DOI: 10.1002/cbic.201000168

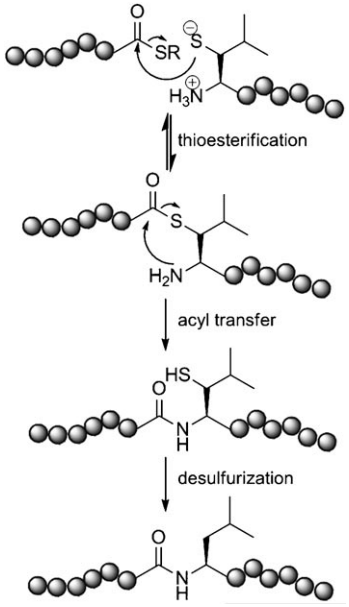
Protein Synthesis Assisted by Native Chemical Ligation at Leucine

Ziv Harpaz, Peter Siman, K. S. Ajish Kumar, and Ashraf Brik*[a]

Total chemical synthesis of proteins offers exceptional opportunities for preparing targets with exquisite control over the covalent structure, high purity and large quantities for functional and structural analysis.[1] In this regard, native chemical ligation (NCL) continues to be the method of choice for joining two unprotected peptides through thiol capture followed by S–N acyl transfer to form the amide bond at the ligation site.[2] The desulfurization reaction introduced by Yan and Dawson[3] as a post-NCL step greatly expanded the scope of ligation chemistry beyond Xaa-Cys (Xaa is any amino acid) by making ligation at Xaa-Ala sites accessible in the synthesis of functional proteins.[4] This seminal work has prompted several groups to extend this strategy to other β- and γ-mercapto amino acid derivatives (e.g., Phe and Val) as part of the ongoing efforts to extend NCL to essentially any ligation junction.[5] In addition, by positioning a thiol handle at the Lys side chain and at the 2-acetamido group of the glycan moiety in a glycopeptide allowed NCL to be used for peptide ubiquitylation[6] and glycopeptide ligation[7] respectively, and after a desulfurization step, the native structures could be achieved.

The above-described examples testify to the power of desulfurization when combined with NCL to assist the synthesis of naturally occurring proteins. In addition, recent reports on the mild and highly versatile free-radical Cys reduction protocol,[5a,b,8] as well as the compatibility of these methods in the presence of other thiol functionalities[9] permit the use of this two-step approach in the synthesis of a variety of protein targets. Despite the introduction of mercaptoPhe and mercaptoVal to assist efficient peptide ligation,[5] their utility in protein synthesis has not yet been demonstrated. The preparation of proteins by using ligation methods is often challenging because decreases in rate and chemoselectivity only come to light in more complex systems. To date, only ligation at Xaa-Cys followed by desulfurization has been demonstrated in the synthesis of full-length proteins,[4] this diminishes the generality of the desulfurization approach in the total chemical synthesis of proteins. Here we report an innovative strategy for ligation at Xaa-Leu sites by using β-mercaptoleucine combined with desulfurization (Scheme 1) and its application in the total chemical synthesis of HIV-1 Tat protein.

