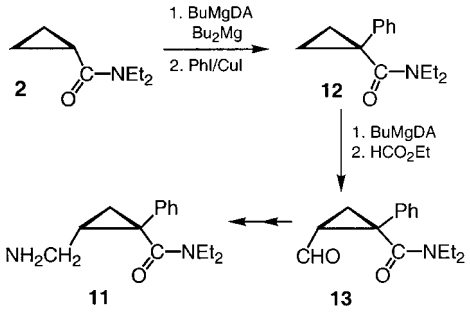
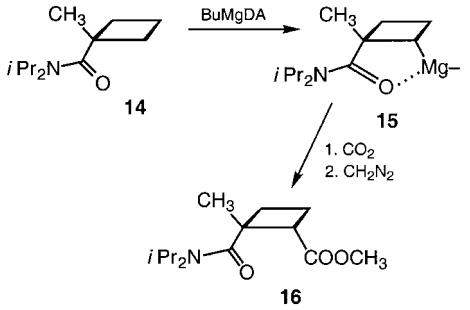


Scheme 5. A new approach to the synthesis of Milnacipran (**11**).

ethyl formate to give **13** (73% yield of isolated product), a known Milnacipran precursor.[12]

Aromatic and benzyl CH groups are more acidic than those of cyclopropanes and cubanes, so it is not surprising that BuMgDA stoichiometrically metalates an amide-activated CH group appropriately positioned thereupon. What if the CH group is less acidic? Cyclobutane CH is kinetically 2500 times more weakly acidic than cyclopropane CH.[13] Under conditions as harsh as most base/solvent combinations can withstand (e.g. LiTMP/THF or LiTMP + $Hg(TMP)_2$/THF at 0° C (TMP = 2,2,6,6-tetramethylpiperidine) or LDA/THF at reflux), there is no observed metalation of **14**. However, under the conditions of our preliminary trials, reaction of **14** with excess BuMgDA in THF at reflux for 5 h gave about 20% conversion[14] to the *cis*-β-metalated compound **15**, identified by formation of ester **16**[10] (17% yield of isolated product, Scheme 6). Although still to be optimized, this β deprotonation/metalation of a cyclobutane amide is extraordinary and unprecedented.



Scheme 6. Deprotonation/magnesiation of an amide-activated cyclobutane using BuMgDA.

Received: January 16, 2002
Revised: April 11, 2002 [Z18529]

[1] a) P. E. Eaton, C.-H. Lee, Y. Xiong, *J. Am. Chem. Soc.* **1989**, *111*, 8016–8018; b) P. E. Eaton, K. A. Lukin, *J. Am. Chem. Soc.* **1993**, *115*, 11370–11375; c) For a review: K. W. Henderson, W. J. Kerr, *Chem. Eur. J.* **2001**, *7*, 3430–3437, and references therein; d) Cf. Y. Kondo, M. Shilai, M. Uchiyama, T. Sakamoto, *J. Am. Chem. Soc.* **1999**, *121*, 3539–3540.
[2] Z. Hantosi, PhD Thesis, The University of Chicago, **1999**.
[3] a) As far as we can determine, amido-Grignards have not been examined elsewhere for their deprotonation/metalation abilities. b) We ignore, for the present, consideration of Schlenk equilibria and/or aggregation of magnesiated species.
[4] Aldrich Chemical sells a mixture of *n*-butyl and *s*-butyl compounds; this generates a mixture of *n*BuMgDA/*s*BuMgDA; this was what we used. The pure compounds behave similarly as bases.
[5] See: V. Snieckus, *Chem. Rev.* **1990**, *90*, 879–933.
[6] The magnesiated products have useful (but reduced) Grignard-like reactivity; cf. ref. [1].
[7] The addition of TMEDA does not improve the proton removal ability of Bu_2Mg.
[8] a) For example: H. W. Pinnick, Y.-H.Chang, S. C. Foster, M. Govindan, *J. Org. Chem.* **1980**, *45*, 4505–4507.
[] b) The α-lithio derivative of **2**, prepared by reaction of **2** with *n*BuLi at −78 °C, is stable only below 0 °C.
[9] A small excess of Bu_2Mg was used to remove adventitious water.
[10] Stereochemical assignments: aq. 30% H_2SO_4 amide hydrolysis (no isomerization; demonstrated by re-amidation) followed (e.g. for **9**) by NMR comparison to the known acid or (e.g. for **10**) by X-ray analysis; for **6** by X-ray analysis of the crystalline dicyclohexylamide analogue; for Scheme 4, ^1H and ^{13}C NMR; for **16**, X-ray single crystal analysis. CCDC-173837 (**10**), CCDC-173837 (**6**), CCDC-173571 (**16**) contain the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/conts/retrieving.html (or from the Cambridge Crystallographic Data Centre, 12, Union Road, Cambridge CB2 1EZ, UK; fax: (+44) 1223-336-033; or deposit @ccdc.cam.ac.uk).
[11] Prolonged exposure of the products to the generating conditions also leads to isomerization.
[12] S. Shuto, S. Ono, H. Imoto, K. Yoshii, A. Matsuda, *J. Med. Chem.* **1998**, *41*, 3507–3514, and references therein.
[13] A. Streitwieser, Jr., R. A. Caldwell, W. R. Young, *J. Am. Chem. Soc.* **1969**, *91*, 529.
[14] By NMR: 20% **16**, 80% unreacted starting material.

