



ACADEMIC  
PRESS

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Analytical Biochemistry 318 (2003) 71–79

ANALYTICAL  
BIOCHEMISTRY

[www.elsevier.com/locate/yabio](http://www.elsevier.com/locate/yabio)

## Screening methods to determine biophysical properties of proteins in structural genomics

Esmeralda A. Woestenenk, Martin Hammarström, Torleif Härd,\* and Helena Berglund

*Department of Biotechnology, Royal Institute of Technology (KTH), SE-106 91 Stockholm, Sweden*

Received 3 December 2002

### Abstract

We have developed and tested a simple and efficient protein purification method for biophysical screening of proteins and protein fragments by nuclear magnetic resonance (NMR) and optical methods, such as circular dichroism spectroscopy. The method constitutes an extension of previously described protocols for gene expression and protein solubility screening [M. Hammarström et al., (2002), *Protein Science* **11**, 313]. Using the present purification scheme it is possible to take several target proteins, produced as fusion proteins, from cell pellet to NMR spectrum and obtain a judgment on the suitability for further structural or biophysical studies in less than 1 day. The method is independent of individual protein properties as long as the target protein can be produced in soluble form with a fusion partner. Identical procedures for cell culturing, lysis, affinity chromatography, protease cleavage, and NMR sample preparation then initially require only optimization for different fusion partner and protease combinations. The purification method can be automated, scaled up or down, and extended to a traditional purification scheme. We have tested the method on several small human proteins produced in *Escherichia coli* and find that the method allows for detection of structured proteins and unfolded or molten globule-like proteins.

© 2003 Elsevier Science (USA). All rights reserved.

*Keywords:* High throughput; NMR spectroscopy; CD spectroscopy; Structural genomics; Heterologous gene expression; Protein purification

Structural genomics and related high-throughput biology require the possibility to screen candidate proteins and protein fragments to assess the suitability of these for structural studies. Depending on the outcome of the screen together with the priority level of the target the decision whether to produce and purify a larger amount of protein for structure determination, to test different fragments or introduce point mutations to improve stability and solubility, or to choose a better-behaving homologous protein can then be made. More traditional protein research can also benefit from screening methods for determination of domain boundaries or the effect of point mutations or for comparison of solubilizing fusion partners or purification methods.

Here we describe a straightforward and versatile method that allows for rapid biophysical character-

ization of proteins produced in *Escherichia coli*. The objective has been to establish a method that, without any protein-specific optimization, is applicable for purification of most proteins that can be produced with medium or high yields. The method should include a minimum number of reagents and purification steps but still yield protein samples that are sufficiently pure and concentrated so that a clean NMR<sup>1</sup> spectrum can be recorded to deduce the biophysical status of the product, for instance whether it is folded or unfolded.

<sup>1</sup> *Abbreviations used:* GST, glutathione *S*-transferase; IgG, immunoglobulin G; TEV, Tobacco Etch Virus; NMR, nuclear magnetic resonance; CD, circular dichroism; ANS, 8-anilino-1-naphthalene-sulfonic acid; HSQC, heteronuclear single quantum coherence; LB, Luria broth; DTT, dithiothreitol; TPPI, time-proportional phase incrementation; EDTA, ethylenediaminetetraacetic acid; TCTP, translationally controlled tumor protein; GFP, green fluorescent protein; CBF $\beta$ , core binding factor  $\beta$ ; HSP1, heat shock factor binding protein; IPKA, protein kinase inhibitor alpha; HAC, acetic acid; SDS-PAGE, sodium dodecylsulfate-polyacrylamide gel electrophoresis.

\* Corresponding author. Fax: +46-8-5537-8358.

*E-mail address:* [Torleif.Hard@biotech.kth.se](mailto:Torleif.Hard@biotech.kth.se) (T. Härd).

The method should be scalable and it should be possible to extend it to a traditional purification scheme and to use it in high-throughput applications.

We base our method on production of fusion proteins containing the target protein and different N-terminal fusion partners. N-terminal fusions can enhance the expression level and solubility of proteins produced in bacteria and we have previously described a procedure for high-throughput expression and solubility screening which is based on comparison of several different N-terminal fusion partners [1]. Here we take advantage of the possibility to also use the fusion partners as affinity tags for purification and extend the solubility screening to a general purification scheme as shown in Fig. 1. The proteins are produced with Gb1, GST, or the Z domain as N-terminal fusion partners. Starting with a small-scale culture (typically 50 ml) in  $^{15}\text{N}$ -labeled minimal medium, the cells are lysed by sonication and the product protein is bound on an IgG or glutathione Sepharose affinity matrix. Cleavage of the fusion is performed on the column using TEV or 3C (PreScission) proteases, and the target protein is eluted. A concentration step is added if needed. All steps can be performed for many samples in parallel using standard laboratory equipment.

We use NMR spectroscopy for biophysical characterization and complement the NMR data with circular dichroism (CD) and ANS fluorescence spectroscopy. The line widths and dispersion of resonances in a  $^1\text{H}$ - $^{15}\text{N}$ -HSQC NMR spectrum is a superior determinant of whether a protein is folded and exists in a single well-defined conformation, while a CD spectrum gives information about the secondary structure content. ANS fluorescence can be used in ambiguous cases to decide whether a protein is a potential molten globule.

