DOI: 10.1002/cbic.201200045

A Universal Expression Tag for Structural and Functional Studies of Proteins

Vladimir V. Rogov,*[a, b] Alexis Rozenknop,[a, c] Natalia Yu. Rogova,[a] Frank Löhr,[a] Suhas Tikole,[a] Victor Jaravine,[a] Peter Güntert,[a, d] Ivan Dikic,[c, e, f] and Volker Dötsch*[a]

The expression, isolation, and purification of peptides and proteins are the crucial initial steps for investigation of their structures and functions. In recent years, many reports have shown strong progress in efficient expression and isolation of peptides and proteins by use of fusion technology with the aid of a variety of expression and solubility tags: GST, MBP, NusA, thioredoxin, ubiquitin, SUMO, GB-1, etc.[1]

In the majority of these applications the fusion tag is removed by proteolytic or chemical cleavage, and only the purified peptide/protein is used for further investigations. In cases of proteins and peptides that show a tendency to aggregate, however, the removal of the fusion tag can cause severe problems with solubility, preventing investigation with methods that require higher concentrations, such as NMR spectroscopy or isothermal titration calorimetry (ITC). For peptides, because of their small sizes relative to proteins, removal of a fusion tag can lead to further problems with regard to their detection and purification and to reaching sufficient concentrations.

To overcome this unfavorable situation, it could be beneficial to keep the fusion tag attached during the investigations, although this requires that the structural and functional properties of the protein or peptide not be changed. The concept of such non-cleavable solubility-enhancement tags (SETs) for studies of "poorly behaved" proteins by solution NMR was introduced by Zhou et al.[2,3] More than 30 proteins based on the small highly soluble protein GB-1[4] were successfully studied structurally.[3] However, the idea of SETs can be significantly extended for broader usage with alternative "universal" SETs, and can be implemented by combining expression, solubility, and stability enhancement tags in a single leading protein.

In this work we have tried to design a "universal" tag, which could be useful for most biophysical and biochemical applications and could be implemented both for in vitro and for in vivo studies of peptides/proteins and their interactions. We have tested the application of a designed ubiquitin-based (Ub-based) tag in structural and functional studies of a number of target peptides and proteins. In contrast with the "non-interacting" GB-1 tag most commonly used so far, Ub is known for its ability to interact nonspecifically with almost all proteins with $K_D > 0.3$ mM. The wild-type Ub possessing an N-terminal His tag has already been successfully used as an expression and solubility tag in several studies.[5] The major drawbacks of these fusion constructs were lower solubility enhancement than with the SUMO or NusA tags[6] and a spontaneous degradation of the fused construct in bacteria,[7] which is connected to the Ub-fold recognition and nonspecific cleavage of the Ub–GG–X bond by the ElaD protease.[8] In order to overcome these problems, we have redesigned the Ub sequence to shift the isoelectric point of the leading part containing a His_6 tag and Ub from 7.1 to 5.4 and simultaneously to increase the thermodynamic stability of the resulting protein with the aid of reported data.[9] Additionally, we have positioned the His_6 tag behind the Ub sequence in order to use the enhanced expression level provided by the Ub 5′-mRNA sequence (see Scheme 1 A and Figure S1 in the Supporting Information for details). The purpose of the presence of a TEV cleavage site is to enable additional protein purification steps by reverse purification if necessary. The reasons for the choice of a TEV cleavage site are the broad usage of this protease in research groups and the potential to use a range of protease inhibitors simultaneously with TEV cleavage. The substitution of the two C-terminal Gly units at positions 75 and 76 in Ub with Ser and Ala, respectively, reduced the spontaneous cleavage of the fusion constructs (Ub3, Scheme 1 A) in *E. coli*. However, to allow for in vitro and in vivo cleavage of the Ub tag with deubiquitination enzymes for various functional applications, we also created one version of our expression plasmid (Ub2) that still possesses the Ub-GG-X linkage. Two of our Ub fusion constructs (Scheme 1 B and C) contain internal His_{10} purification tags inserted into loop regions either between Pro19 and Ser 20 (Ub19 constructs) or between Glu63 and Lys64 (Ub63 constructs). These constructs are designed to increase Ni-NTA affinity (ten vs. six His units), to reduce the length of the linker between Ub and the target protein/peptide, and to provide