New Methods for Proteomic Research: Preparation of Proteins with N-Terminal Cysteines for Labeling and Conjugation**

Thomas J. Tolbert and Chi-Huey Wong*

In memory of Sun Fong

Proteins with N-terminal cysteines are useful in a wide range of biotechnological applications ranging from protein semi-synthesis to site-specific N-terminal labeling. Peptides and proteins with N-terminal cysteines undergo native chemical ligation and expressed protein ligation reactions with thioesters to form native peptide bonds.[1, 2] These reactions have been used to extend the size of proteins that can be synthesized chemically and to incorporate synthetic peptides with modifications and labels into expressed proteins.[3–8] In addition, proteins with N-terminal cysteines also react chemoselectively with aldehydes to form thiazolidines, and this reaction has been utilized to label and immobilize peptides and proteins.[9–12] Here we present a novel method (Figure 1) to produce proteins with N-terminal cysteines by

[*] Prof. C.-H. Wong, Dr. T. J. Tolbert
The Department of Chemistry and Skaggs Institute for Chemical Biology
The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
Fax: (+1) 858-784-2409
E-mail: wong@scripps.edu

[**] This research was supported by the NIH (R37 GM44154).
Supporting information for this article is available on the WWW under http://www.angewandte.com or from the author.

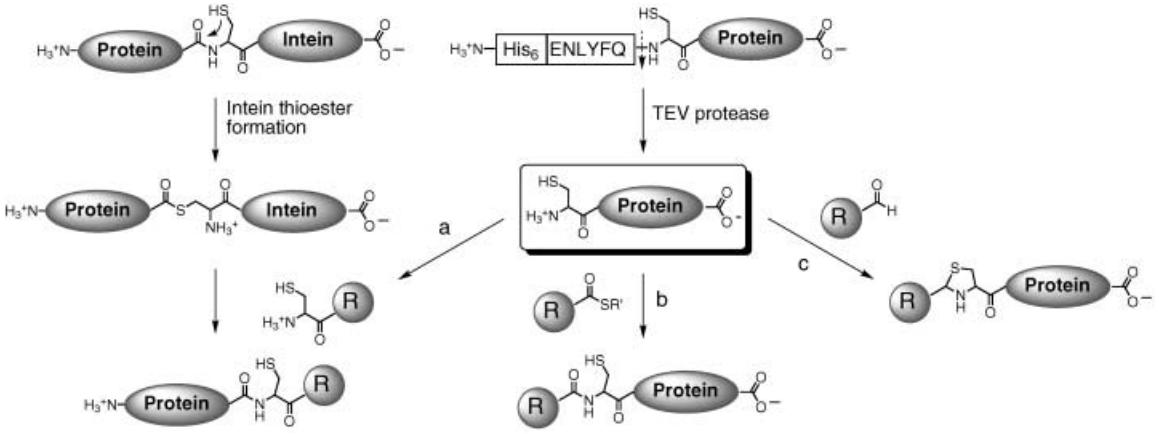


Figure 1. Generation of proteins with N-terminal cysteines using TEV protease and some examples of uses for proteins with N-terminal cysteines: a) expressed protein ligation reactions with intein generated thioesters; b) native/expressed protein ligation reactions with synthetic peptides and labels; c) chemoselective ligation with aldehydes to form thiazolidines.

using the highly selective tobacco etch virus NIa protease (TEV protease). We also demonstrate the subsequent labeling of proteins with N-terminal cysteines produced in this way using thioesters of two common protein labeling reagents. We believe that the use of TEV protease to generate N-terminal cysteines overcomes some of the limitations of existing methods, and this should make it generally useful for chemoselective ligation, proteomic labeling, and other biotechnological applications.

