw Ultra 11.0[48]) for the side chains of L-amino acid substrates, relates hydrophobicity and catalytic efficiency. The substrates were divided into four classes according to their side chain properties: natural substrate [■: L-Phe (**1**)], aliphatic side chains smaller than L-Phe (**1**) [●: β-cyclohexyl-L-alanine (**3**), L-norleucine (**5**), L-Ile (**10**), L-Leu (**14**), L-norvaline (**15**), L-Val (**17**), L-2-aminobutyric acid (**18**) and L-Ala (**21**)], heteroatom-containing aliphatic side chains smaller than L-Phe (**1**) [○: L-Met (**12**), L-homocysteine (**19**), L-His (**20**), L-Thr (**22**), L-Ser (**23**) and L-Cys (**24**)], and side chains bigger than L-Phe (**1**) [**X**: 4-amino-L-Phe (**4**), L-Tyr (**6**), L-homophenylalanine (**7**), L-Trp (**8**) and 3-nitro-L-Tyr (**11**)]. Amino acid substrates exhibiting substrate inhibition (**29**, **30**) are excluded. The numbers in the plot refer to the compound numbers listed in Table 1. The subset of the substrates that contain aliphatic side chains smaller than L-Phe (●) shows a linear correlation with a slope of 2.3. The correlation coefficient *R* of the corresponding fit is 0.97.

2) Shape complementarity and van der Waals interactions (dispersion forces): Deviations from the above-mentioned correlation (Figure 4) can be rationalized by invoking additional factors. Nonpolar interactions such as van der Waals interactions will contribute significantly to binding and catalysis when specific, close-packed interactions are involved, because the proximity of substrate and residues lining the binding pocket provides more opportunities for interaction. For example, the isoleucyl-tRNA synthetase binds valine, with a side chain lacking only one CH_3 group compared to isoleucine, 150 times more weakly because of the loss of the van der Waals interaction of this methylene group with the binding pocket.[49] In the case of the TycA A-domain, van der Waals interactions can explain the deviation of the natural substrate L-Phe (**1**) from the linear correlation observed with the hydrocarbon aliphatic side chains (Figure 4) and account for a gain of more than two orders of magnitude in catalytic efficiency. The stacking of the substrate phenyl ring between two walls of the binding pocket leads to a high degree of shape complementarity and the potential for productive van der Waals interactions.[26] Following this argument, the large deviation of L-Met (**12**), de-

spite its lower shape complementarity, can be ascribed to increased van der Waals interactions due to the high polarizability of the sulfur ether group.[41] Van der Waals interactions might also compensate for the low hydrophobicity of the heteroatom-containing substrates, L-Ser (**23**), L-Thr (**22**) and L-His (**20**) that deviate from the log *P* correlation (Figure 4), giving their side chain an overall quasi-neutral contribution to amino acid binding; the substrate without a side chain, Gly (**29**), shows similar activity (Table 2).

3) Steric exclusion: Substrates larger than the natural substrate are disfavoured, although they share most of the features of L-Phe (**1**). A pronounced decrease in catalytic efficiency is observed, even if only one heavy atom is added. For example the k_{cat}/K_M for L-Tyr (**6**) is 800-fold lower than that for L-Phe (**1**). A more hydrophilic substrate with a similar size, 4-amino-L-Phe (**4**), shows a similar activity to that of L-Tyr (**6**), indicating that substrate side-chain size can be predominant over hydrophobicity. This is consistently the case for all substrates larger than L-Phe (**1**), but not for those that are smaller. Thus, the observed discrimination is a consequence of steric exclusion. 3-Nitro-L-Tyr (**11**), a derivative of L-Tyr (**6**) with an additional nitro group in the *meta* position of the phenyl ring, also shows a very similar activity compared to L-Tyr (**6**), indicating that the steric exclusion is more stringent at the *para* position, and that steric fit (or shape complementarity) is looser along the plane of the L-Phe (**1**) phenyl ring. This is confirmed by the observation that L-Trp (**8**), the second-best natural substrate analogue of the TycA A-domain, shows an 1800-fold decrease in catalytic efficiency compared to that of L-Phe (**1**). This decrease is only twice as large as that observed for L-Tyr (**6**), despite a much larger side chain. For L-homophenylalanine (**7**), an L-Phe (**1**) derivative in which an additional methylene group has been inserted at the C_α, k_{cat}/K_M decreases by 1700-fold (that is, well below the "hydrophobic baseline"), although all the recognition elements of L-Phe (**1**) are present. In all cases, the steric exclusion of the larger side chains might upset the orientation of the ammonium and carboxylate centres leading to suboptimal positioning of the substrate for nucleophilc attack on ATP.