We have tested the procedures on a number of human proteins and found that in most cases it is possible to perform all steps from protein production to the recording of an NMR spectrum in 1 day.

## Materials and methods

### Gene expression and protein production

Gene subcloning, expression, and solubility screening were performed as described [1]. The three expression vectors used in this study were pDEST-TH3 (N-terminal Gb1 domain fusion with a TEV protease cleavage site), pDEST-TH6 (N-terminal GST with a 3C protease cleavage site), and pDEST-TH10 (N-terminal Z domain with a 3C cleavage site). The solubilizing effect of these fusion partners was tested in the earlier study [1], which also describes the preparation of pDEST-TH3. The pDEST-TH6 vector is derived from the Gateway vector

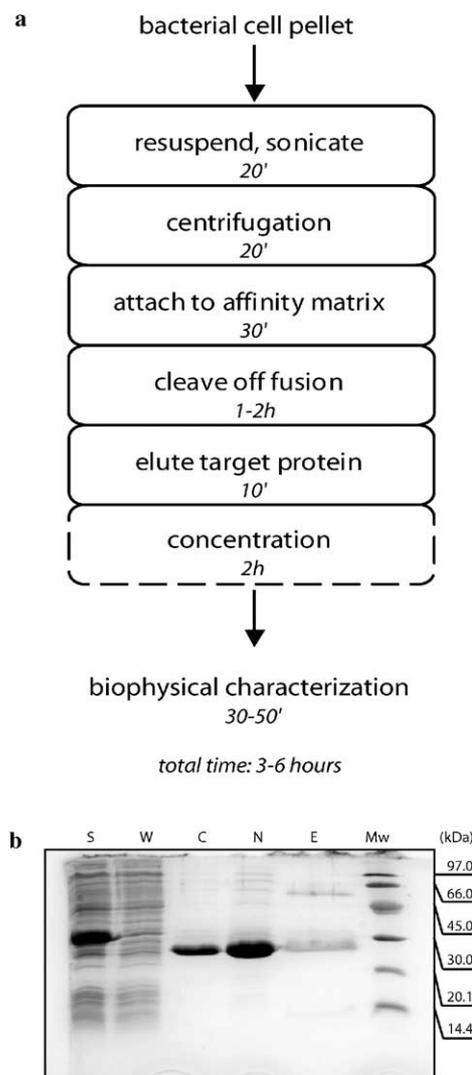


Fig. 1. (a) Schematic representation of the purification scheme. The duration of the different steps is shown in italic, and the total time including measurement of the NMR (and/or CD) spectra accumulates to 3 to 6 h, depending on the protease that is used and whether a concentration step is included. (b) SDS-polyacrylamide gel showing a typical purification, in this case of the human translationally controlled tumor protein (TCTP). Lane S contains the soluble fraction of the sonicated cells, with the target protein fused to Gb1 as a 28-kDa product (apparent size 35 kDa); lane W is the flow-through after binding of the fusion protein to the IgG column; lane C is the cleaved TCTP protein (19.5 kDa; apparent size 27 kDa); lane N is the concentrated sample that was subsequently used for characterization by NMR and CD spectroscopy; lane E contains the eluate from the IgG column. (The cleaved Gb1 fusion partner is not observed and it is often not visible on 16% polyacrylamide gels due to its small size.) The right lane contains molecular weight markers for which the sizes have been indicated.

pDEST-15 (Invitrogen) by the insertion of the nucleotide sequence CTGGAAGTTCTGTTCCAGGGGCC coding for the recognition site of 3C (or PreScission) protease, located 3' to GST and 5' to the ATTR1 recombination site. The vector pDEST-TH10 is derived

from pDEST-TH5 [1] in which the second of the two Z domains was replaced by a 3C cleavage site.

Proteins were produced in the *E. coli* strain BL-21 (DE3) Codon Plus RP (Stratagene). Overnight cultures were grown at 37 °C from a glycerol stock or from a colony on a LB agar plate in 3–5 ml LB medium containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. The next day, 50 ml M9 medium [2], containing ampicillin, chloramphenicol, and 0.8 g/L <sup>15</sup>NH<sub>4</sub>Cl as the only nitrogen source (Martek Biosciences), was preheated to 37 °C, inoculated with 1.0 to 1.5 ml of the overnight culture, and grown at 37 °C. When the OD<sub>600</sub> reached 0.8 to 1.0, expression of the heterologous gene was induced by addition of isopropyl-β-D-thiogalactoside (Boehringer Mannheim) to a final concentration of 1 mM. After 2 to 3 h, the cells were harvested by centrifugation at 3000g for 20 min.

#### *Protein purification and on-column fusion cleavage*

The cells were resuspended in 2 ml of the wash buffers recommended for the two affinity matrices: 50 mM Tris, 150 mM NaCl, 0.05% Tween 20 (TST) at pH 7.6 for IgG Sepharose 6 Fast Flow (Amersham Biosciences) and 140 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, and 1.8 mM KH<sub>2</sub>PO<sub>4</sub> (PBS) at pH 7.3 for glutathione Sepharose 4b (Amersham Biosciences). Lysis was performed in a Branson Digital Sonifier with six 10-s pulses, alternated with pauses of 5 s, while cooling the sample in ice water. Alternatively, a pulse time of 3 min was used, with pulses and pauses of 1 s. The cell debris were pelleted by centrifugation at 8000g for 20 min and the supernatant was collected. Home-packed columns containing 1 to 2 ml affinity matrix were run by gravity flow and equilibrated using the wash buffers described above. The sonicate was bound to the affinity matrix and a subsequent washing step removed cellular proteins. The outlet of the column was closed and protease was added; 150 µl 0.25 mg/ml TEV protease for Gb1 fusion proteins and 30 µl 10 mg/ml 3C protease (provided by My Hedhammar, Department of Biotechnology, KTH) or 10 units PreScission protease (Amersham Biosciences) for GST and Z fusion proteins. The protease solutions were diluted to a total volume of 1 ml (5 mM dithiothreitol) to enhance distribution over the column. The top of the column was closed and it was left to incubate at room temperature for 1 h under gentle agitation. The flow-through was collected and subsequently 2 ml wash buffer was added, of which one column volume was collected. The two fractions were pooled and, if necessary, concentrated in Centricon YM-3 concentrators for 2 h. Remaining bound material was eluted for SDS-PAGE analysis with 0.5 M acetic acid (HAc), with pH adjusted to 3.4 using NH<sub>4</sub>Ac, for IgG Sepharose and 10 mM glutathione in 50 mM Tris-HCl at pH 8.0 for glutathione Sepharose.

#### *Protein purification using multiwell filter plates*

Purification of proteins in 96-well filter plates followed essentially the same protocol as that described in the previous paragraph. The supernatant recovered after centrifugation of the lysed cells was bound to 1.0–1.5 ml affinity matrix in a Whatman UniFilter 2000 96-well polypropylene filter plate (GF/B or PP Mesh 10 µm; 2-ml wells). Subsequent washing and elution steps were performed by centrifugation at 100g for 1 min in a swing-out rotor. Flow-throughs were collected in 96-well plates (Greiner polypropylene Masterblock; 2-ml wells). For the protease digestion step, the inlets and outlets of the filter plate were closed with Parafilm, and the plate was incubated with gentle shaking.

#### *Alternative batch cleavage of the fusion protein*

Protease and DTT were directly added to the soluble lysate containing the overproduced fusion protein and the cleavage reaction was performed at the same temperature and for the same time as those of the on-column cleavage. After digestion, the protein–protease mix was run over the columns mentioned above to remove the fusion partner. If necessary, the digested protein sample was concentrated.

#### *NMR spectroscopy*

NMR samples consisted of 500 to 600 µl of the protein solutions, to which 10% (vol/vol) D<sub>2</sub>O (Cambridge Isotope Laboratories) was added. The pH was in some cases adjusted to below 7 using diluted HCl. <sup>1</sup>H-<sup>15</sup>N heteronuclear single quantum coherence (HSQC) [3] and fast HSQC (FHSQC) [4] experiments were recorded at 30 °C on Bruker AVANCE 500- and 600-MHz spectrometers equipped with 5-mm triple-resonance probes with Z gradients. The sensitivity of these TXI probes are 830:1 and 1200:1 on 500 and 600 MHz, respectively, as measured using the standard ethylbenzene S:N test with a 2-ppm noise region. HSQC experiments were carried out with 128 increments using TPPI (HSQC) or States-TPPI (FHSQC) for phase sensitivity in the indirect dimension and with 8 to 32 transients per increment, unless a longer data collection is mentioned. For comparison we also recorded NMR data for some of the samples on a Bruker DRX 600 MHz spectrometer equipped with a 5-mm triple-resonance Z-gradient cryoprobe with a sensitivity of 3800:1 (2 ppm noise). Data were processed using XWIN-NMR on SGI workstations.

#### *CD spectroscopy*

Circular dichroism was measured on the protein samples that were also used for NMR. Depending on

the concentration, CD samples were prepared either in a 0.1-mm cuvette, with a total volume of 20  $\mu$ l undiluted NMR sample, or in a 1.0-mm cuvette, following three-fold dilution of the NMR sample to a total volume of 200  $\mu$ l. CD in the wavelength range 200 to 250 nm was then measured at room temperature on a Jasco J-810 spectropolarimeter.

#### *ANS fluorescence measurements*

Protein concentrations were measured using the Bradford assay and samples were diluted to 0.37 mg/ml with phosphate buffer (20 mM KPi, pH 6.3; 1 mM EDTA). Protein samples (45  $\mu$ l) were then mixed with 5  $\mu$ l of a 0.6 mM solution of ANS (Sigma) to final concentrations of 0.33 mg/ml protein and 60  $\mu$ M ANS. Fluorescence emission was recorded at room temperature on a Perkin–Elmer LS50B luminescence spectrometer using an excitation wavelength of 385 nm and scanning the emission wavelength between 400 and 600 nm with a slit width of 6 nm on both excitation and emission.

## **Results**

#### *Protein production and purification*

Growth and gene expression in 50-ml cultures require about 3 1/2 h if the cell mass at the beginning of growth in minimal medium is high and the protein is produced at medium to high levels. Heterologous gene expression is in this case induced 1 h after inoculation. The different steps in the purification protocol and the times required to perform these are shown in Fig. 1a. About 40 min are allowed for cell resuspension, lysis by sonication, and centrifugation to remove cell debris. Binding of the protein to the affinity matrix and washing out endogenous *E. coli* proteins is carried out in 30 min. The pDEST-TH3 (Gb1) fusion protein is digested while bound to the IgG Sepharose column with TEV protease (1 h at room temperature), while the pDEST-TH6 (GST) and pDEST-TH10 (Z domain) fusion proteins, bound to glutathione and IgG Sepharose, respectively, are cleaved with 3C protease (2 h at room temperature) or PreScission (2 h at 4°C). After digestion the target protein is eluted from the column while the fusion partner and any undigested fusion protein stay bound. If necessary, samples are concentrated to a suitable NMR volume (500 to 600  $\mu$ l) to use all protein obtained from the 50-ml cultures. This optional concentration step adds 2 h to the sample preparation scheme, but production and biophysical characterization of a protein sample still can be performed within a day when concentration is needed. In cases where a large number of cultures are handled simultaneously, special growth conditions are required, or the chosen culture volume is

larger than 50 ml, it is more convenient to do the protein production during one day and the protein purification and analysis during the next day.

An example of a purification procedure as followed by SDS–PAGE is shown in Fig. 1b. The produced fusion protein consisting of target (human TCTP) and Gb1 can in this case be seen as the most abundant protein in the sonicate. It is not present in the flow-through collected after loading on the affinity matrix, indicating strong binding to IgG Sepharose. Digestion with TEV protease on the column releases cleaved TCTP and the eluted fraction is sufficiently pure for further biophysical characterization. The enzymatic cleavage is almost quantitative because no fusion protein can be detected in the eluate obtained with 0.5 M HAc. The protease is co-purified with the product, but at low concentrations. (The molar ratio of product to protease is on the order of 50 or higher.)

#### *Application on a test set of proteins*

In this study we tested the purification protocols on 19 proteins in the size range 7 to 27 kDa including 15 human proteins for which the function and structure were unknown at the time of selection. To these we added two positive controls, i.e., human proteins that we had purified and studied by NMR in the lab previously, and two proteins that were either known to be well-behaving (DsbA from *E. coli*) or fluorescent (GFP from *Aequorea victoria*) and therefore in both cases convenient to include for the purpose of method development. The 15 unknown proteins were chosen based on their high expression and solubility levels in *E. coli* and 12 of these come from a set of proteins described previously [1]. A total of 17 of the 19 proteins including the four controls were successfully purified, whereas 2 proteins in the set precipitated on the affinity matrix column. All 17 purified proteins also yielded an interpretable NMR spectrum, in this case a  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum in which all or most resonances are visible. A standard HSQC experiment with eight transients and 128 increments is carried out in 15 to 20 min, while about 30 min are needed to prepare the sample and the spectrometer for the experiment. Using our standard room-temperature probes for 500- and 600-MHz spectrometers we found that in favorable cases it is possible to obtain a high-quality HSQC spectrum of the purified protein in less than 20 min without concentration of the 1- to 2-ml volume in which the samples were eluted. About two thirds of the proteins required concentration and a few samples required data collection times of more than 2 h. The protein concentration needed to record an NMR spectrum for the purpose of biophysical characterization is considerably lower than that needed for a full structure determination. As an example, a spectrum of human CBF $\beta$  at a concentration of 60  $\mu$ M was recorded

using a cryoprobe at 600 MHz in less than 20 min (Fig. 2a). We found that the salt concentration and buffers used here do not attenuate the nominal sensi-

tivity of the cryoprobe compared to a room-temperature probe and that a fourfold sensitivity gain is achieved when using a cryoprobe.

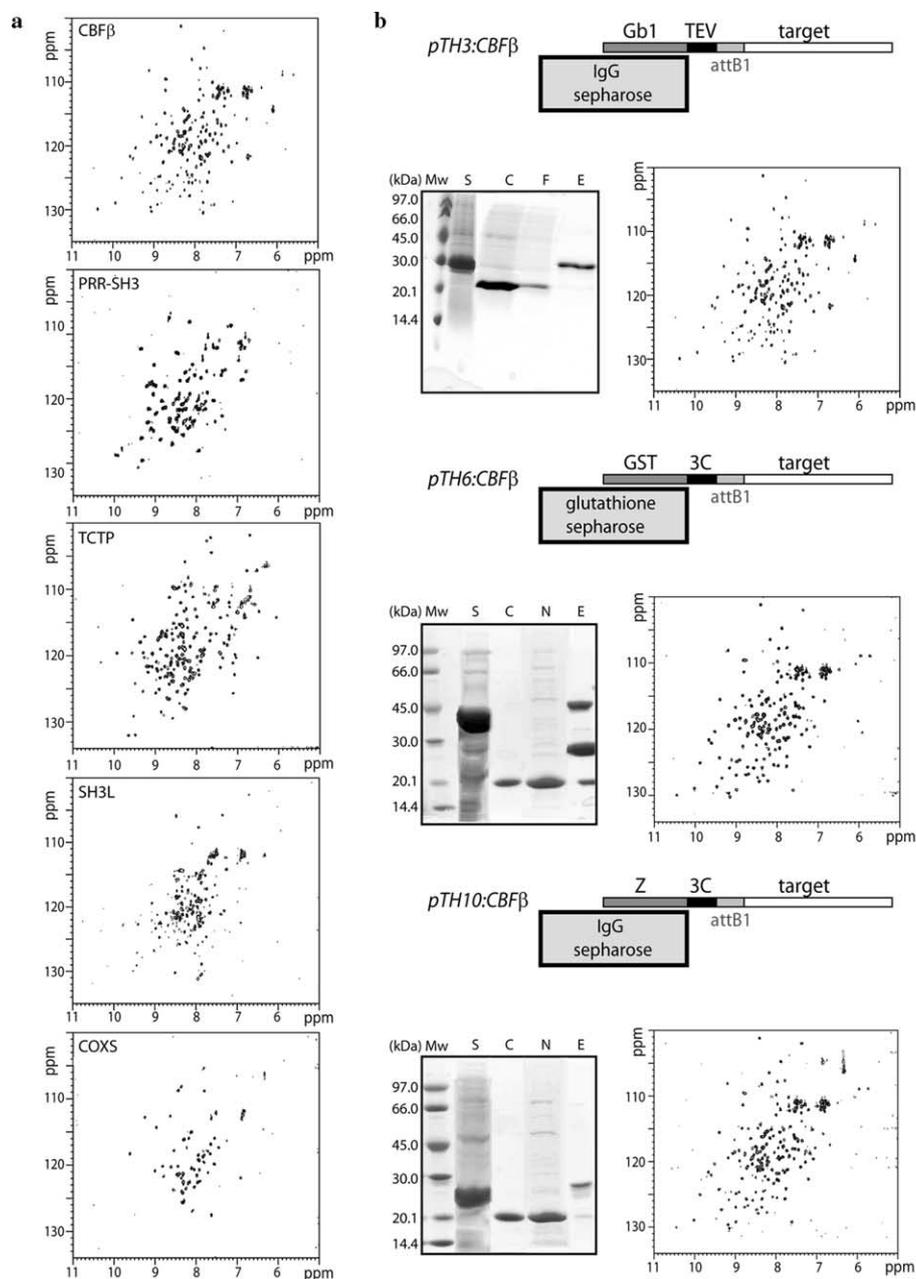


Fig. 2. (a) <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra of structured target proteins. Human CBFβ [14] and PRR-SH3 [15], which have been studied previously in our laboratory, were produced as Gb1 fusions. The spectrum of unconcentrated CBFβ (approx 60 μM) was recorded in less than 20 min on a 600-MHz spectrometer equipped with a cryoprobe. The remaining spectra are of proteins for which the structures were not known at the time of selection: human TCTP (produced as Gb1 fusion), human SH3L (SH3 domain-binding glutamic acid-rich protein; Z fusion; concentrated sample), and human COXS (cytochrome *c* oxidase copper chaperone; Z fusion; concentrated sample). (b) CBFβ produced with different expression vectors. The expression constructs with protease sites and affinities used for purification are shown schematically. The four lanes of the SDS-PAGE gel in the top panel represent the soluble fraction after cell lysis (S), the flow-through after TEV digestion (C), the second flow-through (upon addition of 1 ml TST buffer) (F), and the eluate of undigested protein (E). The NMR spectrum is identical to that shown in (a) and was run on the unconcentrated C fraction. The SDS-PAGE gels in the middle and bottom panels show the concentrated flow-through (lanes N), which were used to record the NMR spectra. Some differences in chemical shifts of arginine side chain resonances in NMR spectrum in the top panel compared to the middle and bottom panels are due to different spectral widths in the indirect <sup>15</sup>N dimension. The results shown in the middle panel were obtained by purification in a 96-well filter plate (see Materials and methods). In this case, the eluate (E) lane demonstrates the consequences of incomplete cleavage and flow-through processes, showing undigested fusion protein, free GST, and some cleaved CBFβ. With the GST system requiring relatively low flow rates, the incomplete release of cleaved protein is likely to be the consequence of high flow rates during the centrifugation step.

Examples of NMR spectra of folded human proteins are shown in Fig. 2a. The favorable dispersion and sharp resonances show that these proteins are folded into a single stable conformation and are probably monomeric. Three of the five NMR spectra in Fig. 2a are of unknown human proteins and the high quality of the data indicates that it is going to be possible to determine their structures. A number of proteins gave good results with more than one fusion partner. For instance, CBF $\beta$ , a human transcription factor-associated protein, was found to express at high levels with all three fusion partners tested in this study (Fig. 2b). The different steps in the purification process are shown on the SDS–polyacrylamide gels in the figure. The resulting NMR spectrum is practically identical in the three cases, indicating that the structural properties of this protein

are independent of the fusion partner. The top panel in Fig. 2b also shows that it is possible to perform a batch protease digestion (as described under Materials and methods) followed by a simple removal of the fusion partner on an affinity column. The middle panel is an example of the application of the purification method to the 96-well format (see figure legend and Materials and methods).