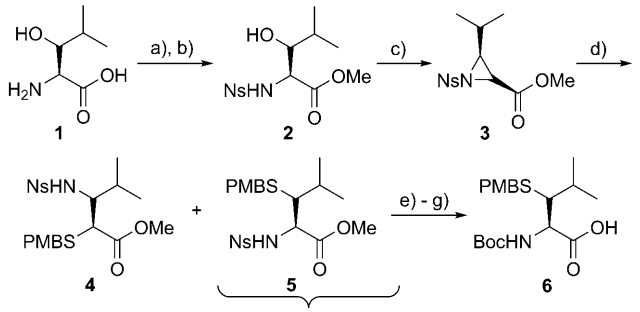
To implement the approach outlined in Scheme 1, we first had to design a synthetic strategy for β-mercaptoleucine, the key residue in this strategy. We envisioned two routes to achieve the synthesis of this building block by starting from the commercially available *threo*-β-hydroxy-L-leucine. In the first route, the β-hydroxy group could be converted to a good leaving group followed by nucleophilic substitution with a thiol nucleophile, whereas in the second one an aziridine intermediate could be employed to introduce the thiol functionality through a regioselective ring opening. Applying the first strategy by activation of the β-hydroxy group as a mesylate or tosylate derivative, then substitution with PMB-SH or thioacetic acid as a source of the thiolate functionality led to formation of the β-elimination product in a substantial amount (>60%). Additionally, it is very likely that under these conditions the thiol-substituted product could also be formed through a Michael addition, which would lead to a racemization at the α-carbon, rendering this method less desirable for us. Inspired by the recent work on the racemization-free synthesis of β-methylcysteine by starting from Thr,[10] we decided to follow a similar path to prepare the desired building block (Scheme 2). Thus, β-hydroxy-L-leucine was converted to the methyl ester derivative followed by treatment with *p*-nitrophenylsulfonyl chloride (NsCl) to give the *p*-nitrosulfonamide **2** in 70% yield (two steps). Ring closure of **2** under Mitsunobu conditions afforded the aziridine **3** in 96% yield. Subsequently, $BF_3{\cdot}OEt_2$-mediated ring opening with PMB-SH furnished the two regioisomers **4** and **5** in a quantitative yield in a 6:4 ratio respectively. Apparently, the additional steric hindrance in our case compared to Thr (isopropyl vs. methyl) decreased the desired regioselectivity. Notwithstanding the unsatisfactory regioselectivity of the aziridine opening, the regioisomers **4** and **5** were



Scheme 1. General strategy for NCL at Leu.

[a] *Z. Harpaz, P. Siman, Dr. K. S. A. Kumar, Dr. A. Brik*
Department of Chemistry, Ben-Gurion University of the Negev
Beer Sheva 84105 (Israel)
Fax: (+972) 8-6472943
E-mail: abrik@bgu.ac.il

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201000168.

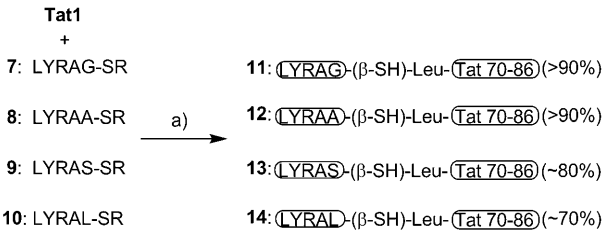


Scheme 2. Synthesis of β-mercaptoleucine **6**; Reagents and conditions: a) $SOCl_2$, MeOH, reflux; b) NsCl, Et_3N, CH_2Cl_2, RT, 70% over two steps; c) PPh_3, diisopropylazodicarboxylate, THF, 0 °C, 96%; d) *p*-$MeOC_6H_4CH_2SH$, $BF_3{\cdot}OEt_2$, CH_2Cl_2, 0 °C, quantitative yield with 6:4 ratio of **4** and **5**; e) *p*-$MeOC_6H_4SH$, K_2CO_3, RT, MeCN/DMSO; f) $(Boc)_2O$, iPr_2EtN, CH_2Cl_2, 0 °C; g) 1 M LiOH, THF/H_2O, 0 °C, 40% over three steps.

straightforwardly separated by using flash chromatography and analyzed by NMR spectroscopy for structure determination (Supporting Information); this allowed us to continue our study of examining the efficiency of ligation at Leu and achieving the total synthesis of HIV-1 Tat. The regioselectivity of the aziridine-ring-opening reaction is being optimized in our laboratory. Compound **5** was then transformed to the desired product through a three-step process, which included the replacement of the Ns with the benzyloxycarbonyl (Boc) derivative and hydrolysis of the methyl ester to give the protected β-mercaptoleucine **6** in 40% yield (for three steps), ready for SPPS. Notably, The *p*-methoxybenzyl (PMB) group is compatible with Boc- and Fmoc-SPPS. Moreover, its stability to the cleavage conditions in the Fmoc-SPPS should allow its use in sequential ligation employing $Hg(OAc)_2$ for its removal.

At this stage, we started the synthesis of a model peptide containing β-mercaptoleucine to examine the scope of the ligation junction at Xaa-Leu and its applicability in the synthesis of HIV-1 Tat protein. The C-terminal peptide of HIV-1 Tat(69–86), **Tat1**, was chosen to present the β-mercaptoleucine for the ligation study with thioester peptides **7**–**10** bearing Gly, Ala, Ser, and Leu at the C terminus, respectively. These thioester peptides represent the four groups that divide the 20 amino acids based on their ligation rates with Cys (Scheme 3).[11]

Gratifyingly, HPLC coupled with MS analysis revealed that all ligation reactions were successful and gave the ligation products **11**–**14** in 70–90% yield (Figure 1A and in the Supporting



Scheme 3. Model thioester peptides used in our ligation study with **Tat1** and the corresponding ligation products. a) 6 M Gn·HCl, pH 7, 37 °C, 2% thiol additives (*v*/*v*).

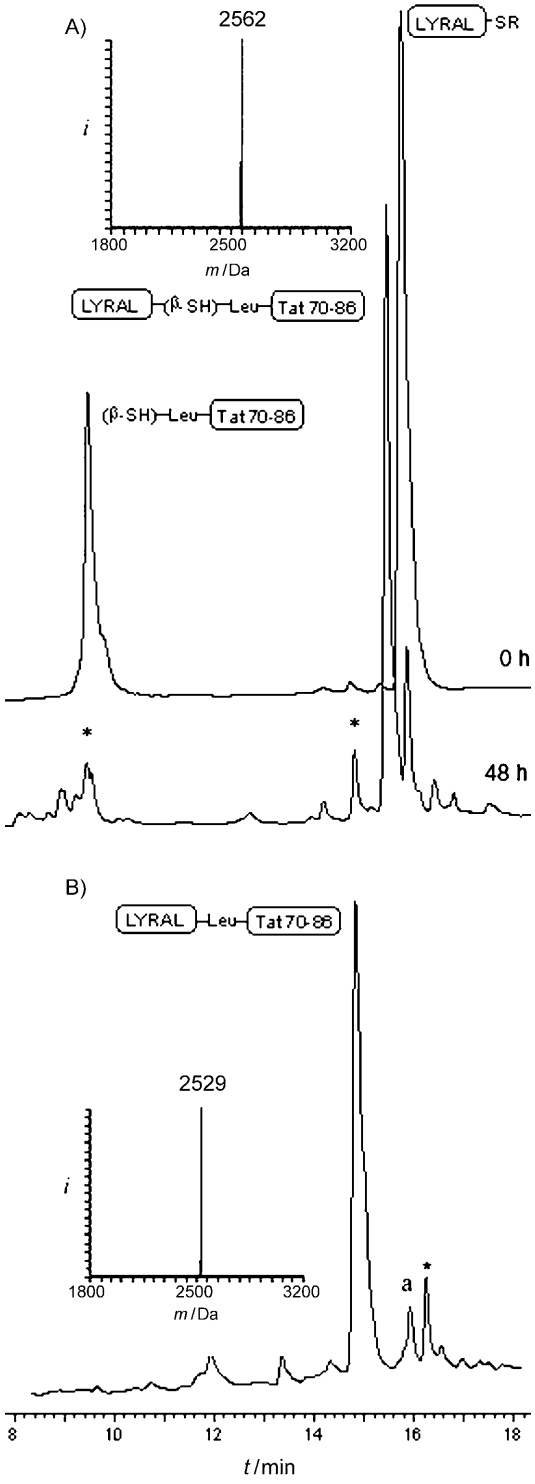
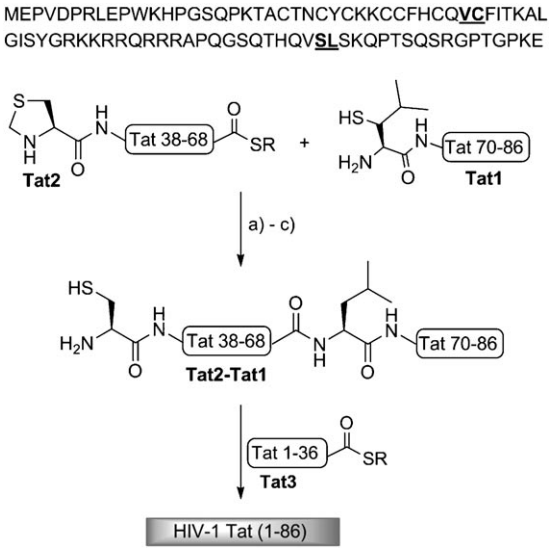


Figure 1. Model study of NCL assisted by Leu: A) Analytical HPLC and mass spectrometry analysis monitoring the ligation reaction between **Tat1** and LYRAL-SR (**10**), representing the most difficult ligation junction in our study, to give, after 48 h, the ligation product **14** with the observed mass 2562 Da (calcd *m*/*z*: 2561 Da). B) Analytical HPLC and mass spectrometry analysis of the desulfurization reaction of peptide **14** to afford the desulfurized product, after 8 h, with the observed mass of 2529 Da (calcd *m*/*z*: 2530 Da). Peak * corresponds to unidentified masses and peak a, is the unreacted peptide **14**.

Information). As expected, ligation rates were slower when compared with Xaa-Cys due to the additional steric bulkiness at the β-mercaptoleucine. In our study, the Gly thioester **7** reacted the fastest, and the Leu thioester **10** was the slowest. The ligation products were then subjected to metal-free desulfurization conditions to give the desulfurized products in excellent yields (Figure 1 B in the Supporting Information).

To examine the utility of NCL at Xaa-Leu in protein synthesis, we focused on the total synthesis of HIV-1 Tat. This protein plays a regulatory role in the life cycle of the HIV-1 virus by activating HIV transcription through complex interactions with RNA and host cell factors.[12] Moreover, HIV-1 Tat is known to undergo multiple post-translational modifications that regulate the dynamics and complexity of these interactions. Chemical synthesis of HIV-1 Tat and its analogues should assist the ongoing efforts to understand the role of these post-translational modifications on its function, in particular to the studies aiming to have a combinatorial view of these modifications.[12] Previous attempts by our group to synthesize this protein by using side-chain-assisted ligation were unsuccessful due to an inefficient removal of the auxiliary.[13] Our current synthetic strategy for the synthesis of HIV-1 Tat, assisted by Leu, is depicted in Scheme 4. Despite the presence of several Cys residues in the protein sequence, their close location to the N ter-



Scheme 4. Sequence and synthetic strategy of HIV-1 Tat (the ligation junctions are highlighted). a) Leu-assisted NCL, b) desulfurization, c) Thz→Cys.

minus limits the use of NCL at Xaa-Cys for synthesis. Moreover, this adds additional synthetic challenges of maintaining these residues intact, if a desulfurization step is required. In this case the use of a protecting group, such as Acm, would be necessary to avoid the desulfurization of the native thiol groups. On the other hand, the presence of Leu69 should facilitate the ligation between **Tat1** and HIV-1 Tat(37–68) thioester, **Tat2**.

The three precursor peptides, **Tat1**–**3**, were prepared by using Fmoc-SPPS. The syntheses of the thioester fragments, **Tat2** and HIV-1 Tat(1–36), **Tat3**, were carried out by using the

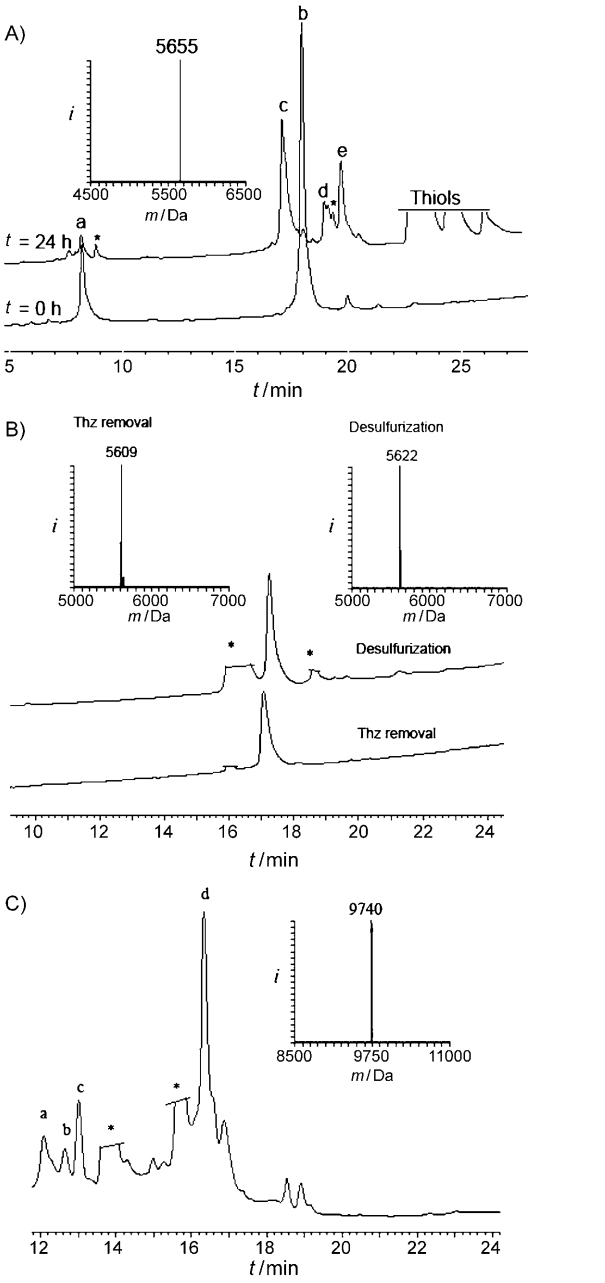


Figure 2. Synthesis of HIV-1 Tat assisted by Leu: A) Analytical HPLC monitoring the ligation reaction between **Tat1** and **Tat2** (1.2 equiv). Reaction at 0 h: Peak a corresponds to **Tat1** with the observed mass 1944 Da (calcd *m/z*: 1944 Da); Peak b corresponds to **Tat2** with the observed mass 3830 Da (calcd *m/z*: 3830 Da). Reaction at 24 h: Peak c corresponds to the ligation product **Tat2**–**Tat1** with the observed mass 5655 Da (calcd *m/z*: 5654 Da; see inset, deconvoluted ESI-MS), in addition to a slight contamination of the hydrolyzed thioester with the observed mass 3727 Da (calcd *m/z*: 3728 Da). Peak d corresponds to the thiophenol exchange intermediate of **Tat2** with the observed mass 3820 Da (calcd *m/z*: 3820 Da); Peak e corresponds to benzyl mercaptan exchange intermediate of **Tat2** with the observed mass 3833 Da (calcd *m/z*: 3834 Da). Peak * corresponds to unidentified masses. B) Analytical HPLC trace and mass spectrometry analysis of desulfurization (calcd *m/z*: 5622.1 Da) and Thz to Cys conversion (calcd *m/z*: 5609 Da) of **Tat2**–**Tat1**, peak * corresponds to desulfurzation reagents. C) Analytical HPLC trace and mass spectrometry analysis of ligation reaction of **Tat2**–**Tat1** and **Tat3** at 48 h. Peaks a, b, and c correspond to cyclization of **Tat3** to form cyclized thioester with the mass of 4130 Da, whereas peak d corresponds to the ligation product HIV-1 Tat(1–86) with the observed mass of 9740 Da (calcd *m/z*: 9740 Da). Peak * corresponds to thiol additives.

N-acyl-urea-based chemistry developed recently by Dawson (30–35% isolated yield) and co-workers (Supporting Information).[14] The ligation between **Tat1** and **Tat2** peptides was performed in a similar manner to our model study, affording the desired product **Tat2**–**Tat1** in 40% isolated yield (Figure 2A). The ligation product was then subjected to the desulfurization conditions in which a full conversion to the desired product was observed within 7 h (Figure 2B). The Thz group (1,3-thiazolidine-4-carboxo) resisted desulfurization under these conditions, and the absence of other Cys residues in **Tat1** and **Tat2** avoided the necessity for protection of SH side chains. Subsequently, the N-terminal Thz was fully converted to Cys within 5 h by using methoxylamine (Figure 2B).[4a] Final Cys-NCL step with the N-terminal fragment **Tat3**, furnished the full-length wild-type HIV-1 Tat (Figure 2C). It should be noticed that the final ligation involves a thioester (**Tat3**)-bearing Val residue at the C terminus, which is known to slow the ligation rate.[11] Furthermore, the presence of seven Cys residues at this stage led to the cyclized thioester intermediates that are less reactive than the aryl thioester, which further decreased the efficiency of the synthesis.

In summary, we have demonstrated that NCL can be extended to Xaa-Leu sites. Notably, Leu is the most abundant amino acid with approximately 9.8% frequency in nature.[15] The new ligation chemistry in combination with desulfurization enabled the first chemical synthesis of the HIV-1 Tat protein. Our synthetic strategy should allow the synthesis of various analogues of post-translationally modified HIV-1 Tat for structural and functional analysis.[12] The ligation at Xaa-Leu site expands the repertoire of the available ligation junctions, thus adding flexibility to the synthetic scheme when attempting total chemical synthesis of proteins.

Acknowledgements

This work was supported by the Israel Science Foundation.

Keywords: desulfurization · leucine · native chemical ligation · peptides · protein synthesis

[1] a) C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem.* **2008**, *120*, 10182–10228; *Angew. Chem. Int. Ed.* **2008**, *47*, 10030–10074; b) S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338–351.
[2] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776–779.
[3] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526–533.
[4] a) D. Bang, G. I. Makhatadze, V. Tereshko, A. A. Kossiakoff, S. B. H. Kent, *Angew. Chem.* **2005**, *117*, 3920–3924; *Angew. Chem. Int. Ed.* **2005**, *44*, 3852–3856; b) R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder, T. W. Muir, *Nature* **2008**, *453*, 812–816.
[5] a) C. Haase, H. Rohde, O. Seitz, *Angew. Chem.* **2008**, *120*, 6912–6915; *Angew. Chem. Int. Ed.* **2008**, *47*, 6807–6810; b) J. Chen, Q. Wan, Y. Yuan, J. Zhu, S. J. Danishefsky, *Angew. Chem.* **2008**, *120*, 8649–8652; *Angew. Chem. Int. Ed.* **2008**, *47*, 8521–8524; c) D. Crich, A. Banerjee, *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065.
[6] a) K. S. Ajish Kumar, M. Haj-Yahya, D. Olschewski, H. A. Lashuel, A. Brik, *Angew. Chem.* **2009**, *121*, 8234–8238; *Angew. Chem. Int. Ed.* **2009**, *48*, 8090–8094; b) R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *J. Am. Chem. Soc.* **2009**, *131*, 13592–13593.
[7] a) A. Brik, Y.-Y. Yang, S. Ficht, C.-H. Wong, *J. Am. Chem. Soc.* **2006**, *128*, 5626–5627; b) A. Brik, C.-H. Wong, *Chem. Eur. J.* **2007**, *13*, 5670–5675.
[8] Q. Wan, S. J. Danishefsky, *Angew. Chem.* **2007**, *119*, 9408–9412; *Angew. Chem. Int. Ed.* **2007**, *46*, 9248–9252.
[9] a) B. L. Pentelute, S. B. H. Kent, *Org. Lett.* **2007**, *9*, 687–690; b) Y. Y. Yang, S. Ficht, A. Brik, C. H. Wong, *J. Am. Chem. Soc.* **2007**, *129*, 7690–7701.
[10] R. S. Narayan, M. S. van Nieuwenhze, *Org. Lett.* **2005**, *7*, 2655–2658.
[11] T. M. Hackeng, J. H. Griffin, P. E. Dawson, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068–10073.
[12] C. Hetzer, W. Dormeyer, M. Schnölzer, M. Ott, *Microbes & Infection* **2005**, *7*, 1364–1369; Infection **2005**, 7, 1364–1369.
[13] a) M. Y. Lutsky, N. Nepomniaschiy, A. Brik, *Chem. Commun.* **2008**, *10*, 1229–1231; b) K. S. Ajish Kumar, M. Haj-Yahya, Z. Harpaz, A. Brik, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3870–3874.
[14] J. B. Blanco-Canosa, P. E. Dawson, *Angew. Chem.* **2008**, *120*, 6957–6961; *Angew. Chem. Int. Ed.* ***2008***, *47*, 6851–6855.
[15] Based on the latest released statistics by the European Bioinformatics Institute: http://www.ebi.ac.uk/uniprot/TrEMBLstats/.

Received: March 12, 2010
Published online on April 30, 2010