It is difficult to generalize these observations, because there are few or no examples where a set of k_{cat}/K_M values are available. At least one A-domain (of enniatin synthetase) exhibits similar characteristics: acceptance of various substrates but also evidence for size exclusion and lack of acceptance of charged substrates, although in this case, only product yields (roughly corresponding to k_{cat} data) were available.[50] It will be interesting to probe by measuring k_{cat}/K_M values whether these specificity features are more generally shared amongst other A-domains.

Specificity and selective pressure

The substrate acceptance exhibited by the TycA A-domain raises the question whether such a profile represents an enzyme under intense selection pressure or one that was not evolved to be specific. Aminoacyl-tRNA synthetases that catalyze the same chemical process (that is, the adenylation of amino acids) provide a system for comparison, despite the lack of sequence or structure homology.[26,51,52] Aminoacyl-tRNA synthetases have evolved under substantial selective pressure ensuring high fidelity in protein synthesis.[53] For example, the discrimination between valine and isoleucine by valyl-tRNA synthetase measured by the ATP/PP_i-exchange assay, is worth more than a 10^5-fold reduction in k_{cat}/K_M,[54] although the substrates differ only by one CH_3 group. For phenylalanyl-tRNA synthetase, the discrimination between phenylalanine and tyrosine is less pronounced, but still 1800-fold,[55] as a consequence of the addition of just one hydroxyl group. The latter example roughly matches the discrimination between L-Phe (**1**) and larger amino acids in the TycA A-domain. By contrast, the discrimination for unnatural amino acids with the same number of heavy atoms is much less pronounced; L-Phe (**1**) is only a 45-fold better substrate than β-cyclohexyl-L-alanine (**3**), and the binding of D-Phe (**2**) is only decreased by twofold compared to that of L-Phe (**1**). Likewise, unnatural L-norleucine (**5**) exhibits the same activity as L-Tyr (**6**) despite a structurally very different side chain. This weaker discrimination for unnatural amino acids is likely to be due to a lack of selective pressure and is mirrored in aminoacyl-tRNA synthetases, which activate a range of unnatural substrate analogues but discriminate stringently between natural substrates.[56,57]

Collectively, these considerations suggest that the discrimination between substrates by the TycA A-domain is comparable with that by the corresponding aminoacyl-tRNA synthetase (phenylalanyl-tRNA synthetase) at the level of the amino acid adenylation reaction and imply that there is high selective pressure on the adenylation activity specificity of TycA. Nonribosomal peptides are often not essential for growth and proliferation of the producer organism (for example, tyrocidine and gramicidin S),[58] and in some cases, the peptide sequence depends on the concentrations of the amino acids added to the growth media (e.g. for tyrocidine,[59] surfactin[60] and cyclosporine).[61] Some NRPS modules have been shown to be promiscuous in vivo, for example, in tyrocidine synthesis where Phe and Tyr residues at positions 3, 4 and 7 can be replaced by Trp.[59] These examples suggest low evolutionary pressure for accurate substrate recognition by some NRPS A-domains, in contrast to the apparently higher selective pressure on TycA for the specific activation of L-Phe (**1**) . The reason for this discrimination could lie in downstream processing; the thioesterase domain (TE-domain) that cyclizes the linear precursor of tyrocidine is intolerant to amino acid substitution at the first position of the peptide, which depends on TycA specificity.[62] In elongation modules, both the A-domain and the C-domain acceptor site are selectivity filters,[63] so an incorrect aminoacyl-*S*-phosphopantetheine(Ppant) intermediate is not processed further and eventually hydrolyzed by type II thioesterases (TEIIs) after an increased half-life.[64] However, in TycA, such an intermediate is processed almost normally, since the E-domain has a broad substrate tolerance[65] and the downstream C-domain donor site is stereoselective but not side chain selective.[63] As a consequence, the downstream proofreading of TycA substrate selectivity only occurs at the level of the TE-domain. The high energetic cost of the hydrolysis of an incorrect decapeptidyl-*S*-Ppant intermediate would then explain why there is substan-

tial selective pressure on the TycA A-domain to only activate amino acids that can be finally cyclized.