[a] Dr. V. V. Rogov, Dr. A. Rozenknop, N. Y. Rogova, Dr. F. Löhr, S. Tikole, Dr. V. Jaravine, Prof. Dr. P. Güntert, Prof. Dr. V. Dötsch
Institute of Biophysical Chemistry and
Center for Biomolecular Magnetic Resonance
Goethe University Frankfurt
Max-von-Laue Strasse 9, 60438 Frankfurt (Germany)
E-mail: rogov@bpc.uni-frankfurt.de
vdoetsch@em.uni-frankfurt.de

[b] Dr. V. V. Rogov
Institute of Protein Research
142290 Puschino (Russia)

[c] Dr. A. Rozenknop, Prof. Dr. I. Dikic
Institute of Biochemistry II and
Cluster of Excellence Macromolecular Complexes
Goethe University Frankfurt
Max-von-Laue Strasse 7, 60438 Frankfurt (Germany)

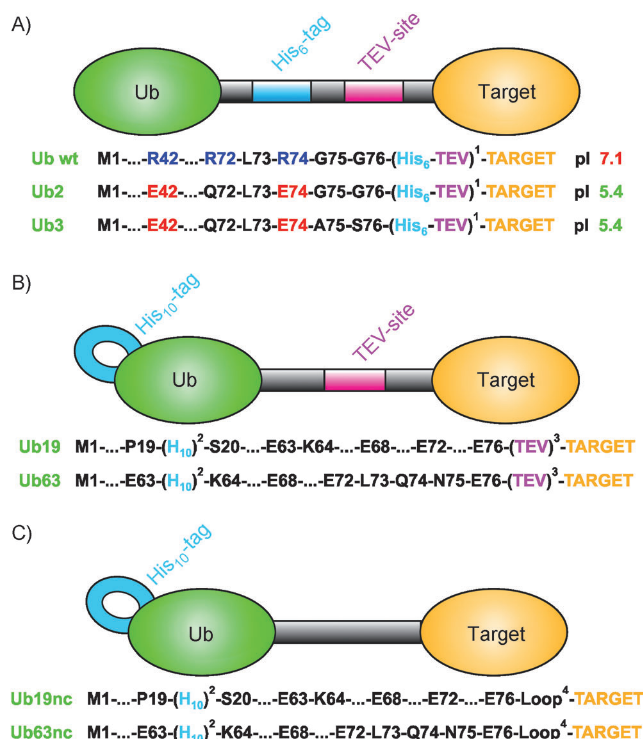
[d] Prof. Dr. P. Güntert
Frankfurt Institute for Advanced Studies, Goethe University Frankfurt
Ruth-Moufang-Strasse 1, 60438 Frankfurt am Main (Germany)

[e] Prof. Dr. I. Dikic
Mediterranean Institute for Life Sciences
Mestrovicevo setaliste bb, 21000 Split (Croatia)

[f] Prof. Dr. I. Dikic
School of Medicine, University of Split
Soltanska 2, 21000 Split (Croatia)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201200045.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim



Scheme 1. Representations of the designed Ub constructs. A) Ub2 and Ub3 constructs. B) Ub19 and Ub63 TEV-cleavable constructs. C) Ub19nc and Ub63nc TEV^{-} constructs. The most important features and differences within protein sequences are indicated. (His$_6$-TEV)^1: SGSGHHHHHHSAGENLYFQGA. (His$_{10}$)^2: GSAHHHHHHHHHHAGS. (TEV)^3: TQSATSDASGGENLYFQGA. (Loop)^4: TQSATSDASA.

Table 1. Expression yields of ubiquitin-fused proteins and peptides.