Current methods for producing proteins with N-terminal cysteines from affinity-tagged fusion proteins include using protein expression systems that utilize the self-cleaving activity of inteins[13] and selective proteolysis.[4, 5] Intein expression systems utilize the protein-splicing activity of inteins to catalyze the intramolecular cleavage of the intein from its fusion partner.[14] The difficulty in using the intein cleavage reaction to produce proteins from affinity-tagged intein fusion proteins is that it may cause in vivo cleavage during expression and complicate the preparation process. Many factors can affect in vivo and in vitro cleavage such as the expression conditions, structure of the desired protein, and identity of the N-terminal amino acids at the cleavage junction.[14] To obtain a construct that expresses well, screening of several expression conditions and multiple expression constructs with different amino acids at the cleavage junction is frequently required. Selective proteolysis of affinity-tagged fusion proteins to produce N-terminal cysteines generally does not suffer from expression problems since the affinity tags do not cleave themselves in vivo and can be chosen to optimize expression yield and protein folding. Unfortunately, the most common protease to be used to generate N-terminal cysteines, factor Xa,[4, 5] is not completely sequence-specific, sometimes cleaving peptide bonds after positively charged residues in the absence of its primary recognition sequence.[15, 16] This often requires careful monitoring of factor Xa cleavage reactions to prevent internal cleavage of the desired protein. For many protein applications a more selective protease than factor Xa with a lower probability of cleaving internally during affinity tag removal would be advantageous.

We decided to investigate the use of TEV protease to generate N-terminal cysteines from affinity-tagged fusion proteins because it has several advantages for labeling applications. It is a highly selective cysteine protease with a seven amino acid recognition site, recognizing six amino acids N-terminal to the scissile bond between the P1 glutamine and the P1′ serine or glycine (Figure 2a).[17, 18] Another advantage of TEV protease is that it can be expressed in active form in *E. coli* and yeast without affecting cell viability,[19] making production of TEV protease for large-scale applications inexpensive.[20] Additionally, TEV protease has already proven useful in proteomic applications, as demonstrated by a recent yeast proteomic study in which cleavage of a natural TEV recognition site, resulting in an N-terminal glycine, was used for purification of over 580 protein assemblies.[21] Finally, cleavage of the natural TEV protease recognition site can produce proteins with N-terminal serines that can be utilized in labeling through selective oxidation of the N-terminal serine to an aldehyde.[22]

Previous studies of TEV protease specificity indicated that it might be possible to use TEV protease to produce N-terminal cysteines. Genetic and biochemical analysis has shown that the P6, P3, and P1 amino acid residues are conserved in all known natural TEV cleavage sites and are major determinants of protease specificity.[17, 23] The remain-

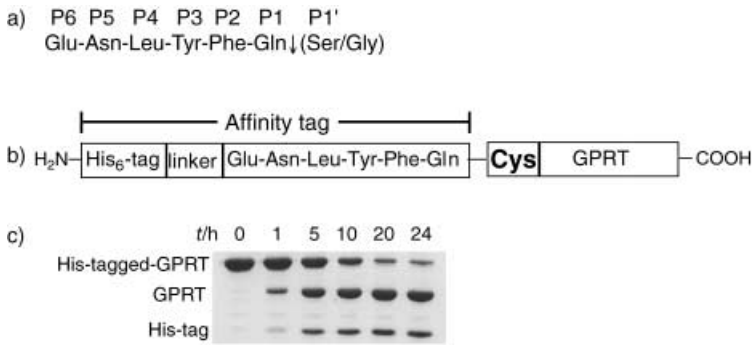


Figure 2. a) TEV protease cleavage site. b) Diagram of the histidine (His)-tagged GPRT fusion protein with mutated TEV cleavage site. c) Time course of cleavage of GPRT fusion protein (5 mg) with TEV protease (100 units; 1 unit of TEV protease will cleave 3 μg of a control substrate per hour).