Conclusions

This work defines TycA as a surprisingly specific enzyme comparable in its amino acid adenylation selectivity to the corresponding phenylalanyl-tRNA synthetase, despite its ability to accept a large number of substrates in principle. The enzymes of secondary metabolism such as TycA have been considered more tolerant than those of primary metabolism.[66] Promiscuity in the biosynthetic pathways of secondary metabolites has been identified as an evolutionarily advantageous feature,[67] and many examples of pathway engineering (most notably in carotenoid biosynthesis[68]) exist. However, this work suggests that in the case of TycA, specificity (measured by k_{cat}/K_M) is substantial, despite an underlying level of promiscuity with a "baseline activity," for which the key prerequisite is hydrophobicity of the substrate side chain. For these substrates, the TycA A-domain can still show detectable, and thus evolvable, activity for a range of substrates. In this context, the potential of catalytic promiscuity based on hydrophobic interactions could be harvested for the directed evolution of A-domains that are the entry points for enzymatic NRPS assembly lines. The baseline level of relatively nonspecific activity that hydrophobicity confers might be exploited as a head start for further specialization, while already conferring a selectable advantage. However, the barrier that has to be overcome can still be considerable, highlighting the obstacles to combinatorial biosynthesis—where multiple enzymatic processing of alternative building blocks is required—even after successfully passing a gatekeeper domain such as TycA. Therefore, the engineering of A-domains for the recognition of alternative substrates would represent the first important step in the specificity change of a multistep biosynthetic machinery. This raises the question of what the best strategy for evolutionary improvement will be and where the most convenient entry points for functional adaptation of specificity are. These considerations are important given that directed evolution rarely delivers improvements of more than a few orders of magnitude. A comparison between natural substrates and some unnatural substrates seems to indicate that the selective pressure under which TycA evolved has lead to stronger discrimination against potential substrates available in the cell. If this observation can be generalized, the incorporation of the unnatural building blocks desired in combinatorial biosynthesis would present a smaller challenge than the use of natural amino acids. Strategically, small changes to the substrate structure like the gradual introduction of polar or charged functional groups, might allow an "adaptive walk"[69] that gradually approaches the desired target reaction. As in natural evolution aided by catalytic promiscuity, this strategy will make the maintenance of a detectable (and thus selectable) level of activity more likely and could render directed evolution approaches feasible.

Experimental Section

Chemicals: All chemicals were purchased from Sigma–Aldrich unless stated otherwise. β-cyclohexyl-L-alanine (**3**), D-Met (**13**), L-Phe amide (**31**), L-phenylglycine (**32**) and D-phenylglycine (**33**) were purchased from Novabiochem. D-Luciferin (**34**) and L-homocysteine (**19**) were purchased from Promega and Biosynth (Staad, Switzerland), respectively. Radiolabeled tetrasodium [^{32}P]pyrophosphate ([^{32}P]PP$_i$) was purchased from Perkin–Elmer.