Target	Target MW [kDa]	Yield[a] [mg L^{-1}]	Yield increase (times)
LC3A	13.8	35	3[b,c]
LC3B	13.6	25	5[b,c]
GABARAPL-1	13.7	35	12[b,c]
TBK1_ULD	9.9	8	3[b]
RcsF	14.2	n.d.	n.d.
Δ30 RcsF	10.9	20	2[c]
Gao	40.3	30	–
p62-LIR	1.075	120	–
NBR1-LIR	1.524	110	–
NIX-W140	1.846	120	–
optineurin-LIR	1.716	110	–
GoLoko1	4.016	80	–

[a] Protein yields were calculated from pure NMR quality fractions. Comparisons are given for those proteins that have already been expressed in other fusion constructs (GST-[b] or NusA-[c]), where available. Peptide yields were calculated for the pure fractions of the Ub-fused constructs after Ni-NTA chromatography; n.d.: not determined.

a more restricted orientation of the targets coupled to Ni-NTA chips for surface plasmon resonance experiments. Through the positioning of the His$_{10}$ tag in a loop region close to the N terminus of Ub, the Ub 5′-mRNA sequence that is important for high expression yields remains intact.

The expression levels of the Ub19 and Ub63 fusion constructs remain similar to those of Ub2 and Ub3, and their abilities to bind to the Ni-NTA matrix and their TEV cleavage rates are indistinguishable as well. Products of TEV cleavage could be purified by reverse Ni-NTA chromatography. According to NMR and CD spectroscopy, Ub19 and Ub63 have structural and thermodynamic features equivalent to those of wild-type Ub.

To test the increases in expression achievable by use of these constructs we fused them with several proteins and peptides. The expression levels of all our targets fused to these new Ub tags were enhanced, with a few exceptions (Table 1). The final yields of the purified targets were higher in all cases, however, because of the fast and efficient isolation/purification scheme enabled by these constructs (Table 1 and Figure S2). With these constructs we achieved expression of different peptides with molecular weights ranging from 1.1 to 4.0 kDa for interaction studies with their specific receptor proteins. A typical yield of a Ub-peptide fusion construct was 100–150 mg L^{-1} of LB medium, which corresponds to 10 to 20 mg of the pure peptide. Such yields allowed us to optimize the peptide isolation and purification protocol in order to obtain NMR samples for structure determination of ≈90% pure peptide within 24 h (Figure S3). In cases of proteins fused to our Ub constructs, the yields depend on the stabilities of the target proteins and their tendencies to aggregate and to be degraded in the cells (Figure S2), varying from 5 to 50 mg of purified proteins per liter of M9 medium.

The use of Ub-fused proteins was a crucial step in our structural, thermodynamic, and functional studies of several proteins. Below we compare the results obtained with several biophysical methods for purified proteins and their Ub-fused forms.

CD studies of the fusion constructs

The small size of Ub and its extraordinary thermal stability under biologically relevant conditions (around pH 7) are crucial advantages for this system over the NusA or GST tags in CD studies. The CD contribution of a target protein is proportional to the molecular weight target/leader ratio, which is more favorable in the case of a Ub leader than for the larger NusA or GST tags. Therefore, deconvolution of CD spectra into partial secondary structure elements is more precise (Figure S4), providing more accurate information on the secondary structure composition of the target under given conditions. Indeed, previously reported data indicate that even a Ub-interacting moiety (yeast VPS27 ubiquitin-interacting motif, residues 258–279) can be analyzed by this approach with an error of less than 10%.[10] When subjected to thermal unfolding monitored by CD, the Ub-fusion constructs each show a single thermal transition—that of the target protein (Figure S4), because the melting temperature of Ub is ≥100 °C (90.0° at pH 4.0).[11] Thus, through the use of intact Ub-target fusion constructs, one can easily estimate the secondary structure contents and the stabilities of targets and follow the influence of target mutations on these parameters.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ITC studies of the fusion constructs

ITC experiments can provide important information on the thermodynamics of protein–ligand interactions. The accuracy of the results, however, depends on the absence of protein aggregation and on the ability to determine the precise concentrations of peptides even without Trp/Tyr residues. In order to test whether the Ub-fusion constructs could be used directly for ITC experiments in place of isolated and purified (or synthesized) peptide targets, we compared thermodynamic parameters of the interaction of the human autophagy modifiers[12] LC3A and LC3B with peptides spanning a LC3-interaction region (LIR) of the autophagy receptors p62 (residues 333–343) and NBR1 (residues 722–739). These peptides were used either as Ub fusion constructs or as synthetic peptides (Figure S5). In all cases the discrepancies between thermodynamic parameters did not exceed the concentration errors (Table S1), suggesting that the Ub leader has a negligible influence on these interactions. The measured K_D values ranged from 1 to 100 μM and were confirmed by NMR titrations.[13,14] The results indicate that this class of interactions can be studied without removal of the expression tags.