  1433-7851/02/4112-2172 $ 20.00+.50/0 *Angew. Chem. Int. Ed.* **2002**, *41*, No. 12

ing residues can be mutated individually with some substitutions. Mutation of the P1′ residue to cysteine has been reported to reduce the TEV cleavage rate to 20% and only allow about 50% processing when tested in proteins produced by in vitro translation in rabbit reticulocyte lysate.[17, 18] We hypothesized that the incomplete processing and reduced cleavage rates previously observed for a P1′ cysteine mutation may be due to disulfide bond formation in the proteins produced in rabbit reticulocyte lysate. To test if mutation of the TEV site P1′ residue to cysteine would be efficiently processed, we expressed an affinity-tagged test protein, *E. coli* xanthine–guanine phosphoribosyltransferase (GPRT), with a mutated TEV cleavage site containing a cysteine replacement at P1′ (Figure 2b). A time course of the cleavage of 5 mg of the test protein with 100 units of TEV protease under reducing conditions revealed that the mutated TEV site was cleaved at approximately 70% the rate of the natural cleavage site (Figure 2c). In addition, the mutated cleavage site was processed to 95% completion in 24 h, indicating that TEV protease is synthetically useful in generating proteins with N-terminal cysteines. This procedure has also been applied to human interleukin-2 and *E. coli* sialic acid aldolase, and both proteins were cleaved to 90–95% completion at the correct sites to yield the desired N-terminal cysteines.

Next we set out to explore protein N-terminal labeling using TEV protease to generate N-terminal cysteines for chemoselective ligation. Two water-soluble thioester labeling reagents were synthesized for this purpose. Biotin and 5-carboxyfluorescein are labeling reagents commonly used to randomly label protein amines that are soluble at pH 7 and above, but will precipitate at more acidic pH because of protonation of their carboxylic acids. Reaction of biotin and 5-carboxyfluorescein with uncharged thiols produced water-insoluble thioesters that are not able to react with proteins in solution. By using *tert*-butyl mercaptoacetate to form the thioesters of biotin and 5-carboxyfluorescein, thioester labeling reagents **1** and **2** (Figure 3) were synthesized that were water soluble and able to react with proteins in aqueous solution in the absence of denaturants and organic cosolvents.

Initially simultaneous protease cleavage and thioester labeling of the affinity-tagged fusion proteins was attempted. Unfortunately the thioester labels are inhibitors of the TEV protease, possibly because the TEV protease is a cysteine protease with an active site cysteine that can be acylated by the thioesters. A two-step cleavage/labeling reaction was developed to overcome the inhibition, that is removing the affinity tag first, and then adding the thioester labels after the TEV protease cleavage was complete (Figure 3). Using this procedure it was found that 4 mM of the thioester labeling reagents were able to label 95–100% of the test protein in 12 h (Figure 4).

In summary, we have demonstrated the use of TEV protease to remove N-terminal affinity tags from fusion proteins to produce proteins with N-terminal cysteines and the subsequent use of those N-terminal cysteines for labeling. Advantages of using TEV protease to generate N-terminal cysteines include the high selectivity of the protease which makes it generally applicable to the cleavage of a wide range of fusion proteins and the ability to overexpress TEV protease for economical large scale applications. Cloning and expression of affinity-tagged fusion proteins with mutated TEV cleavage sites is straightforward and it should be simple to