Expression and purification of TycA: The production of His-tagged TycA (PheATE-His) was performed in *E. coli* BL21(DE3) bearing the vector pSU18-*tycA-PheATE*-His.[70] In a baffled flask (2 L), 2xYT medium (500 mL) containing chloramphenicol (20 μg mL^{-1}) was inoculated (1:100) with an overnight culture made from a fresh colony of the above strain and grown at 37 °C to $OD_{600} \approx 0.6$. IPTG was added (final concentration of 0.5 mM), and the culture was grown for another 4 h at 30 °C. The cells were pelleted (30 min, 4000 rpm, 4 °C), resuspended in binding buffer (0.5 M NaCl, 20 mM Tris-HCl, 5 mM imidazole, pH 7.8) and stored at −20 °C, if necessary. For cell lysis, DNaseI (NEB, 0.2 U mL^{-1}) was added, the cells were broken with an emulsifier, centrifuged (1 h, 18,000 rpm, 4 °C) and filtered (0.45 μm). The protein was affinity-purified on a HisTrap FF column on an ÄKTA chromatography system (GE Healthcare) with binding buffer with a linearly increasing imidazole concentration (5–250 mM). The pure protein was then dialyzed against the assay buffer (50 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 8.0) and stored in aliquots at −20 °C. Protein concentration was determined by the Bradford assay (M_W of TycA PheATE-His = 123 868.4 g mol^{-1}) and by measuring the absorbance at 280 nm (ε_{280} = 142 685 M^{-1} cm^{-1}).[71]

Kinetic measurements using the ATP/PP$_i$-exchange assay in a 96-well format: The reactions (60 μL) were carried out in the assay buffer (50 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 8.0) at the following final concentrations: 10 mM $MgCl_2$, 0.0007 μCi μL^{-1} [^{32}P]PP$_i$, 2 mM ATP, 0.1 mM $Na_4P_2O_7$, 0.1 or 1 μM TycA PheATE-His. A concentrated solution (3×, 20 μL) of enzyme (also containing $MgCl_2$ and [^{32}P]PP$_i$ in the assay buffer) was used to start the reaction with a multichannel pipette in microtubes (1.2 mL, Greiner Bio One) by adding it to a prereaction mixture (1.5×, 40 μL) containing the substrate, ATP and PP$_i$ in the assay buffer. The enzyme concentration, substrate concentration and reaction time were optimized for every substrate. The reaction was stopped at different time points by transferring the reaction mixture (60 μL) to Stop Mix (60 μL, 150 mM $Na_4P_2O_7$, 840 mM perchloric acid, 3.6% (*w*/*v*) charcoal) in a Black & White plate (Perkin–Elmer). Water (130 μL) was then added to each well, and the plate was centrifuged (3 min, 4000 rpm, 20 °C). The fluid containing the excess [^{32}P]PP$_i$ was removed by inverting the plate. After the charcoal was washed twice with H_2O (220 μL), it was resuspended in scintillation fluid (150 μL, Optiphase Supermix, Perkin–Elmer), and the radioactivity was measured after 48 h of incubation by reading in top mode for 1 min per well in a MicroBeta, 96-well plate, liquid scintillation counter (Perkin–Elmer). The amount of radiolabeled ATP formed (adsorbed to charcoal) was calculated from counts per minute (cpm) values, corrected for the counts measured in the absence of amino acid substrate. The Michaelis–Menten parameters K_M and k_{cat}, and the substrate inhibition parameters K'_M, k'_{cat} and K_{SI}[32] were calculated from the initial velocity of time curves with different substrate concentrations by a nonlinear least-squares fit to the Michaelis–Menten equation or the analogous formula taking account of substrate inhibition,[32] using Kaleidagraph (see the Supporting Information for a detailed calculation of k_{cat} from the radioactivity measurements). A second ATP binding site has been postulated for one A-domain of ACV synthe-

tase.[72] There is no evidence that this is the case for TycA. However, should there be a two-substrate reaction involving a second ATP binding site, then the K_M values calculated would be apparent constants (specific for the fixed concentration of MgATP).

Acknowledgements

This work was supported by the MRC and the EU networks ENDIRPRO and ProSA. B.R.M.V. is supported by the EU Early-stage Training Site ChemBioCam. F.H. is an ERC Starting Investigator. We thank Sijbren Otto for helpful discussions, Linda Otten and Michelle Schaffer for preliminary experiments and Torsten Stachelhaus, Dan Tawfik, Pat O'Brien, Tony Kirby, Stefanie Jonas, Linda Otten and other colleagues for comments on the manuscript.