NMR studies of the fusion constructs

The SET approach has been successfully used for structural studies of a number of "badly behaving" proteins.[3] We fused the Ub-like domain (ULD) of TANK binding kinase 1 (TBK1) to Ub in order to protect this target from aggregation. Unlike Ub, TBK1-ULD has such a strong tendency to aggregate that the NMR study of this domain was only possible at concentrations of less than 200 μM, so only backbone resonances could be assigned.[15] Once fused to Ub, TBK1-ULD showed significantly better stability as well as less aggregation due to solubility enhancement and could be further characterized with NMR methods both structurally and functionally (Figure S6 A). We have not observed significant chemical shift perturbations (CSPs) on comparing the HN resonances of free TBK1-ULD and of TBK1-ULD fused to Ub. NMR titration experiments also showed the same CSP pattern for commercially synthesized and for expressed Ub-fused peptides (Figure S6 B). As well as time and cost benefits in relation to peptide synthesis, Ub-fused peptides additionally provide full buffer control and the potential to determine the exact concentration of a peptide moiety without internal Trp or Tyr residues.

Direct detection of protein–peptide interactions in cell lysates

The structural and biophysical characterization of many proteins is limited by the problem of their degradation and aggregation. The high expression levels of the Ub-fused targets enabled us to investigate ^{13}C,^{15}N-labeled proteins/peptides directly in the cell lysates without any isolation, purification, and/or concentration enhancement steps. In vivo, in the cellular environment, many proteins interact nonspecifically with other cellular components, supported by the high level of molecular crowding.[16] After cell lysis, crowding is significantly decreased, often enabling the direct detection of the overexpressed protein in the cell lysate. For a standard HSQC experiment we have found that 1–3 mL of isotope-enriched cell culture were enough to visualize all detectable resonances. For 3D experiments 7.5–20 mL were necessary (Figure 1). The "cell lysate" NMR sample can utilize up to 100 mL M9 cell culture without significant loss of spectral quality due to crowding. Although detection of proteins and peptides in lysates solves the problem of nonspecific interaction, this method also suffers from proteolytic degradation and the instability of the cell lysate itself. Detailed analysis showed that the redox equilibrium plays the most important role in the instability of cell lysate conditions.

We therefore optimized the protocol for cell lysate preparation (see the Supporting Information for methods and protocols). These modifications increased the cell lysate NMR sample