### Biophysical characterization

NMR can be combined with different biophysical techniques to provide a more complete picture of the structural state of a protein. Techniques that can be used for screening purposes include circular dichroism and fluorescence spectroscopy, as illustrated in Fig. 3. The

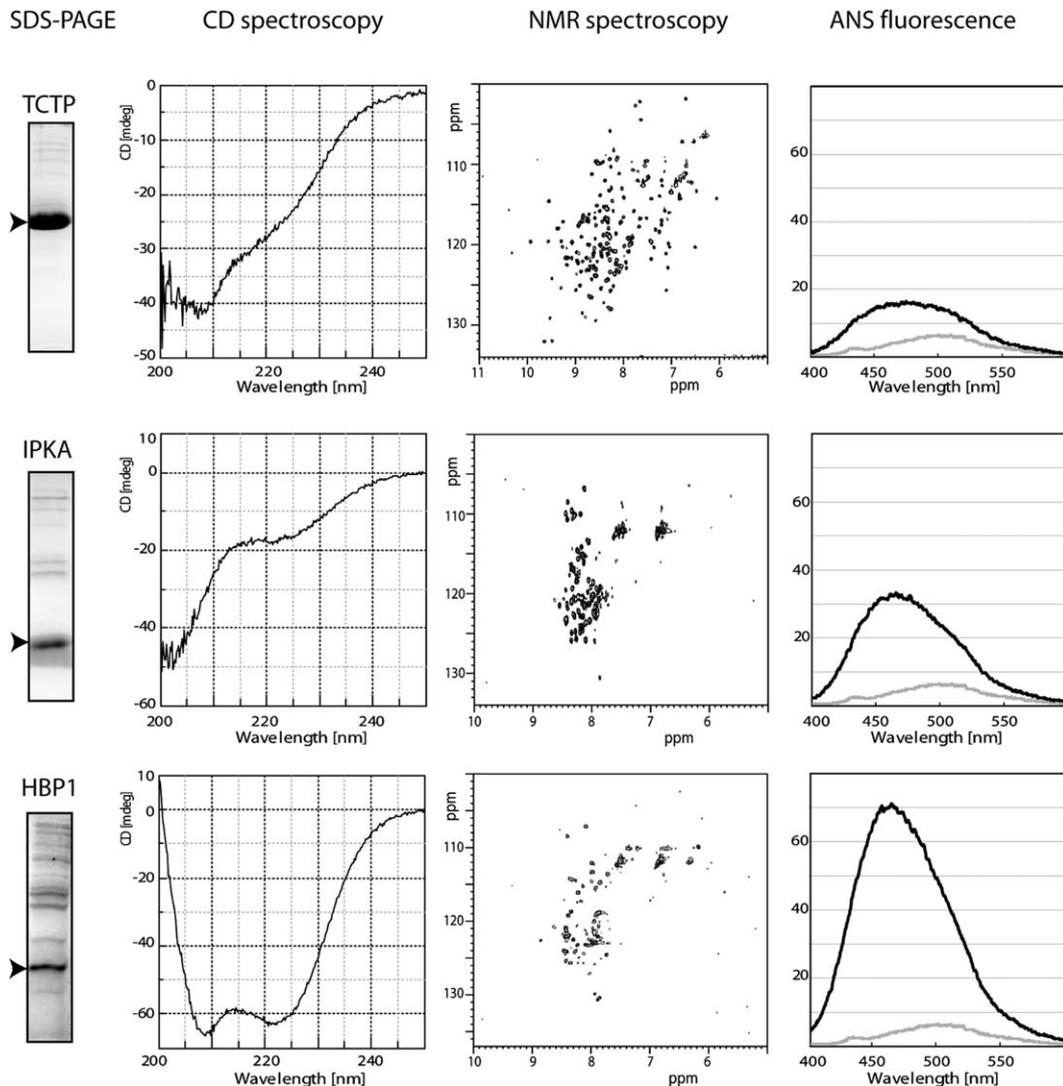


Fig. 3. Biophysical characterization. CD, HSQC, and ANS fluorescence emission spectra of the 19.5-kDa TCTP (top; classified as “folded”), the 8.0-kDa human protein IPKA (protein kinase inhibitor alpha; center; classified as “unstructured”), and the 8.5-kDa human protein HBP1 (heat shock factor binding protein; bottom; classified as “molten globule”). For ANS fluorescence, gray and black emission spectra were recorded in buffer and in the presence of protein, respectively. The fluorescence scale is in arbitrary units. The fluorescence of TCTP is the same as the one shown in Fig. 2a.

CD spectrum of the human HBPI protein shows a typical  $\alpha$ -helical profile with CD minima at 208 and 222 nm, while we had initially judged from the poor dispersion of the NMR spectrum that this protein was disordered. We therefore tested HBPI for molten globule behavior by binding it to the fluorescent dye ANS and observed the characteristic increase of fluorescence and shift in emission wavelength that is indicative of binding of ANS to molten globules [5]. From these data we conclude that HBPI is not suited for further structural studies under the conditions currently used. The middle panel in Fig. 3 shows typical CD and NMR spectra of an unstructured protein—human IPKA; there is a very narrow dispersion of backbone and side-chain NMR resonances and the CD spectrum approaches a sharp minimum at 200 nm. An unfolded state is further supported by a significantly lower increase in ANS fluorescence upon binding to IPKA than that observed with HBPI. As a comparison we also show corresponding data obtained with TCTP (see SDS gel in Fig. 1). As expected, ANS does not bind well to this well-folded protein. The CD spectrum is not completely conclusive, albeit indicative of  $\beta$ -sheet content, but the NMR spectrum nevertheless confirms a well-defined structure. The biophysical characterization of a larger set of human proteins with unknown structure and function will be published elsewhere.

## Discussion

We have developed a protein preparation procedure that allows for rapid biophysical characterization of proteins produced with good yield in *E. coli*. One important feature of the method is that it is general in the sense that it does not depend on any particular physical properties of the target protein, as long as it can be produced in soluble form with a fusion protein. Another important feature is that it constitutes a natural extension of our previously published subcloning and expression and solubility screening protocols [1].

All steps prior to biophysical characterization are easily performed in parallel and on a 96-well-plate format, even with standard laboratory equipment. Since the purification procedure consists of standard methods, it is straightforward to extend it to full-scale batch purification when high-concentration samples are needed (e.g., for structural studies by NMR or X-ray crystallography). The fact that identical steps are used with different fusion and protease combinations should also allow for a general automation procedure, possibly using 24-well plates. We find that, in cases with high expression yield, it is possible to obtain enough protein from 50 ml minimal medium cultures. Expression in minimal medium is often considered difficult for the cells, but  $^{15}\text{N}$ -rich media that may yield protein

expression levels comparable to growth in LB medium are commercially available.

For many applications, for instance biochemical assaying, it might not be necessary to remove the fusion partner by cleavage. However, it is necessary to remove it if the overall objective is to study structure. In addition, the sharp lines from small fusion proteins tend to dominate the NMR spectrum when working with larger proteins and the CD spectrum of the stably folded fusion partner might overshadow that of the target protein. An important issue has been whether proteins that are screened for solubility with different fusion partners are also soluble following protease cleavage and concentration. Of the 19 proteins tested for this study only 2 precipitated during purification on the affinity matrix column. These results concern a limited set of small proteins, but they are nevertheless encouraging for further studies of the effect of solubilizing fusion partners for protein production in *E. coli*.

It is conceivable that on-column digestion of the fusion protein in some cases can cause problems, e.g., due to inaccessibility of the protease cleavage site or interaction of the protease with the column. In such cases we suggest that the digestion step should be done directly on the lysate. Since this prevents removal of the cellular proteins, the background protein level may be a problem but there are methods to avoid interfering signals from cellular proteins [6,7]. Batch cleavage instead of on-column cleavage might also be feasible even without any attempts to decrease the background protein level (top panel in Fig. 2b).

A number of approaches to obtain an NMR spectrum without elaborate purification have been described previously. For instance, several groups have acknowledged the possibility of recording a high-quality NMR spectrum of an overproduced protein in cell lysate without any extensive purification. The success has been attributed to the surplus of the overproduced protein and that the large size of many proteins and protein complexes in the cell make them invisible in the NMR spectrum due to line broadening. Gronenborn and Clore [8] show how informative  $^{15}\text{N}$ -HSQCs of two proteins can be obtained after simply subjecting the cell lysate to concentration and buffer exchange. There have also been reports on different labeling schemes to enhance the labeling of the target, keeping the background proteins unlabeled or labeled at a lower level. Addition of rifampicin at the time for induction of expression in connection with a change from unlabeled to labeled medium is a way to increase the specificity of the labeling of overproduced protein and has been shown to yield NMR data of good quality [6]. Specific amino acid labeling has also been suggested as a way to reduce the need for sample purification [7]. Most examples of NMR studies on lysate samples have, however, been

demonstrated on only one or two proteins, while other studies do not make use of solubilizing fusion partners [9].

We favor NMR and CD as the techniques of choice for biophysical screening of proteins because of the high information content of the data. The drawbacks are that neither of these is as yet a true high-throughput technique and that there is a physical size limit on proteins studied by NMR. However, it must be noted that this latter constraint is less severe if the purpose is to use NMR for characterization rather than for full structure determination. In fact, an HSQC spectrum on a  $^{15}\text{N}$ -labeled, but nondeuterated, sample can give enough information to judge the content of well-defined structure in proteins well into the 40- to 50-kDa size range (see for instance [10]). The HSQC NMR spectrum provides solid and quantitative information in several ways. First, and most important for the present screening protocol, it is possible to judge from the dispersion of resonances whether a protein is well folded. Completely unfolded proteins and, sometimes, molten globules show a typically narrow chemical shift dispersion in the amide region of the NMR spectrum due to the fact that many or all resonances occur at or close to random coil values [11]. Well-folded proteins, on the other hand, normally show a dispersion of the resonances within the  $^1\text{H}$  chemical shift range 6 to 10 ppm. Proteins with only helical secondary structure are sometimes exceptions, but a folded state can in this case still be securely deduced from a dispersion of the side-chain amide resonances of asparagine and glutamine residues in the  $^1\text{H}$  and  $^{15}\text{N}$  chemical ranges of 6.5 to 7.5 and 110 to 114 ppm, respectively. Second, the number of resonances observed in the HSQC spectrum should agree with the expected number of amide–proton pairs in the protein. If this is not the case then one can expect either multiple stable conformations (too many resonances) or proteolytic digestion of the product (too few resonances). Such scenarios are not infrequent but are readily deduced by counting the number of peaks in the HSQC spectrum. Third, the line shapes (sharpness) of the resonances are good indicators of the total molecular weight and thereby of any oligomerization. A qualitative inspection will reveal whether the protein is monomeric or possibly dimeric or whether it is in a higher-order aggregation state, and further quantification can be made by straightforward measurements of the average  $^1\text{H}$  transverse relaxation time (T2) to obtain an estimate of the rotational correlation time [12,13]. Such measurements are normally also part of the optimization of the conditions if it is decided that an NMR structure determination of the protein should be carried out.

To reduce the NMR measuring time, we prefer to concentrate our protein samples rather than to record spectra on lower-concentration samples. Another way to

save time in NMR experiments is to use a cryoprobe. Under the sample conditions currently used, a signal to noise ratio that is four times higher than that with a normal probe can be obtained. This means that the biophysical characterization can be faster or that the concentration of the protein can be lowered, either of which possibilities have their advantages.

## Conclusions

The general procedure for sample preparation described in this article combines established methods for expression, purification, and biophysical characterization of proteins. The amount of soluble protein produced in a small *E. coli* culture can be enhanced using a suitable fusion tag, which then allows for easy one-step purification that can be automated. While CD and NMR spectroscopy by themselves are both fast and highly informative methods for biophysical characterization of proteins, combining the two techniques increases the level of certainty of the obtained results. We show that we get clear information on the structural status of the proteins that we have tested, i.e., whether they are folded or unfolded. We also show that these two spectroscopic techniques can be complemented by fluorescence measurements of added ANS dye to detect potential molten globules.

## Acknowledgments

This work was supported by the Knut and Alice Wallenberg Foundation (KAW) and the EU project “Structural Proteomics in Europe” (SPINE). We thank Susanne van den Berg for construction of expression vectors and Dr. Juergen Schleucher at the University of Umeå for assistance.

## References

- [1] M. Hammarström, N. Hellgren, S. van den Berg, H. Berglund, T. Härd, Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*, *Protein Sci.* 11 (2002) 313–321.
- [2] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, second ed., Cold Spring Harbour Press, Cold Spring Harbour, NY, 1989.
- [3] L.E. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc.* 114 (1992) 10663–10665.
- [4] S. Mori, C. Abeygunawardana, M. O’Neil Johnson, P.C.M. van Zijl, Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation, *J. Magn. Reson. Ser. B* 108 (1995) 94–98.

- [5] G.V. Semisotnov, N.A. Rodionova, O.I. Razgulyaev, V.N. Uversky, A.F. Gripas, R.I. Gilmanshin, Study of the “molten globule” intermediate state in protein folding by a hydrophobic fluorescent probe, *Biopolymers* 31 (1991) 119–128.
- [6] F.C.L. Almeida, G.C. Amorim, V.H. Moreau, V.O. Sousa, A.T. Creazola, T.A. Américo, A.P.N. Pais, A. Leite, L.E.S. Netto, R.J. Giordano, A.P. Valente, Selectively labeling the heterologous protein in *Escherichia coli* for NMR studies: a strategy to speed up NMR spectroscopy, *J. Magn. Reson.* 148 (2001) 142–146.
- [7] H.D. Ou, H.C. Lai, Z. Serber, V. Dötsch, Efficient identification of amino acid types for fast protein backbone assignments, *J. Biomol. NMR* 21 (2001) 269–273.
- [8] A.M. Gronenborn, G.M. Clore, Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy, *Protein Sci.* 5 (1996) 174–177.
- [9] A. Yee, X. Chang, A. Pineda-Lucena, B. Wu, A. Semesi, B. Le, T. Ramelot, G.M. Lee, S. Bhattacharyya, P. Gutierrez, A. Denisov, C.-H. Lee, J.R. Cort, G. Kozlov, J. Liao, G. Finak, L. Chen, D. Wishart, W. Lee, L.P. McIntosh, K. Gehring, M.A. Kennedy, A.M. Edwards, C.H. Arrowsmith, An NMR approach to structural proteomics, *Proc. Natl. Acad. Sci. USA* 99 (2002) 1825–1830.
- [10] T. Otomo, N. Ito, Y. Kyogoku, T. Yamazaki, NMR observation of selected segments in a larger protein: central-segment isotope labelling through intein-mediated ligation, *Biochemistry* 38 (1999) 16040–16044.
- [11] H.J. Dyson, P.E. Wright, Equilibrium NMR studies of unfolded and partially folded proteins, *Nature Struct. Biol.* 5 (NMR Suppl.) (1998) 499–503.
- [12] J. Anglister, S. Grzesiek, H. Ren, C.B. Klee, A. Bax, Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS, *J. Biomol. NMR* 3 (1993) 121–126.
- [13] D. Plaksin, S. Chacko, P. McPhie, A. Bax, E.A. Padlan, D.H. Margulies, A T cell receptor V $\alpha$  domain expressed in bacteria: does it dimerize in solution? *J. Exp. Med.* 184 (1996) 1251–1258.
- [14] M. Wolf-Watz, T. Grundström, T. Härd, Structure and backbone dynamics of apo-CBF $\beta$  in solution, *Biochemistry* 40 (2001) 11423–11432.
- [15] H. Hansson, M.P. Okoh, C.I.E. Smith, M. Vihinen, T. Härd, Intermolecular interactions between the SH3 domain and the proline-rich TH region of Bruton’s tyrosine kinase, *FEBS Lett.* 489 (2001) 67–70.