Keywords: adenylation domain · biosynthesis · hydrophobic effect · nonribosomal peptide synthesis · substrate specificity

[1] J. Berdy, *J. Antibiot.* **2005**, *58*, 1–26; J. Clardy, M. A. Fischbach, C. T. Walsh, *Nat. Biotechnol.* **2006**, *24*, 1541–1550; G. L. Challis, D. A. Hopwood, *Curr. Opin. Biotechnol.* **2007**, *18*, 475–477.
[2] S. W. Haynes, G. L. Challis, *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 203–218; M. A. Fischbach, C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468–3496; S. Lautru, G. L. Challis, *Microbiology* **2004**, *150*, 1629–1636; U. Linne, S. Doekel, M. A. Marahiel, *Biochemistry* **2001**, *40*, 15824–15834; L. E. Quadri, *Mol. Microbiol.* **2000**, *37*, 1–12.
[3] H. G. Floss, *J. Biotechnol.* **2006**, *124*, 242–257.
[4] S. D. Copley, *Curr. Opin. Chem. Biol.* **2003**, *7*, 265–272.
[5] P. J. O'Brien, D. Herschlag, *Chem. Biol.* **1999**, *6*, R91–R105.
[6] U. T. Bornscheuer, R. J. Kazlauskas, *Angew. Chem.* **2004**, *116*, 6156–6165; *Angew. Chem. Int. Ed.* **2004**, *43*, 6032–6040; S. Jonas, F. Hollfelder in *The Handbook of Protein Engineering, Vol. I* (Eds.: U. T. Bornscheuer, S. Lutz), Wiley-VCH, Chichester, **2009**, pp. 47–72; K. Hult, P. Berglund, *Trends Biotechnol.* **2007**, *25*, 231–238.
[7] O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2006**, *10*, 498–508.
[8] M. E. Glasner, J. A. Gerlt, P. C. Babbitt, *Adv. Enzymol. Relat. Areas Mol. Biol.* **2007**, *75*, 193–239, xii–xiii.
[9] R. A. Jensen, *Annu. Rev. Microbiol.* **1976**, *30*, 409–425.
[10] P. J. O'Brien, D. Herschlag, *J. Am. Chem. Soc.* **1998**, *120*, 12369–12370.
[11] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493–505.
[12] G. L. Challis, J. Ravel, C. A. Townsend, *Chem. Biol.* **2000**, *7*, 211–224.
[13] S. G. Peisajovich, D. S. Tawfik, *Nat. Methods* **2007**, *4*, 991–994; S. Bershtein, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2008**, *12*, 151–158.
[14] K. T. Nguyen, D. Ritz, J. Q. Gu, D. Alexander, M. Chu, V. Miao, P. Brian, R. H. Baltz, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17462–17467.
[15] D. Butz, T. Schmiederer, B. Hadatsch, W. Wohlleben, T. Weber, R. D. Süssmuth, *ChemBioChem* **2008**, *9*, 1195–1200.
[16] H. D. Mootz, N. Kessler, U. Linne, K. Eppelmann, D. Schwarzer, M. A. Marahiel, *J. Am. Chem. Soc.* **2002**, *124*, 10980–10981.
[17] D. E. Cane, C. T. Walsh, C. Khosla, *Science* **1998**, *282*, 63–68.
[18] B. Shen, *Sci. STKE* **2004**, *2004*, pe14; J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380–416; M. A. Fischbach, C. T. Walsh, *Science* **2006**, *314*, 603–605; G. L. Challis, J. H. Naismith, *Curr. Opin. Struct. Biol.* **2004**, *14*, 748–756.
[19] R. H. Baltz, *Nat. Biotechnol.* **2006**, *24*, 1533–1540.
[20] M. Hans, A. Hornung, A. Dziarnowski, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2003**, *125*, 5366–5374; A. Tanovic, S. A. Samel, L. O. Essen, M. A. Marahiel, *Science* **2008**, *321*, 659–663; D. P. Frueh, H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh, G. Wagner, *Nature* **2008**, *454*, 903–906; S. Kapur, C. Khosla, *Nature* **2008**, *454*, 832–833.
[21] T. Stachelhaus, A. Schneider, M. A. Marahiel, *Science* **1995**, *269*, 69–72; A. Schneider, T. Stachelhaus, M. A. Marahiel, *Mol. Gen. Genet.* **1998**, *257*, 308–318.
[22] M. A. Fischbach, J. R. Lai, E. D. Roche, C. T. Walsh, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11951–11956.
[23] K. Eppelmann, T. Stachelhaus, M. A. Marahiel, *Biochemistry* **2002**, *41*, 9718–9726.
[24] S. G. Lee, F. Lipmann, *Meth. Enzymol.* **1975**, *43*, 585–602.
[25] M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem. Rev.* **1997**, *97*, 2651–2674.