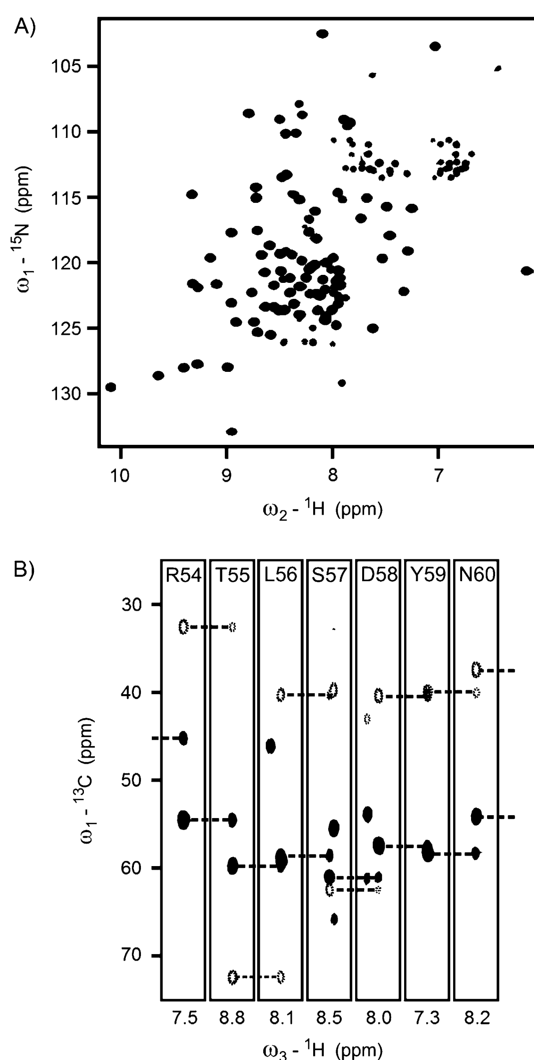


Figure 1. NMR spectroscopy of Ub-fused peptides in cell lysate. A) ^{1}H,^{15}N HSQC spectrum of the Ub2_p62-LIR fusion construct in cell lysate from M9 medium (2 mL). B) Slices of an HNCACB spectrum of the Ub2_p62-LIR fusion construct in cell lysate from M9 medium (7.5 mL). HNCACB correlations for R54–N60 of the Ub2 moiety are indicated by dashed lines. Cβ resonances are shown as dotted lines.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

lifetime from an initial 6 h at 25 °C to three days at 30 °C (Figure S7). The combination of the optimized protocols for cell lysate preparation, non-uniform sampling (NUS),[17] and hyperdimensional (HD)[18] NMR methods allowed us to use 10–15 mL of M9 cell culture for the automated assignment[19] of the backbone and side-chain resonances of the Ub3_NBR1-LIR fusion construct with the aid of an NMR dataset recorded during two days (see related paper by Tikole et al.[20]).

An additional application of the Ub-fused peptide expression coupled with NMR spectroscopy in cell lysate is direct screening for interaction partners or functional activity of the target proteins without time-consuming purification steps. We have investigated the interaction of the LC3(A,B,C) autophagy modifiers with peptides containing so-called LIR motifs (peptides with the aromatic amino acid-X-X-hydrophobic amino acid basic sequence). In the most effective LIRs the position of the aromatic residue is occupied by Trp (Figure 2, left).

There is no Trp residue present in the designed Ub tag, so the only Trp residue in the Ub-LIR-peptide fusion protein is directly involved in LC3 binding. The single Trp side-chain resonance in the $^{1}H,^{15}N$ HSQC spectrum is a reporter signal for monitoring the protein–peptide interaction. Indeed, upon titration of the cell lysate containing $^{13}C,^{15}N$-labeled Ub3_p62-LIR with unlabeled LC3B protein, we observed a distinct titration profile for this resonance. The experiments indicated a slow-exchange mode of interaction with a K_D value in the low-μM range, which corresponds to the published value for titration of $^{13}C,^{15}N$-labeled LC3B with synthetic non-labeled p62-LIR peptide.[14] For p62-LIR (and potentially for most peptide targets containing a single Trp residue), 1D NMR is enough for observation of the state of this residue, which makes these fusion constructs amenable for fast screening of binding events for structural studies.

In conclusion, the Ub tag described in this work combines properties for expression and solubility enhancement, like standard tags such as NusA, GST, SUMO or GB1, but with better efficiency. Whereas most of the tags have to be cleaved for further studies of the protein/peptide targets, we have shown the designed Ub fusion construct to be suitable for CD, ITC, and NMR experiments without tag removal. We have thus shown that protein–protein interaction, structural, and functional studies can be performed with the aid of target protein/peptide fused to Ub with the same results as with the "pure" protein/peptide but without loss of material over numerous purification steps. Moreover, proteins that could not be expressed and characterized by use of "traditional" tags could be studied efficiently as Ub fusions due to stabilization of the target fold and polypeptide chain integrity, presumably through unspecific interactions with Ub. This feature can be exploited for structural and functional investigations of targets without any isolation and purification in minimal experimental times (in cell lysate). Further improvements of the Ub tag and the design of other efficient expression tags should warrant study for even better functionality and universality.