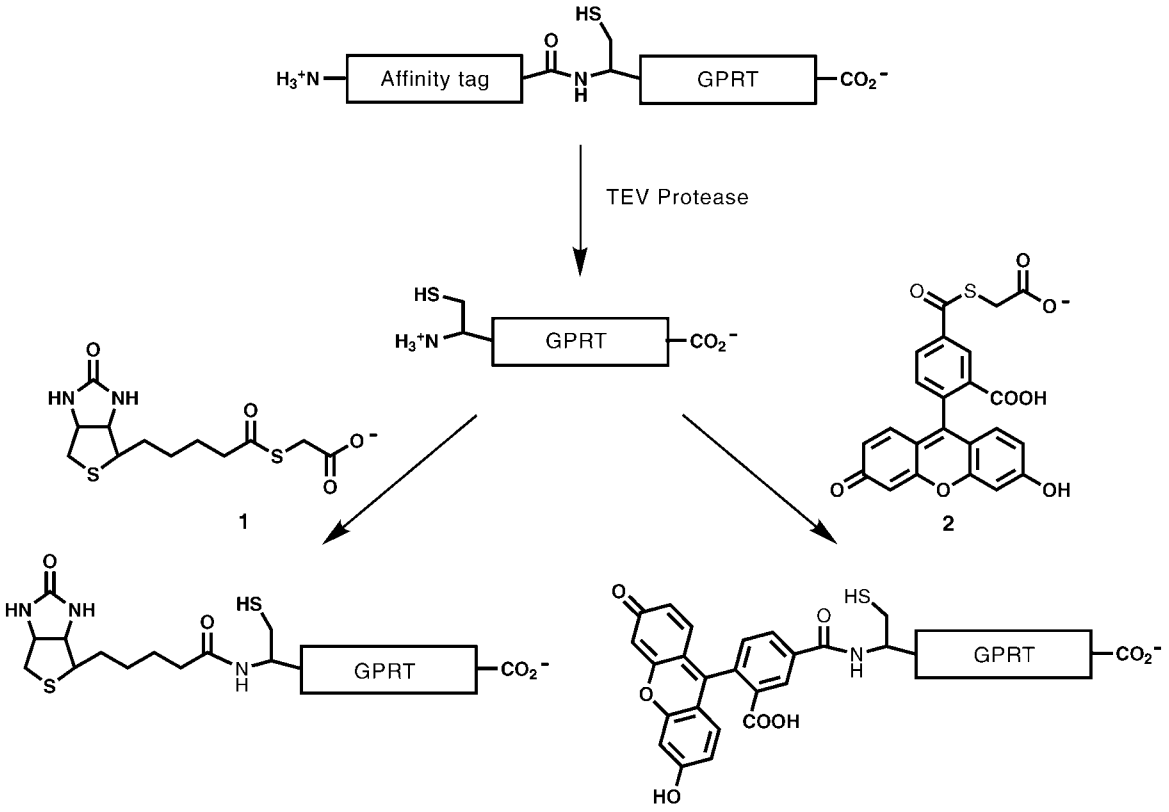


Figure 3. Two-step cleavage/labeling of GPRT using TEV protease and thioester labeling reagents **1** and **2**.

  1433-7851/02/4112-2173 $ 20.00+.50/0

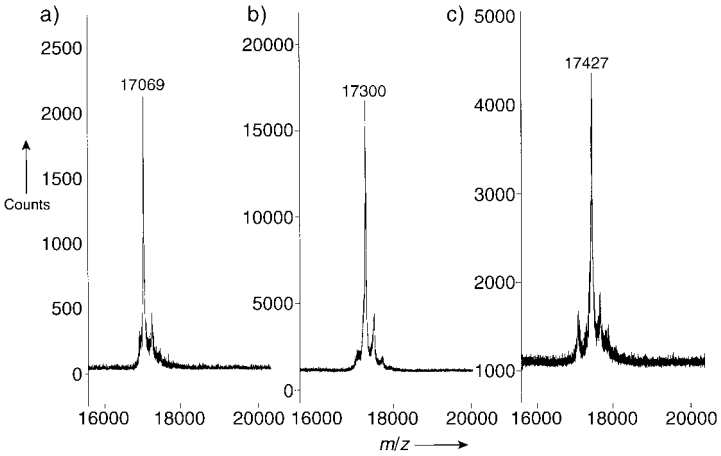


Figure 4. MALDI-TOF spectra of cleaved and labeled GPRT: a) GPRT: expected 17074, observed 17069; b) biotin-labeled GPRT: expected 17300, observed 17300; c) fluorescein-labeled GPRT: expected 17432, observed 17427.

apply to a large number of proteins. Synthesis of water-soluble thioester labels can be accomplished using *tert*-butyl mercaptoacetate, and proteins can be cleaved and labeled with these reagents in a two-step cleavage/labeling reaction under non-denaturing conditions. The affinity-tagged fusion proteins described herein are not significantly different from affinity-tagged fusion proteins produced in expression systems that have already been used to express and purify nearly all of the proteins of yeast for construction of functional protein arrays and microarrays.[24, 25] The combination of those types of expression strategies with removal of the affinity-tag and subsequent labeling with synthetic molecules should prove useful for many in vitro biochemical and proteomic applications. We hope that TEV protease generation of proteins with N-terminal cysteines will be of use in future protein labeling and semi-synthesis applications, and are currently investigating its application to a broader variety of proteins and glycoproteins, including their micro- and nano-fabrication to produce arrays with uniform orientation.