[26] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997**, *16*, 4174–4183.
[27] L. Luo, M. D. Burkart, T. Stachelhaus, C. T. Walsh, *J. Am. Chem. Soc.* **2001**, *123*, 11208–11218.
[28] R. Calendar, P. Berg, *Biochemistry* **1966**, *5*, 1690–1695; R. K. Airas, *Eur. J. Biochem.* **1990**, *192*, 401–409.
[29] U. Linne, M. A. Marahiel, *Methods Enzymol.* **2004**, *388*, 293–315.
[30] L. M. Miller, M. T. Mazur, S. M. McLoughlin, N. L. Kelleher, *Protein Sci.* **2005**, *14*, 2702–2712.
[31] L. G. Otten, M. L. Schaffer, B. R. Villiers, T. Stachelhaus, F. Hollfelder, *Biotechnol. J.* **2007**, *2*, 232–240.
[32] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, 3rd ed., Portland, London, **2004**, pp. 113–144.
[33] R. Eisenthal, M. J. Danson, D. W. Hough, *Trends Biotechnol.* **2007**, *25*, 247–249.
[34] V. Bučević-Popović, M. Pavela-Vrančić, R. Dieckmann, H. von Döhren, *Biochimie* **2006**, *88*, 265–270; M. Pavela-Vrancic, R. Dieckmann, H. von Döhren, H. Kleinkauf, *Biochem. J.* **1999**, *342 Pt 3*, 715–719.
[35] L. Luo, C. T. Walsh, *Biochemistry* **2001**, *40*, 5329–5337.
[36] T. A. Tatusova, T. L. Madden, *FEMS Microbiol. Lett.* **1999**, *174*, 247–250.
[37] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, *Bioinformatics* **2006**, *22*, 195–201.
[38] D. E. Ehmann, J. W. Trauger, T. Stachelhaus, C. T. Walsh, *Chem. Biol.* **2000**, *7*, 765–772; T. Stachelhaus, C. T. Walsh, *Biochemistry* **2000**, *39*, 5775–5787.
[39] S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, *105*, 715–738; R. A. Lew, E. Boulos, K. M. Stewart, P. Perlmutter, M. F. Harte, S. Bond, S. B. Reeve, M. U. Norman, M. J. Lew, M. I. Aguilar, A. I. Smith, *FASEB J.* **2001**, *15*, 1664–1666; R. Galati, A. Verdina, G. Falasca, A. Chersi, *Z. Naturforsch. C* **2003**, *58*, 558–561.
[40] W. Blokzijl, J. B. F. N. Engberts, *Angew. Chem.* **1993**, *105*, 1610–1650; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1545–1579.
[41] A. Fersht in *Structure and Mechanism in Protein Science: Guide to Enzyme Catalysis and Protein Folding* 3rd ed., Freeman, New York, **1999**, pp. 324–348.
[42] A. Davis, S. J. Teague, *Angew. Chem.* **1999**, *111*, 778–792; *Angew. Chem. Int. Ed.* **1999**, *38*, 736–749.
[43] A. Z. Dudek, T. Arodz, J. Galvez, *Comb. Chem. High Throughput Screening* **2006**, *9*, 213–228.
[44] H. Kubinyi, *QSAR, Hansch Analysis and Related Approaches (Methods and Principles in Medicinal Chemistry)*, VCH, Weinheim, **1993**.
[45] C. Hansch, E. Coats, *J. Pharm. Sci.* **1970**, *59*, 731–743.
[46] M. D. Toney, J. F. Kirsch, *Science* **1989**, *243*, 1485–1488; B. C. Smith, J. M. Denu, *J. Biol. Chem.* **2007**, *282*, 37256–37265; L. C. James, D. S. Tawfik, *Protein Sci.* **2001**, *10*, 2600–2607.
[47] V. N. Dorovska, S. D. Varfolomeyev, N. F. Kazanskaya, A. A. Klyosov, K. Martinek, *FEBS Lett.* **1972**, *23*, 122–124.
[48] CambridgeSoft, ChemBioDraw Ultra 11.0, **2007**.
[49] R. B. Loftfield, E. A. Eigner, *Biochim. Biophys. Acta Gen. Subj.* **1966**, *130*, 426–448.
[50] S. C. Feifel, T. Schmiederer, T. Hornbogen, H. Berg, R. D. Süssmuth, R. Zocher, *ChemBioChem* **2007**, *8*, 1767–1770.
[51] O. Kotik-Kogan, N. Moor, D. Tworowski, M. Safro, *Structure* **2005**, *13*, 1799–1807.
[52] P. O'Donoghue, Z. Luthey-Schulten, *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 550–573.
[53] A. Fersht in *Structure and Mechanism in Protein Science: Guide to Enzyme Catalysis and Protein Folding* 3rd ed., Freeman, New York, **1999**, pp. 377–400.
[54] A. R. Fersht, C. Dingwall, *Biochemistry* **1979**, *18*, 2627–2631.
[55] S. X. Lin, M. Baltzinger, P. Remy, *Biochemistry* **1983**, *22*, 681–689.
[56] M. C. Hartman, K. Josephson, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4356–4361.

[57] M. C. Hartman, K. Josephson, C. W. Lin, J. W. Szostak, *PLoS ONE* **2007**, *2*, e972.
[58] N. Rajendran, M. A. Marahiel in *Comprehensive Natural Products Chemistry, Vol. 4: Amino Acids, Peptides, Porphyrins, and Alkaloids* (Vol. ed.: J. W. Kelly), Pergamon, Oxford, **1999**, pp. 195–220.
[59] H. D. Mootz, M. A. Marahiel, *J. Bacteriol.* **1997**, *179*, 6843–6850.
[60] G. Galli, F. Rodriguez, P. Cosmina, C. Pratesi, R. Nogarotto, F. de Ferra, G. Grandi, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1994**, *1205*, 19–28.
[61] A. Lawen, R. Traber, *J. Biol. Chem.* **1993**, *268*, 20452–20465.
[62] J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel, C. T. Walsh, *Nature* **2000**, *407*, 215–218; J. W. Trauger, R. M. Kohli, C. T. Walsh, *Biochemistry* **2001**, *40*, 7092–7098.
[63] P. J. Belshaw, C. T. Walsh, T. Stachelhaus, *Science* **1999**, *284*, 486–489.
[64] E. Yeh, R. M. Kohli, S. D. Bruner, C. T. Walsh, *ChemBioChem* **2004**, *5*, 1290–1293.
[65] D. B. Stein, U. Linne, M. A. Marahiel, *FEBS J.* **2005**, *272*, 4506–4520.
[66] R. D. Firn, C. G. Jones, *Nat. Prod. Rep.* **2003**, *20*, 382–391; R. D. Firn, C. G. Jones, *Mol. Microbiol.* **2000**, *37*, 989–994.
[67] M. A. Fischbach, J. Clardy, *Nat. Chem. Biol.* **2007**, *3*, 353–355.
[68] D. Umeno, A. V. Tobias, F. H. Arnold, *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 51–78.
[69] H. A. Orr, *Nat. Rev. Genet.* **2005**, *6*, 119–127; J. Maynard Smith, *Nature* **1970**, *225*, 563–564.
[70] S. Gruenewald, H. D. Mootz, P. Stehmeier, T. Stachelhaus, *Appl. Environ. Microbiol.* **2004**, *70*, 3282–3291.
[71] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch in *The Proteomics Protocols Handbook* (Ed.: J. M. Walker), Humana, Totowa, **2005**, pp. 571–607.
[72] W. Kallow, M. Pavela-Vrancic, R. Dieckmann, H. von Döhren, *Biochim. Biophys. Acta Proteins Proteomics* **2002**, *1601*, 93–99.

Received: August 18, 2008
Published online on February 2, 2009