Acknowledgements

This study was supported by the Centre for Biomolecular Magnetic Resonance and Cluster of Excellence Macromolecular Complexes, Frankfurt, Germany, the Deutsche Forschungsgemeinschaft (grant DFG JA1952/1-1), and the Lichtenberg program of the Volkswagen Foundation. We thank Dr. Sebastian Wagner (University of Copenhagen, Denmark) for providing us initial Ub_ABIN-1$_{(495–545)}$ construct and Dr. Lukas Zidek (Masaryk University, Czech Republic) and Jiri Novacek (Masaryk University, Czech Republic) for helpful suggestions and discussions.

Keywords: NMR spectroscopy · protein expression · solubility · tags · ubiquitin

[1] a) D. Waugh, *Trends Biotechnol.* **2005**, *23*, 316 – 320; b) D. B. Smith, K. S. Johnson, *Gene*, **1988**, *67*, 31 – 40; c) E. Samuelsson, T. Moks, B. Nilsson, M. Uhlen, *Biochemistry*, **1994**, *33*, 4207 – 4211.
[2] P. Zhou, A. A. Lugovskoy, G. Wagner, *J. Biomol. NMR* **2001**, *20*, 11 – 14.
[3] P. Zhou, G. Wagner, *J. Biomol. NMR* **2010**, *46*, 23 – 31.
[4] J. R Huth, C. A. Bewley, B. M. Jackson, A. G. Hinnebusch, G. M. Clore, A. M. Gronenborn, *Protein Sci.* **1997**, *6*, 2359 – 2364.
[5] a) T. Kohno, H. Kusunoki, K. Sato, K. Wakamatsu, *J. Biomol. NMR* **1998**, *12*, 109 – 121; b) M. Paal, T. Heel, R. Schneider, B. Auer, *Microb. Cell Fact.*

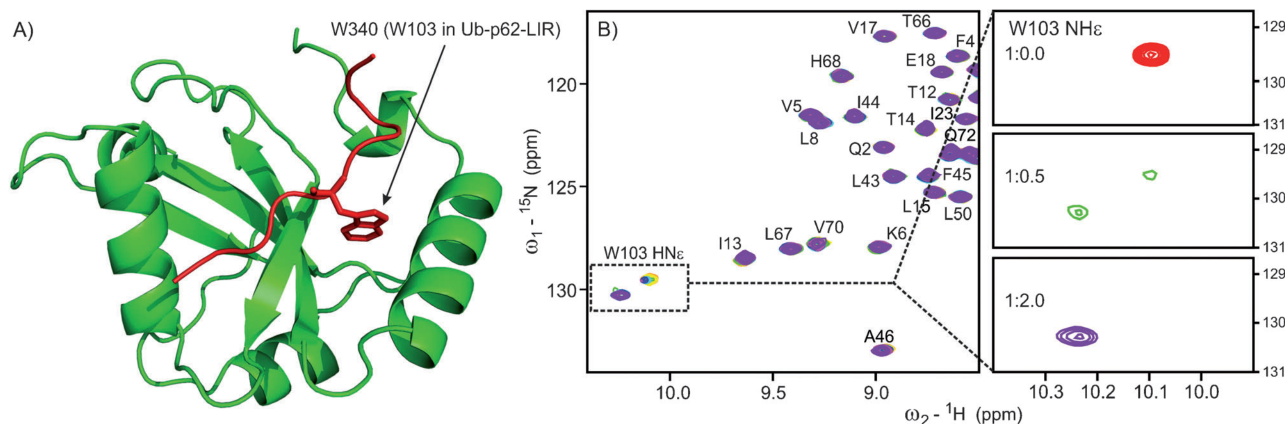


Figure 2. Fast estimation of the p62-LIR/LC3B interaction in cell lysate. A) Structure of the complex of LC3B (green) and p62-LIR (red; PDB ID: 2ZJD).[21] The only Trp residue side chain is shown. B) NMR spectra of the Ub3_p62-LIR cell lysate upon titration with LC3B. The molar ratios were varied from 0 (red) to 2 (violet) and are indicated where necessary.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