Received: March 1, 2002 [Z18809]

[1] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776.
[2] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.
[3] T. W. Muir, P. E. Dawson, S. B. H. Kent, *Methods Enzymol.* **1997**, *289*, 266.
[4] T. W. Muir, *Synlett* **2001**, 733.
[5] D. A. Erlanson, M. Chytil, G. L. Verdine, *Chem. Biol.* **1996**, *3*, 981.
[6] T. C. Evans, Jr., J. Benner, M.-Q. Xu, *Protein Sci.* **1998**, *7*, 2256.
[7] T. J. Tolbert, C.-H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 5421.
[8] L. A. Marcaurelle, L. S. Mizoue, J. Wilken, L. Oldham, S. B. H. Kent, T. M. Handel, C. R. Bertozzi, *Chem. Eur. J.* **2001**, *7*, 1129.
[9] L. Zhang, J. P. Tam, *Anal. Biochem.* **1996**, *233*, 87.
[10] Z. G. Zhao, J. S. Im, K. S. Lam, D. F. Lake, *Bioconjugate Chem.* **1999**, *10*, 424.
[11] M. Villain, J. Vizzavona, K. Rose, *Chem. Biol.* **2001**, *8*, 673.
[12] J. R. Falsey, M. Renil, S. Park, S. Li, K. S. Lam, *Bioconjugate Chem.* **2001**, *12*, 346.
[13] T. C. Evans, Jr., J. Benner, M.-Q. Xu, *J. Biol. Chem.* **1999**, *274*, 3923.
[14] M.-Q. Xu, T. C. Evans, Jr., *Methods (San Diego, CA, US)*, **2001**, *24*, 257.
[15] D. J. Matthews, J. A. Wells, *Science* **1993**, *260*, 1113.
[16] J. L. Harris, B. J. Backes, F. Leonetti, S. Mahrus, J. A. Ellman, C. S. Craik, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7754.
[17] W. G. Dougherty, J. C. Carrington, S. M. Cary, T. D. Parks, *EMBO J.* **1988**, *7*, 1281.
[18] W. G. Dougherty, S. M. Cary, T. D. Parks, *Virology* **1989**, *171*, 356.
[19] T. A. Smith, B. D. Kohorn, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5159.
[20] L. J. Lucast, R. T. Batey, J. A. Doudna, *BioTechniques* **2001**, *30*, 544.
[21] A.-C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzloch, A. Bauer, J. Schultz, J. M. Rick, A.-M. Michon, C.-M. Cruclat, M. Remor, C. Hofert, M. Schelder, M. Brajenovlc, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutweln, M.-A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, G. Superti-Furga, *Nature* **2002**, *415*, 141.
[22] K. F. Geoghegan, J. G. Stroh, *Bioconjugate Chem.* **1992**, *3*, 138.
[23] J. C. Carrington, W. G. Dougherty, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 3391.
[24] M. R. Martzen, S. M. McCraith, S. L. Spinelli, F. M. Torres, S. Fields, E. J. Grayhack, E. M. Phizicky, *Science* **1999**, *286*, 1153.
[25] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, M. Snyder, *Science* **2001**, *293*, 2101.

Icosahedral WAu_{12}: A Predicted Closed-Shell Species, Stabilized by Aurophilic Attraction and Relativity and in Accord with the 18-Electron Rule**

Pekka Pyykkö* and Nino Runeberg

Stable metal clusters are interesting for their own sake and as potential building blocks for nanostructures. This is especially true for gold due to the recently discovered catalytical properties of small or electronically tuned gold clusters.[1, 2] We here predict the existence of a series of isoelectronic clusters that may be particularly stable for three complementary reasons: 1) particularly efficient radial bonding, 2) the aurophilic attraction in the periphery, and 3) relativistic effects:

1) Consider an Au_{12} icosahedron. Its twelve 6s orbitals will span the irreducible representations of point group I_h $a_g + t_{1u} + h_g + t_{2u}$. The first three are spanned by the 6s, 6p, and 5d atomic orbitals of an atom at the center of the icosahedron, respectively. These three levels can hold 18 electrons. This is the number of valence electrons in WAu_{12} or in the isoelectronic ions $[TaAu_{12}]^-$ and $[ReAu_{12}]^+$.
2) As pointed out in ref. [3], covalent bonding and the van der Waals (vdW) bonding behind the metallophilic

[*] Prof. Dr. P. Pyykkö, Dr. N. Runeberg
Department of Chemistry
University of Helsinki
P.O. Box 55 (A.I. Virtasen aukio 1), 00014 Helsinki (Finland)
Fax: (+358) 9-191-50169
E-mail: Pekka.Pyykko@helsinki.fi

[**] This work was supported by The Academy of Finland. The computations were carried out at CSC, Espoo, Finland.

  1433-7851/02/4112-2174 $ 20.00+.50/0