2009, *8*:7; c) C. Zou, S. Kumaran, R. Walser, O. Zerbe, *J. Pept. Sci.* **2009**, *15*, 184–191; d) N. G. Sgourakis, M. M. Patel, G. I. Makhatadze, A. E. Garcia, S. A. McCallum, *J. Mol. Biol.* **2010**, *396*, 1128–1144.
[6] a) C. Lee, C.-G. Lee, S. Takahashi, B.-G. Kim, *Biotechnol. Bioprocess Eng.* **2003**, *8*, 89–93; b) J. G. Marblestone, S. C. Edavettal, Y. Lim, P. Lim, X. Zuo, T. R. Butt, *Protein Sci.* **2006**, *15*, 182–189.
[7] A. M. Catanzariti, T. A. Soboleva, D. A. Jans, P. G. Board, R. T. Baker, *Protein Sci.* **2004**, *13*, 1331–1339.
[8] A. Catic, S. Misaghi, G. A. Korbel, H. L Ploegh, *PLos One* **2007**, *2*, e381.
[9] S. S. Strickler, A. V. Gribenko, A. V. Gribenko, T. R. Keiffer, J. Tomlinson, T. Reihle, V. V. Loladze, G. I. Makhatadze, *Biochemistry* **2006**, *45*, 2761–2766.
[10] M. M. Patel, N. G. Sgourakis, A. E. Garcia, G. I. Makhatadze, *Biochemistry* **2010**, *49*, 8455–8467.
[11] P. L. Wintrode, G. I. Makhatadze, P. L. Privalov, *Protein J.* **1994**, *34*, 246–253.
[12] V. Kirkin, D. G. McEwan, I. Novak, I. Dikic, *Mol. Cell* **2009**, *34*, 259–269.
[13] A. Rozenknop, V. V. Rogov, N. Yu. Rogova, F. Löhr, P. Güntert, I. Dikic, V. Dötsch, *J. Mol. Biol.* **2011**, *410*, 477–487.
[14] I. Novak, V. Kirkin, D. G. McEwan, J. Zhang, P. Wild, A. Rozenknop, V. Rogov, F. Löhr, D. Popovic, A. Occhipinti, A. S. Reichert, J. Terzic, V. Dötsch, P. A. Ney, I. Dikic, *EMBO Rep.* **2010**, *11*, 45–51.
[15] F. Ikeda, C. M. Hecker, A. Rozenknop, R. D. Nordmeier, V. Rogov, K. Hofmann, S. Akira, V. Dötsch, I. Dikic, *EMBO J.* **2007**, *26*, 3451–3462.
[16] G. J. Pielak, C. Li, A. C. Miklos, A. P. Schlesinger, K. M. Slade, G. F. Wang, I. G. Zigoneanu, *Biochemistry* **2009**, *48*, 226–234.
[17] J. C. Hoch, A. S. Stern, *NMR Data Processing*, Wiley, New York, **1996**.
[18] V. A. Jaravine, A. V. Zhuravleva, P. Permi, I. Ibraghimov, V. Y. Orekhov, *J. Am. Chem. Soc.* **2008**, *130*, 3927–3936.
[19] a) B. López-Méndez, P. Güntert, *J. Am. Chem. Soc.* **2006**, *128*, 13112–13122; b) D. E. Zimmerman, C. A. Kulikowski, Y. P. Huang, W. Q. Feng, M. Tashiro, S. Shimotakahara, C. Y. Chien, R. Powers, G. T. Montelione, *J. Mol. Biol.* **1997**, *269*, 592–610.
[20] S. Tikole, V. Jaravine, V. V. Rogov, A. Rozenknop, K. Schmöe, F. Löhr, V. Dötsch, P. Güntert, *ChemBioChem* **2012**; DOI: 10.1002/cbic.201200044 (next paper).
[21] Y. Ichimura, T. Kumanomidou, Y. S. Sou, T. Mizushima, J. Ezaki, T. Ueno, E. Kominami, T. Yamane, K. Tanaka, M. Komatsu, *J. Biol. Chem.* **2008**, *283*, 22847–22857.

Received: January 18, 2012
Published online on March 20, 2012

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim