Journal of Biomolecular NMR (2006) 34: 137–151
DOI 10.1007/s10858-006-0003-0

Article

Rapid determination of protein solubility and stability conditions for NMR studies using incomplete factorial design

Thierry Ducat, Nathalie Declerck, Thierry Gostan, Michel Kochoyan & Hélène Déméné*
A contribution from the Centre de Biochimie Structurale, CNRS UMR 5048/INSERM UMR 554/Universito de Montpellier 1, 29 rue de Navacelles, 34090, Montpellier Cedex, France

Received 1 August 2005; Accepted 23 December 2005

Key words: incomplete factorial design, sample preparation, solubility, stability

Abstract

Sample preparation constitutes a crucial and limiting step in structural studies of proteins by NMR. The determination of the solubility and stability (SAS) conditions of biomolecules at millimolar concentrations stays today empirical and hence time- and material-consuming. Only few studies have been recently done in this field and they have highlighted the interest of using crystallogenesis tools to optimise sample conditions. In this study, we have adapted a method based on incomplete factorial design and making use of crystallisation plates to quantify the influence of physico-chemical parameters such as buffer pH and salts on protein SAS. A description of the experimental set up and an evaluation of the method are given by case studies on two functional domains from the bacterial regulatory protein LicT as well as two other proteins. Using this method, we could rapidly determine optimised conditions for extracting soluble proteins from bacterial cells and for preparing purified protein samples sufficiently concentrated and stable for NMR characterisation. The drastic reduction in the time and number of experiments required for searching protein SAS conditions makes this method particularly well-adapted for a systematic investigation on a large range of physico-chemical parameters.

Abbreviations: BME – beta-mercaptoethanol; DTT – dithiothreitol; EDTA – ethylenediamine tetraacetic acid; FFD – full factorial design; IFD – incomplete factorial design; IPTG – isopropyl-beta-D-thiogalactopyranoside; LB – Luria Bertani; SAS – solubility and stability; SDS-PAGE – sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TRIS – tris-(hydroxymethyl) aminomethane.

Introduction

Remarkable progress has been achieved in recent years for the acquisition and use of NMR data (Pervushin et al., 1997; Tjandra and Bax, 1997; Frydman et al., 2002; Kim and Szyperski, 2003; Kupce and Freeman, 2003). Nevertheless, the limiting step in structural studies of macromolecules often remains the preparation of samples suitable for experimental analysis. The determination of protein structures by NMR spectroscopy still often requires samples that are soluble and stable for several days at millimolar concentrations, at preferably acidic pH and temperature above 30 °C. For this purpose, the determination of protein solubility and stability (SAS) conditions is crucial, yet it remains mostly empirical and hence time- and material-consuming. Very few studies have been conducted in order to develop rapid and general

*To whom correspondence should be addressed. E-mail: helene.demene@cbs.cnrs.fr

methods for optimising NMR samples whereas the need for such methods is growing, especially for high throughput projects (Yee et al., 2002). Most of the work in this field has been devoted to the optimisation of the expression system (Zhou et al., 2001; Shih et al., 2002; Abergel et al., 2003; Folkers et al., 2004). By contrast, increasing efforts have been devoted to the optimisation of crystallisation conditions, the limiting step of protein structure determination by X-ray diffraction methods. From these studies, more rational and systematic approaches of protein crystallogenesis have emerged (Ducruix and Giegé, 1999).

The influence of factors such as temperature, pH, nature and concentration of salts on the solubility diagrams of model proteins has been extensively investigated. Originally, Hofmeister has classified anions and cations according to their ability to precipitate hen egg white proteins (Hofmeister, 1888). It was then proposed (von Hippel and Schleich, 1969) that the structure of proteins is stabilised according to the so-called Hofmeister series, from the least to the most chaotropic ions: sulfate > phosphate > acetate > citrate > tartrate > bicarbonate > chromate > chloride > nitrate≫chlorate > thiocyanate for anions and lithium > sodium > potassium > ammonium > magnesium for cations. Sulfate and phosphate are called lyotropic and tend to stabilise protein structures whereas thiocyanate and chlorate are called chaotropic and tend to denature proteins. The influence of ions on protein solubility also follows the Hofmeister series, though simple and general rules cannot be easily defined (Collins, 2004; Leberman, 1991). Hence, although some theoretical principles governing protein SAS conditions have been elicited, determination of SAS conditions obviously requires a large sampling of nature of salts, concentrations and pH (Bagby et al., 2001). Research work in crystallogenesis has indeed set up sampling methods of the crystallisation space such as the matrix sparse method (Jancarick and Kim, 1991) and incomplete factorial design (Carter and Carter, 1979), aimed at speeding up the crystallisation process.

Since the search of optimised conditions for protein crystallisation and SAS are not conceptually different, the work accomplished in crystallogenesis can be applied for improving the preparation of NMR samples. Bagby et al. (1997) and Lepre and Moore (1998) have indeed, respectively introduced micro-dialysis and hanging drops as experimental techniques to optimise protein solvent conditions for NMR studies. Lindwall et al. (2000) have also introduced a sparse matrix approach for the solubilisation of over-expressed heterologous proteins.

In the present study, we show how a combination of crystallogenesis tools, namely the incomplete factorial design (IFD) and the Hofmeister series can be successfully adapted for the rapid determination of protein SAS conditions for NMR studies. Our aim was to perform NMR studies of two functional domains from the *Bacillus subtilis* regulatory protein LicT (van Tilbeurgh and Declerck, 2001): PRD1 (112 residues) corresponding to the first regulatory domain and CAT-PRD1 (167 residues) comprising the N-terminal RNA binding domain (CAT) linked to PRD1. Both proteins could be overproduced in *E. coli*. In the case of PRD1, most of the protein was rather insoluble in the classical buffers used during the preparative steps (extraction and purification) and we repeatedly failed to obtain a purified sample. In the case of CAT-PRD1, a purified protein sample at millimolar concentration could be prepared, but the protein precipitated at room temperature within a few days, precluding NMR studies. We therefore set up a protocol based on IFD (Carter and Carter, 1979) for fast and rational determination of factors influencing (1) PRD1 solubilisation in cell extracts and (2) the solubility and stability of the purified and concentrated proteins. Compared to a full factorial design, the IFD procedure random sampling drastically reduced the number of experiments required for the finding of optimal conditions. It allowed to rapidly explore the effect of a wide range of pH, nature and concentration of salts in the buffer solutions. Quantification of the physico-chemical parameter factors influencing protein extraction and SAS levels is presented as well as an evaluation of the method on five other proteins.

Materials and methods

Protein purification

The *licT*(1–167) and *licT*(57–167) gene fragments encoding, respectively, the CAT-PRD1 and PRD1 domain from the *Bacillus subtilis* antiterminator LicT (Schnetz et al., 1996) were cloned into a pET15 derivative (Declerck et al., 1999) allowing expression of peptidic domains fused to a

C-terminal His-tag (Leu-G1u-6×His). The fusion proteins were produced in *E. coli* BL21(DE3)-pRep4 and purified by immobilised metal affinity chromatography on a Ni-NTA Superflow resin (Qiagen) followed by gel filtration on a 180 ml Superdex 75 (Pharmacia) column as described in Ducat et al. PlcR is a 35 kDa regulatory protein activating virulence genes in *Bacillus thuringiensis*. PlcR was produced in *E. coil* BL21(DE3)-pRep4 as a His-tagged fusion as described in Slamti and Lereclus (2002). VpG is a 18 kDa Rice Mosaic viral protein (Hebrard et al., 2005). The C-terminal His-tagged protein VpG was produced in *E. coli* M15-pRep4 from a pQE60 derivative (Qiagen). CcpN (18 kDa, regulatory domain alone) and CggR (38 kDa) are *B. subtilis* transcriptional regulators that were produced and purified as His-tagged proteins as described previously (Doan and Aymerich, 2003; Servant et al., 2005).

Sampling and preparation extraction solutions

The incomplete factorial set of experiments was generated using the SamBa software (Audic et al., 1997), available on http://igs-server.cnrs-mrs.fr/samba. This software was originally created to optimise the design of experimental sets of crystallogenesis solutions by an incomplete factorial approach. Three factors and 16 factor levels were used to generate the experimental set of buffer solutions. Factors are the solution parameters to be screened and factor levels the different experimental values taken by these parameters. The factors (and factor levels) tested in the present study were the pH (4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8 and 8.5), the nature of salt sampling the anionic Hofmeister series (Na_2SO_4, KSCN and NaCl) and salt concentration (50, 250, 500 and 750 mM). Thirty-six extraction buffers for the IFD extraction test were randomly generated with the SamBa software and sample the 108 conditions corresponding to the full factorial design (FFD) (Table 1). IFD insures a balanced design for all factor levels. Buffer solutions were prepared at 100 mM with sodium acetate buffer from pH 4.5 to 5.5, with sodium phosphate buffer from pH 6.0 to 7.0 and with TRIS buffer from pH 7.5 to 8.5. Each buffer solution was pH adjusted after salt addition. A slightly different set-up of experiments was developed to screen the extraction solutions of the protein VpG and PlcR. Four factors (and 16 factor levels) were tested: the pH (5, 6, 7, 8), the nature of salt sampling the anionic and cationic Hofmeister series (Na_2SO_4, KSCN, NaCl and KCl), the salt concentration (50, 250, 500 and 1000 mM) and the presence of additive (No, glycerol, NP40 and Triton X-100). Half to one day is needed to prepare such sets. The time to perform sampling, micro-lysis, SDS-PAGE gels and analysis is short and all these steps can be joined in only one day.

Protein extraction and determination of extraction coefficients β_i

The method for the rapid determination of extraction conditions using cell micro-lysis and detection by SDS-PAGE was adapted from Lindwall et al. (2000). Cells from one litre LB culture of BL21(DE3) transformants over-producing the protein of interest were resuspended in 40 ml of a buffer solution containing 10 mM TRIS, pH 7.5 and 100 mM NaCl. Cells overexpressing PRD1 (or VpG and P1cR) were pelletted as 37 (33) 1-ml aliquots in micro-centrifuge tubes and resuspended in 1 ml of one of the 36 (32) extraction buffers or denaturating solution (1% SDS, 2 M urea, 2 mM BME, 2.5% glycerol, 15 mM TRIS, pH 7.5). Cells were then disrupted by sonication on bed ice and centrifuged at $16{,}000 \times g$. An equal volume (30 ± 1.5 µl) of each of the 37 (33) supernatants was run on four (three) 15% SDS gels, each including the extract in denaturating solution, which served as reference. After staining in Coomassie blue, gels were scanned and the NIH Image 1.62 software (available on http://rsb.info.nih.gov) was used to measure the intensity of the bands corresponding to PRD1 (or VpG and PlcR). A constant area box was used to measure the intensity of the background (I_0) and of the targetted protein bands (I_X) on the gels. For each extraction solution (X), the extraction score was calculated using the intensity of the extracted protein band observed in denaturating condition (I_R) as a reference: $S_{obs}(\%) = (I_X - I_0/I_R - I_0) \times 100$. The contribution of each factor level on the solubilisation of PRD1 was evaluated by the minimisation of the energy function E (Carter, 1999):

Table 1. The 36 extraction solutions and extraction scoxes[a] observed for PRD1

Extraction solution	pH	Salts	[Salt] (mM)	S_{obs} (%)[a]	Extraction solution	pH	Salts	[Salt] (mM)	S_{obs} (%)[a]
1	7.0	NaCl	250	59	19	4.5	NaSCN	50	10
2	6.5	Na_2SO_4	750	36	20	7.5	NaSCN	50	64
3	8.0	NaCl	500	47	21	6.0	NaCl	750	56
4	6.5	Na_2SO_4	500	35	22	6.5	NaCl	50	34
5	8.5	NaCl	750	46	23	7.0	NaSCN	50	44
6	5.5	NaSCN	750	14	24	4.5	Na_2SO_4	250	0
7	5.5	NaCl	250	24	25	5.0	NaSCN	750	8
8	7.5	Na_2SO_4	250	58	26	6.5	NaSCN	250	46
9	7.5	NaCl	500	52	27	8.0	Na_2SO_4	250	47
10	5.0	NaSCN	250	4	28	4.5	NaSCN	500	5
11	8.0	NaSCN	750	82	29	5.0	Na_2SO_4	500	20
12	8.5	Na_2SO_4	50	22	30	8.5	NaSCN	500	55
13	5.0	NaCl	50	8	31	8.5	NaCl	250	41
14	7.0	NaSCN	500	58	32	6.0	Na_2SO_4	500	28
15	6.0	NaSCN	250	35	33	4.5	NaCl	750	9
16	6.0	NaCl	50	36	34	5.5	Na_2SO_4	50	18
17	7.0	Na_2SO_4	750	37	35	5.5	NaCl	500	22
18	8.0	Na_2SO_4	50	48	36	7.5	Na_2SO_4	750	36

[a]PRD1 extraction score deduced from the scanning of the SDS gels presented in Figure 1 (mean of three triplicates). The estimated error is about 10% of the extraction score value.

$$E = \sum_{k=1}^{N} (S_{\text{calc}}^{k} - S_{\text{obs}}^{k})^2 \quad (1)$$

with:

$$S_{\text{calc}}^{k} = \beta_0 + \sum_{i=1}^{N_I} \beta_i F_{i,k} \quad (2)$$

where $F_{i,k}$ represents the factor level i in the k-th experiment, N_I the number of factor levels (here $N_I = 16$), and β_0 the mean value of the observed extraction scores S_{obs}^{k} over the N experiments (here $N = 36$ or 32). Equation 2 represents a ''model-free'' approach which does not require any knowledge of the dependence (linear, quadratic or even more complex) of the extraction score on factor levels. This modeling choice amounts to mathematically represent each factor level as a separate factor. The factor levels i are coded 1 (presence) or 0 (absence) in $F_{i,k}$ for each extraction solution k. For instance, the 250 mM NaCl, pH 7 condition (Table 1, experiment 1) is coded by 16 $F_{i,k}$, three equal to unity (those corresponding to the factor level pH 7, the factor level NaCl and the factor level 250 mM) and all others equal to zero. The contribution of factor levels i on the extraction score S_{obs}^{k} is then evaluated by the extraction of β_i coefficients which minimise the energy function E. The β_i coefficient corresponds to the average amount by which the presence of factor level i raises or lowers the score S_{obs}^{k} from the mean score β_0. Minimisation was performed using the non-linear minimisation routine of the R software, available on http://www.R-project.org. The starting values for β_i coefficients were set up to 0. Due to the non-numerical type of coding of factor levels in $F_{i,k}$, the error ε in the derived β_i values can only be *a posteriori* estimated from the variance between the observed and calculated scores (Press et al., 1987):

$$\varepsilon = \sqrt{\sum_{k=1}^{N} (S_{\text{calc}}^{k} - S_{\text{obs}}^{k})/(N - N_I - 1)} \quad (3)$$

The -1 term in the denominator accounts here for the diminution of the number of freedom degrees due to the introduction of the minimised energy in ε estimation.

A mean value score $S_{\text{mean, obs}}(i)$ for each factor level i was also calculated as the mean of all scores over the experiences k where the factor level was present:

$$S_{\text{mean,obs}}(i) = \sum_{F_{i,k}\neq 0, k=1,N} S^{k}_{\text{obs}}.$$

Determination of solubility and stability (SAS) conditions

The hanging drop method described by Lepre and Moore (1998) was adapted to search for protein SAS conditions in crystallisation plates. Purified protein samples were dialysed in 10 mM sodium phosphate buffer pH 6.4, 50 mM NaCl, 2 mM DTT, 100 μM benzamidine, 100 μM EDTA and concentrated using Ultrafree centrifugal filter units (Millipore) until a precipitate started to appear. The precipitate was eliminated by centrifugation and protein concentration was estimated by measuring absorbance at 280 nm (0.4 mM (CAT-PRD1), 0.6 mM (PRD1), 0.2 mM (PlcR), 0.8 mM (CcpN), 0.1 mM (YqfL) and 0.2 mM (CggR)). The 144 buffer solutions corresponding to the FFD set were prepared in 100 mM succinic acid – NaOH (pH 5.0 and 5.5), sodium phosphate (pH 6.0, 6.5 and 7.0) or TRIS–HCl (pH 7.5, 8.0 and 8.5) and contained either NaCl, NaBr, LiCl, KCl, KSCN or Na_2SO_4 at 100, 200, 400 or 600 mM (only LiCl at 600 mM and pH 7.0 was not fully soluble and therefore not tested). Buffer pH was adjusted after salt addition. Hanging drop experiments were set up in 24-well crystallisation plates (Nextal), a system that proved to be more convenient for our purpose than the Linbro plates originally proposed (Lepre and Moore, 1998). Wells were filled with 1 ml of the buffer solutions. Protein aliquots of 2 μl were deposited on the cap, mixed with 1 μl of the well solution and the well was hermetically sealed with the cap. The drops were observed with a binocular microscope after 1, 2.5 and 5 days of incubation at 18 °C and the presence/absence of precipitate was noticed. For the CAT-PRD1 and PRD1 proteins, analysis of the results was done in full factorial design (144 conditions corresponding to 6 pH values and 6 salts at 4 concentrations) as well as in IFD using three independent sampling sets of 36 buffer solutions (generated with the SamBa software). A precipitation score S_{obs} of 0 or 100 was given to each buffer solution for the presence or the absence of visible precipitate in the drop. For

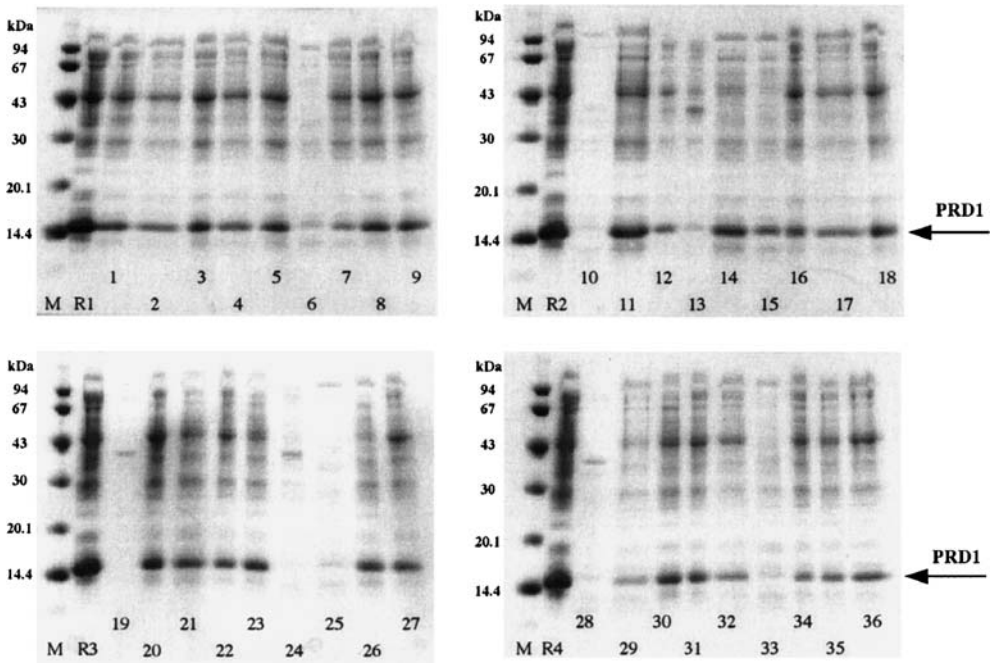

Figure 1. PRD1 solubilisation by each of the 36 buffer solutions observed on 15% SDS-PAGE gels. The number below the lanes indicates the buffer solution number in Table 1. The M lanes correspond to low molecular weight markers. The R lanes correspond to protein solubilisation under denaturating conditions and were used as references for a 100% extraction. The arrows indicate the position of the band corresponding to PRD1.

CcpN, CggR and PlcR, the SAS screening was performed using only 24 conditions (Table S4, 4 salts at 4 concentrations and 6 pH corresponding to 96 conditions in FFD) but 3 values were distinguished for the precipitation score (S_{obs}): 0 (heavy precipitate), 50 (light precipitate) or 100 (no visible precipitate after 2 days). Solubility coefficients β_i were then calculated as described above.

Results

Optimisation of protein extraction conditions

Protein extracts from *E. coli* cells over-producing the PRD1 regulatory domain were prepared in different extraction buffers and analysed by SDS-PAGE. The experimental sets of buffer solutions are presented in Table 1. The number of buffer solutions required for exploring the effect of nine pH values (ranging from 4.5 to 8.5) and three sodium salts (NaCl, Na_2SO_4 and NaSCN) at four concentrations (50, 250, 500 and 750 mM) was reduced from 108 in the FFD (3 factors and 16 factor levels) to only 36 with the application of the IFD. The extraction levels of PRD1 obtained with each of the 36 buffer solutions can be visualised in Figure 1 and the estimated extraction scores are listed in Table 1. For the purpose of this study, it is interesting to note that the total amount of soluble proteins as well as the relative solubility of the different proteins, including PRD1, greatly depends on buffer composition. Total protein recovery is similar in lanes 12 and 13 but the amount of PRD1 is much higher in buffer 12 (pH 8.5, 50 mM Na_2SO_4) than in buffer 13 (pH 5, 50 mM NaCl) whereas it is the contrary for an unknown protein of about 40 kDa visible on the SDS gel. For this unknown protein, the best extraction conditions are obviously at low pH values (lanes 13, 19, 24 and 28), although the total protein yield, this protein excepted, is extremely low under these conditions. For PRD1 extraction, the most efficient buffer appears to be buffer 11 (score 82%, pH 8, 750 mM NaSCN), which improves the extraction score by at least 30% compared with classical buffers such as buffer 31 (score 41%, pH 8.5, 250 mM NaCl) or buffer 1 (score 59%, pH 7.0, 250 mM NaCl).

A quantitative statistitical analysis was performed in order to evaluate the contribution of each factor to protein recovery. Mean values $S_{mean,obs}(i)$ and extraction coefficients β_i were calculated for each variable factor level i for the three proteins (Figure 2 and Tables S1, S2 and S3). The sign of β_i indicates the positive or negative contribution of the factor level i on extraction. The results represented as in Figure 2 enable to unambiguously highlight the relative contribution and the best level of each factor. In the case of PRD1 extraction, the major contributing factor is the pH of the extraction buffer and pH 8 is the optimum. The nature and concentration of salts have little influence on PRD1 recovery level. NaSCN appears however, slightly better than NaCl (and Na_2SO_4 significantly worse) and high salt concentrations (500–750 mM) are better than low salt concentrations. On this basis the best predicted conditions for PRD1 extraction should be at pH 8 and 750 mM NaSCN, which correspond to buffer 11, indeed the best extraction buffer tested.

A similar IFD approach was used for optimising the extraction conditions of two other proteins, P1cR and VpG. A reduced set of experimental buffers was designed in which only four pH values (pH = 5, 6, 7, 8) were tested but which allowed to investigate the effect of four different salts (NaCl, Na_2SO_4, NaSCN and KCl) at four salt concentrations (50, 250, 500 and 1000 mM), as well as the presence of additives that are often used to solubilise and/or stabilise proteins (0.2% NP40, 0.2% Triton X-100, and 10% glycerol). The incomplete factorial design of the 32 experimental conditions listed in Table 2 corresponds to a full factorial design of 256 conditions (4 factors and 16 factor levels). The PlcR and VpG extraction scores observed with each buffer are given in Table 2. For PlcR, extraction levels greater than those obtained under the denaturating condition serving as reference were observed with buffers 7, 10, 20, 23, 30 that all contain a mild detergent (either Triton or NP40). Nevertheless, a high yield of soluble PlcR could also be obtained using buffer 5 (score 99%, pH 6, NaCl 1 M, glycerol) or even buffer 17 containing no additive (score 94%, pH 8, 250 mM Na_2SO_4). In case of VpG, the protein was mostly produced as inclusion bodies that were poorly solubilised by the reference denaturating solution containing 2 M urea. Although the yield of soluble VpG remained low under all conditions, it could be significantly improved by different combination of factors. It was nearly doubled with buffer 16

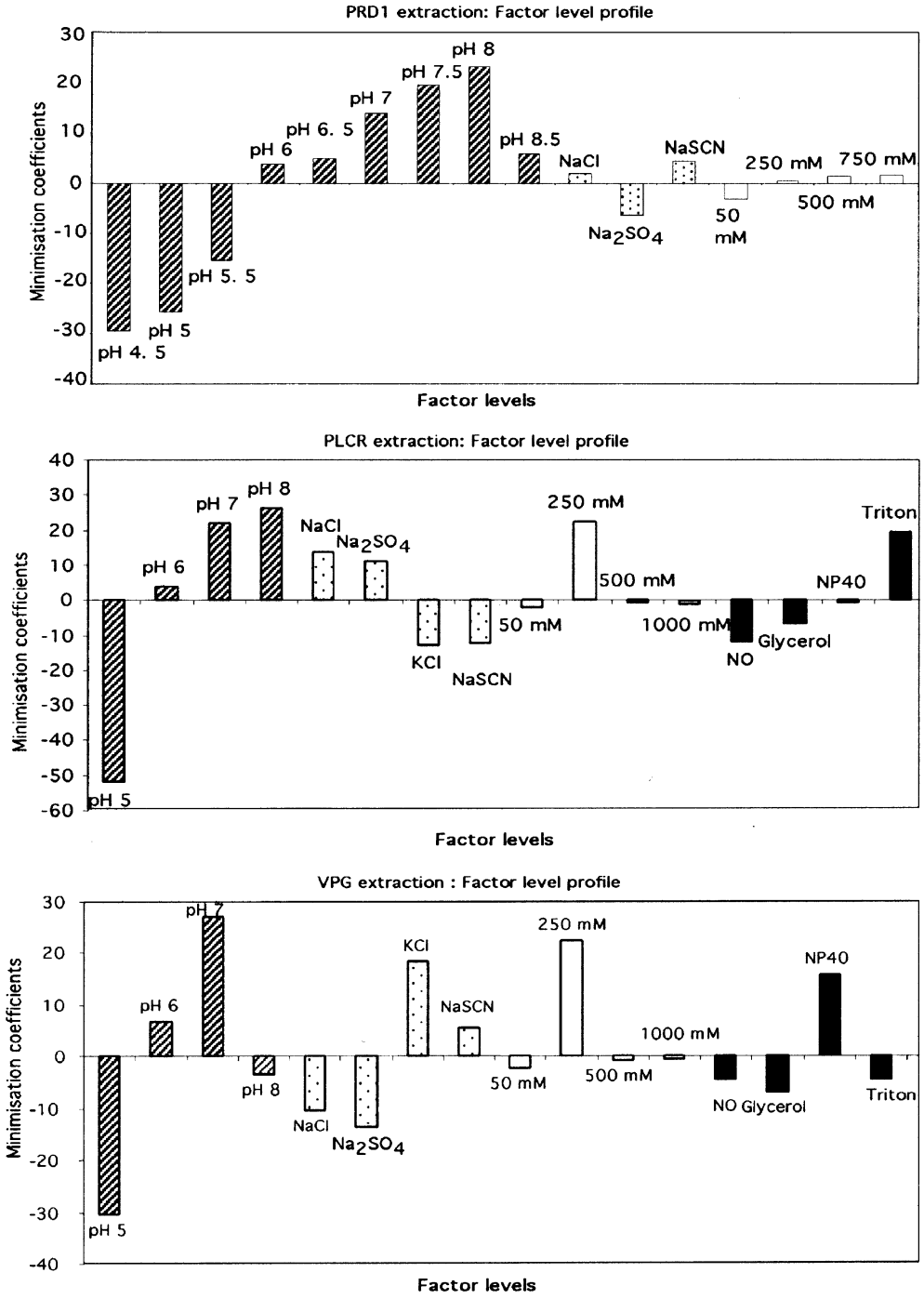

Figure 2. Representation of the extraction coefficients β_i from the minimisation of the sum of the squared differences between observed and calculated scores for the three proteins PRD1, PlcR and VpG. The factor levels are represented for pH (striped sticks), for salt (point filled sticks), for salt concentration (empty sticks) and additives (black sticks). Positive extraction coefficients indicate a positive influence of the factor level on the extraction. At the opposite, negative extraction coefficients indicate negative influence of the factor level on the extraction.

Table 2. The 32 extraction solutions and extraction scores[a] observed for VpG and PlcR

Exp. sol.	pH	salts	[Salt] (mM)	Additive[b]	% PlcR	% VpG	Ext. sol.	pH	Salts	[Salt] (mM)	Additive[b]	% PlcR	% VpG
1	6	NaSCN	1000	NO	20.0	42.0	17	8	Na_2SO_4	250	NO	94.0	53.6
2	7	NaSCN	1000	G	60.0	113.0	18	8	NaSCN	500	NP40	91.5	53.6
3	5	NaSCN	500	T	0.5	15.0	19	6	Na_2SO_4	250	G	80.5	40.0
4	8	NaCl	50	NO	84.5	37.0	20	6	Na_2SO_4	500	T	123.5	46.0
5	6	NaCl	1000	G	99.5	27.5	21	5	Na_2SO_4	1000	NP40	57.5	40.0
6	6	KCl	50	T	78.0	25.0	22	5	KCl	1000	T	13.0	29.0
7	7	NaCl	250	T	120.5	47.5	23	8	NaCl	1000	NP40	111.5	25.0
8	8	Na_2SO_4	1000	T	68.5	117.5	24	5	Na_2SO_4	50	G	45.0	2.5
9	5	NaSCN	50	NO	35.0	14.5	25	7	KCl	500	NP40	71.5	126.0
10	7	Na_2SO_4	50	NP40	103.0	39.0	26	6	NaCl	500	N0	71.0	69.00
11	5	NaCl	250	NP40	2.0	36.0	27	7	NaSCN	250	G	69.0	185.0
12	7	Na_2SO_4	500	NO	67.0	30.0	28	6	NaSCN	50	NP40	27.0	99.0
13	8	KCl	500	G	80.0	38.0	29	6	KCl	250	NP40	40.5	180.0
14	5	KCl	250	NO	0.5	18.5	30	7	NaCl	50	T	118.5	74
15	8	KCl	50	G	50.5	27.5	31	8	NaSCN	250	T	140.5	96.5s
16	7	KCl	1000	NO	74.0	192.0	32	5	NaCl	500	G	11.5	88.5

[a]Extraction score deduced from the scanning of the SDS gels (mean of three triplicates). The estimated error is about 10% of the extraction score value.[b]NO, no additive; G, glycerol; T, Triton; NP40, NP40.

(pH 7, 1000 mM KCl, no additive), 27 (pH 7, 250 mM NaSCN, glycerol) or 29 (pH 6, 250 mM KCl, NP40).

Quantitative analysis of the P1cR and VpG extraction scores was then performed as described above (Figure 2). In the case of PlcR, the analysis highlights the presence of 0.2% Triton as a favourable factor, but also indicates that the pH, the nature and concentration of salt are important contributing factors that need to be adjusted. On the basis of this analysis, it can be predicted that the best extraction conditions for P1cR would be at pH 8, 250 mM NaCl and 0.2% Triton. Since some salts appeared highly detrimental (KCl and NaSCN) to P1cR extraction, we wondered whether this could biase the calculation of the exact contribution of salt concentration. We therefore tested P1cR extraction under the full range of NaCl concentration (50, 250, 500 and 1000 mM) at pH 8 in the presence of 0.2% Triton. Regardless of NaCl concentration, the extraction scores at pH 8 and 0.2% Triton were found to be higher than with the best previously tested conditions (Figure S1, supplementary material). Under these conditions, a concentration of 50 mM NaCl appeared slightly more favourable than the predicted 250 mM concentration. We also tested that at a salt concentration of 50 mM, Triton X-100 was a better additive than NP40 and NaCl a better extraction salt than Na_2SO_4 (data not shown), as finely predicted by the factor profile.

For VpG extraction, our analysis suggests that the pH is the most important contributing factor, pH 7 being the most favourable value. The quantitative analysis further highlights that the nature of salt, the concentration and the presence of an additive also significantly influence the extraction score. Based on this profile, the best conditions are predicted to be: pH 7, 250 mM KCl, 0.2% NP40. Since, as with P1cR, some salts appear to be very detrimental to VpG extraction, we have also tested the full range of KCl concentration (50, 250, 500 and 1000 mM) at pH 7 in presence of 0.2% NP40. The best extraction score was observed at 1 M KCl (Figure S1, supplementary material), with a value 70% higher than with the best condition tested in the incomplete factorial design. We have also tested that, as finely predicted by the factor level profile, pH 7 scores better than pH 8 for this range of conditions (Figure S1).

Determination of solubility and stability (SAS) conditions for CAT-PRD1 and PRD1

The next limiting step in NMR studies of proteins, once produced and purified in sufficient amounts, is to prepare samples that can sustain several days of analysis at high concentration, moderate

temperature and preferably acidic pH. In order to rapidly determine the optimal buffer conditions for PRD1 and CAT-PRD1, we have set up an experimental test procedure using crystallisation plates. The influence of buffer composition on the precipitation of concentrated protein aliquots in hanging drops was tested for six pH values and six different salts at four concentrations. For both PRD1 and CAT-PRD1, a systematic set of 144 experiences corresponding to the full factorial design was performed and analysed. The results were then analysed in three independent sets of 36 experiments generated by automatic sampling of the full experimental set in order to test the robustness of the incomplete factorial design.

The precipitation matrix observed for CAT-PRD1 and PRD1 in crystallisation plates are reported in Table 3a and b, respectively. In case of CAT-PRD1, precipitation occurred systematically below pH 6, except in the presence of at least 200 mM Na_2SO_4. Above pH 6, the protein remained soluble in all buffers for at least 5 days at 18 °C. By contrast, PRD1 precipitation occurred in most buffer conditions tested. The pH range of no precipitation is narrow (essentially around neutrality) and greatly depends on salt nature and concentration. Precipitation occurred in the presence of LiCl or KCl for the whole range of pH, even at low concentration, whereas NaCl up to 400 mM leads to no precipitation around neutral pHs. At pH 8.5, the protein remains soluble only in the presence of high concentration of KSCN (200–600 mM).

The analysis of the precipitation matrices in full or incomplete factorial design is presented in Figure 3 and Table S3 in Supplemental Data. The 36 buffer solutions composing each of the three incomplete samplings generated by the SamBa software are given as additional materials (Table S2). In case of CAT-PRD1 (Table S3a and Figure 4), analysis of the precipitation matrix in full or incomplete sampling gives very similar results. It can readily be deduced that pH is the most important factor determining CAT-PRD1 SAS and that the favourable range extends from pH 6 to at least 7.5. The slight positive effect of Na_2SO_4 is highlighted in 2 out of the 3 samplings and the best SAS conditions predicted by the factorial samplings are: pH 6–7.5, 600 mM (sampling 1) or 400 mM (sampling 3) Na_2SO_4 versus pH 6–7.5, 200 mM Na_2SO_4 (full design). For the PRD1 protein (Table S3b and Figure 4), the individual contribution of each factor level is less accurately evaluated from incomplete factorial design. For instance, β_i coefficients calculated from the full experimental set or from sampling 2 suggest an opposite effect for pH 7. Nevertheless, the most beneficial levels are unambiguously revealed by any of the sampling set analyses. The best predicted SAS conditions for PRD1 are: pH 7 and 100 mM NaCl (sampling 1), pH 7.5 and 100 mM NaCl (sampling 2), pH 7.5 and 100–200 mM NaCl

Table 3a. CAT-PRD1 precipitation matrix[a]

	5.0		5.5	6.0	6.5	7.0	7.5
100 mM	NaCl	Na_2SO_4	NaCl	Na_2SO_4			
	LiCl	NaBr	LiCl	NaBr			
	KCl	KSCN	KCl	KSCN			
200 mM	NaCl	Na_2SO_4	NaCl				
	LiCl	NaBr	LiCl	NaBr			
	KCl	KSCN	KCl	KSCN			
400 mM	NaCl	Na_2SO_4	NaCl				
	LiCl	NaBr	LiCl	NaBr			
	KCl	KSCN	KCl	KSCN	KSCN		
600 mM	NaCl	Na_2SO_4	NaCl				
	LiCl	NaBr	LiCl	NaBr		*	
	KCl	KSCN	KCl	KSCN	KSCN		

[a]Salts keep always the same position in the subsection defined by row and column conditions. The indication of salt name indicates precipitation of CAT-PRD1 in the corresponding solution condition. Conversely, no indication of salt name indicates no precipitation of CAT-PRD1. The * condition corresponds to not tested condition due to incompatibility solution conditions.

Table 3b. PRD1 precipitation matrix[a]

	6.0		6.5		7.0		7.5		8.0		8.5	
100 mM	NaCl	Na_2SO_4							NaCl	Na_2SO_4	NaCl	Na_2SO_4
	LiCl	NaBr	LiCl	NaBr	LiCl		LiCl		LiCl	NaBr	LiCl	NaBr
	KCl	KSCN	KCl	KSCN			KCl		KCl	KSCN	KCl	KSCN
200 mM	NaCl	Na_2SO_4		Na_2SO_4					NaCl	Na_2SO_4	NaCl	Na_2SO_4
	LiCl	NaBr	LiCl	NaBr	LiCl		LiCl		LiCl	NaBr	LiCl	NaBr
	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	
400 mM	NaCl	Na_2SO_4				Na_2SO_4		Na_2SO_4	NaCl	Na_2SO_4	NaCl	Na_2SO_4
	LiCl	NaBr	LiCl	NaBr	LiCl	NaBr	LiCl		LiCl	NaBr	LiCl	NaBr
	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	
600 mM	NaCl	Na_2SO_4	NaCl	Na_2SO_4	NaCl	Na_2SO_4	NaCl	Na_2SO_4	NaCl	Na_2SO_4	NaCl	Na_2SO_4
	LiCl	NaBr	LiCl	NaBr	*	NaBr	LiCl	NaBr	LiCl	NaBr	LiCl	NaBr
	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	KSCN	KCl	

[a]As defined in Table 3a.

or Na_2SO_4 (sampling 3). They are all very similar to those deduced from the full factorial set analysis: pH 7–7.5 and 100 mM NaCl.

Validation of the IFD method on other proteins

In order to validate the IFD approach, we have applied the SAS screening procedure to three other proteins currently studied in our laboratory, namely PlcR, CggR, and CcpN. For the sake of rapidity and convenience, a reduced experimental set of 24 buffers was designed to explore the effect of 6 pH and 4 salts at 4 concentrations (Table S4). The experiment was set in a single 24-wells crystallisation plate and the presence/absence of precipitate in the drops was observed after 12–24 h at 18 °C. Scoring was improved by introducing an intermediate score ($S_{obs} = 50$) when only slight precipitation was observed in the drop. This scoring method was particularly necessary in case of PlcR which was essentially insoluble in all tested buffers. Results of the quantitative analysis are given as supplementary materials (Table S4 and Figure S2). For each protein, a solubility test was then performed under the best predicted condition of pH and at different concentrations of the best predicted salt (Table S5). In all cases, the maximal solubility score was reached under the best conditions of pH and salt deduced from the IFD analysis. For PlcR, best pH and salt nature were correctly predicted, but not the best concentration (100 vs. 200 mM).

Heteronuclear NMR characterisation of PRD1 and CAT-PRD1

The SAS conditions determined using non-labelled protein preparations were applied to prepare NMR samples of the ^{15}N-labelled CAT-PRD1 and PRD1. The maximal concentration that could be reached without apparent precipitation was 2 mM for ^{15}N CAT-PRD1 prepared in phosphate buffer pH 6.4, 50 mM Na_2SO_4, (50 mM was chosen as a compromise between probe sensitivity and protein stability). Only a slight precipitate was observed after 48 h of NMR data acquisition at 32 °C whereas, under the standard conditions (pH 6.5, 200 mM NaCl) the protein heavily precipitated after a few hours of acquisition. The ^{1}H–^{15}N HSQC spectrum recorded under optimised SAS conditions (Figure 4a) showed that the protein was properly folded. Double- and triple-labelled samples of CAT-PRD1 were thus prepared for 2D- and 3D NMR experiments and enabled backbone resonance assignment of this 40 kDa dimeric domain from LicT (Ducat et al., 2002). In the case of PRD1, optimisation of the SAS conditions enabled us to prepare concentrated samples of the ^{15}N protein at up to 1.5 mM. However, no high quality NMR spectra of PRD1 could still be recorded. As illustrated in Figure 4b, the ^{1}H–^{15}N HSQC spectra were always of poor quality, showing essentially large resonance peaks that suggested the presence of protein aggregates. The NMR study of the PRD1 domain alone, which

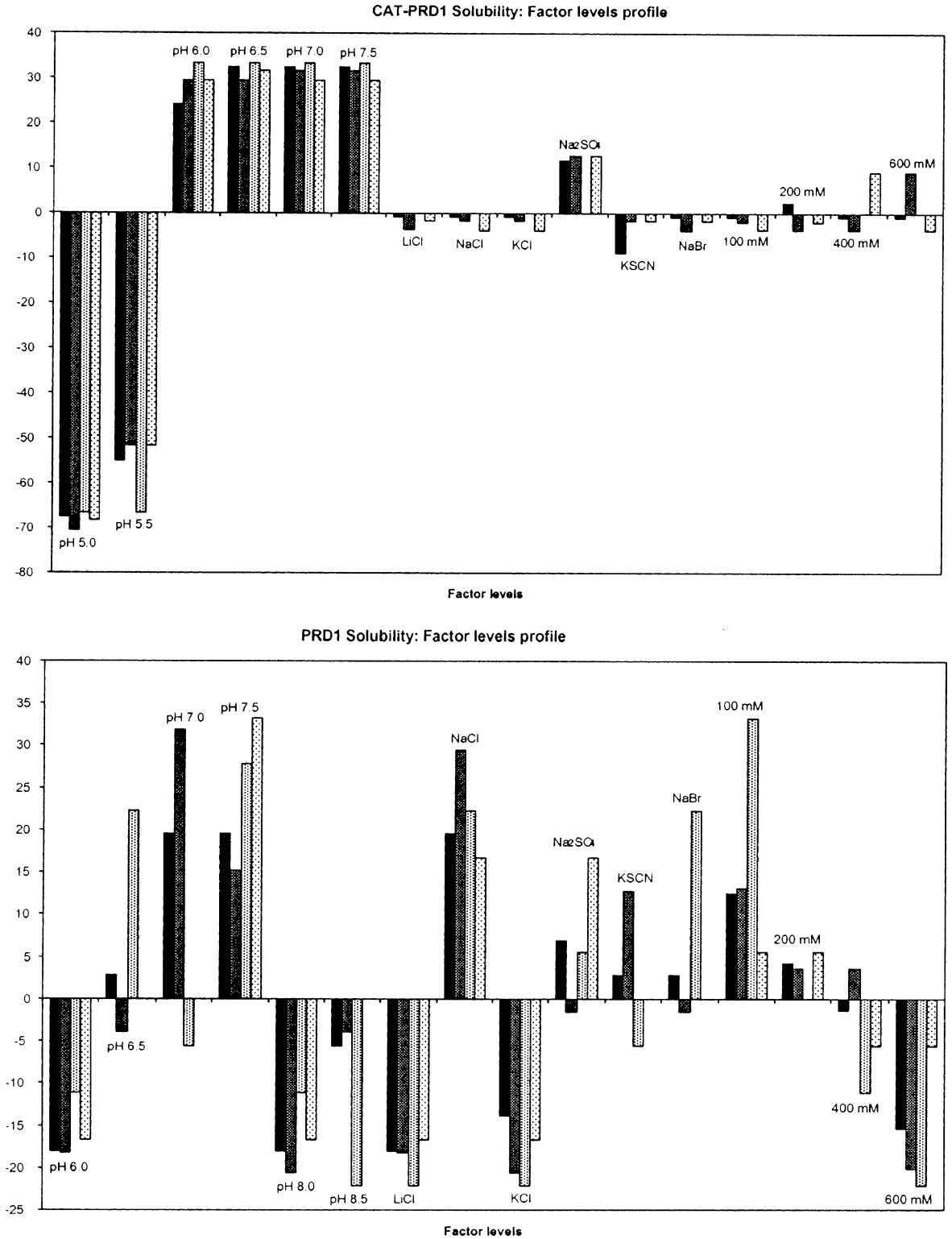

Figure 3. Factor profile of CAT-PRD1 (top) and PRD1 (bottom) solubility coefficients β_i from full factorial design (dark sticks) and from incomplete factorial design, sampling 1 (dark grey sticks), sampling 2 (grey sticks) and sampling 3 (light grey sticks).

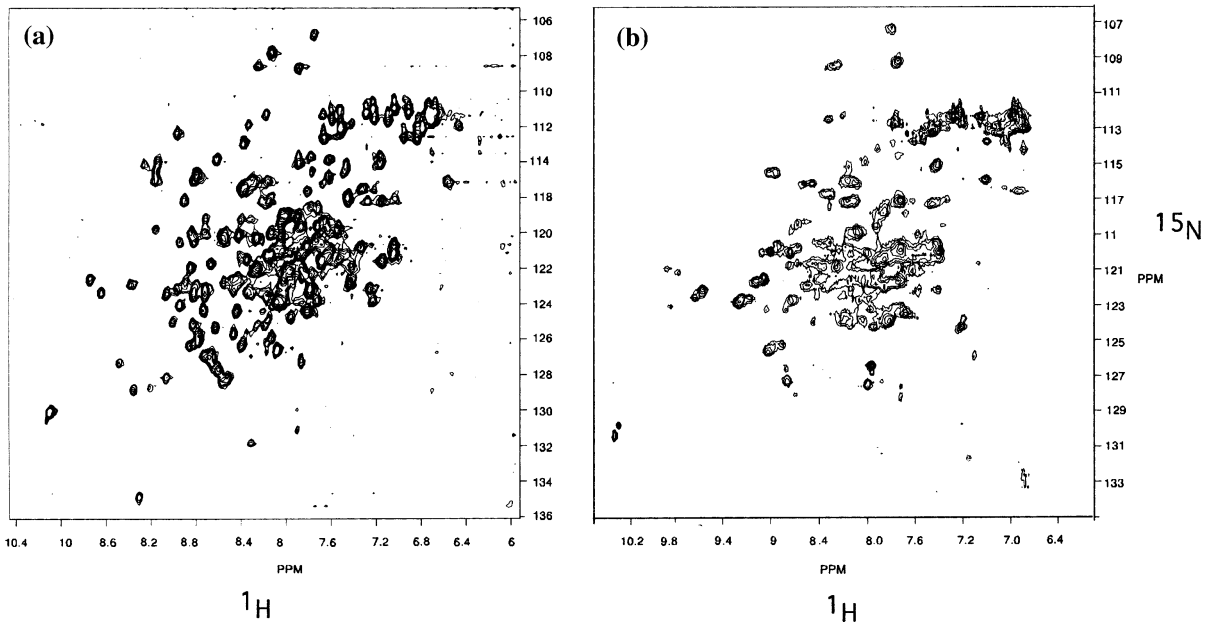

Figure 4. ^{15}N–^{1}H HSQC spectra of Lict-CAT-PRD1 (a) and PRD1 (b) proteins recorded at 32 °C on a 800 and 600 MHz spectrometer, respectively. Both proteins were dissolved in their optimal buffer as determined in this study.

was undertaken in order to facilitate the resonance assignment of CAT-PRD1, was thus rapidly dropped.

Discussion

Incomplete factorial design (Carter and Carter, 1979) and the sparse matrix method (Jancarick and Kim, 1991) have been used for over a decade and have proven their efficiency for sampling the crystallisation space of macromolecules. The extensive and systematic use of the sparse matrix method was mostly due to the success of commercial crystallisation screening kits based on this method. Elaboration of experimental sets by the sparse matrix method is biased towards successful conditions in related problematics whereas sampling by IFD ensures a balanced representation of all factors and factor levels for any kind of experimental procedures without previous knowledge on the protein behaviour. The number of experiments required to explore combinations of variable factors is drastically reduced, speeding up the sampling of a high number of conditions. The simple minimisation model used in the present study describes calculated scores S_{calc} as a linear combination of independent factor levels, i.e.,$\sum \beta_i F_{i,k}$. The power of this "model free" approach comes from the quantification of the influence of each factor level by the coefficients β_i that enable to easily highlight the main effects and major factor levels without any hypothesis about the mathematical dependence of observed scores on variables. Hence, prediction of the best conditions is straigthforward with this approach. By contrast, in the original and subsequent applications – protein crystallisation (Carter and Carter, 1979), protein expression (Abergel et al., 2003) and protein refolding (Chen and Gouaux, 1997; Wu et al., 2004), the factors were allowed to possess only two levels and the calculated score was a linear combination of factors $\sum \beta_i F_i$. As a result, prediction of best conditions required two distinct steps: the identification of the main factors by a two-level IFD followed by a step of extensive screening of the previously identified important factors to obtain the precise mathematical dependence of the energy function on these factor levels.

The present study demonstrates that our sampling method based on incomplete factorial design can be successfully applied for the rapid finding of optimal conditions of protein preparation for NMR experiments. The concentration

figure appears to be the least best determined, by contrast, pH, salt nature and additive composition are always exactly predicted. It is particularly well suited when quantitative scores can be estimated, as described here for the determination of extraction conditions of the PRD1, VpG and PlcR, proteins. For these three proteins, IFD allowed to identify the main favourable factors and to predict conditions that indeed improved protein extraction. Only a couple of days are sufficient to test and analyse the results from the different extraction solutions used for cell microlysis followed by protein detection on three/four SDS-gels, whereas testing the complete range of conditions required for a full factorial analysis is hardly conceivable.

IFD was also successfully applied for optimising solubility and stability conditions of the PRD1 and CAT-PRD1 proteins. For PRD1, whose SAS conditions are very narrow, the best factor levels identified in the full sampling (pH 7 or 7.5, 100 mM NaCl) could be accurately extracted from the analysis of incomplete factorial samplings. In the case of CAT-PRD1, the solubility is predominantely dependent on one factor (pH) as revealed by the full and incomplete factorial samplings. Because of the pre-eminence of pH as stability factor, the slight positive of Na_2SO_4 could not always be detected by incomplete factorial analysis (sampling 2). The profile of solubility of CAT-PRD1 as measured by the 0–1 score appears very flat and independent of the salt nature even in the full factorial plan, since with this type of scoring, the beneficial effect of Na_2SO_4 is apparent in only 2 among 28 potential conditions (200 and 400 mM Na_2SO_4, pH 5.5). Sampling 2 does not sample these particular conditions and as a result, all salts appear equivalent and prediction is therefore useless. The "all-or-nothing" scoring method used here (presence/absence of precipitate in the drop) is indeed not well adapted to find the optimal conditions when a protein is essentially either soluble or insoluble in all tested buffers.

Validation of the IFD method on other proteins (PlcR, CggR and CcpN) showed that the set of solutions that can be used for SAS screening can be conveniently reduced and still enable to accurately determine the optimal conditions of salt and pH. The analysis was significantly improved by using a more quantitative scoring method, distinguishing between heavy or light precipitates. Introducing more intermediate values to quantify the precipitation level could maybe further improve the scoring method but it could also lead to more biased results since protein precipitation evolves with time and quantification by the experimentator can be rather subjective. Other quantitative approaches could also be used to monitor the amount of soluble protein by SDS-PAGE, dot blotting or turbidity measurements (Trésaugues et al., 2004; Vincentelli et al., 2004). Setting up experiments with crystallisation plates is however particularly easy, rapid, and requires little amount of protein sample.

The simplicity of the method and the limited amount of materials required also allows to search SAS conditions for a specific time (by observing the plates after different incubation times) or temperature (by placing the plates in an incubator) required for NMR experiments. For the purpose of this study, we note that investigating a large range of pH and non-protonated salts was sufficient to establish conditions where proteins are soluble for at least a few days. Most of the studies published to date introduce the presence of diverse additives that are not compatible with NMR studies (Armstrong et al., 1999; Trésaugues et al., 2004; Vincentelli et al., 2004). The use of Arg and Asp at high concentrations (≥50 mM) has also been recently described for NMR studies of soluble proteins (Golovanov et al., 2004). Our experience with our DRX Bruker spectrometers equiped with Cryoprobes is however, that the sharp signals of such small molecules at so high concentrations can significantly worsen the multidimensional spectra of proteins because of t_1 noise, making the use of the costly deuterated corresponding compounds mandatory. Whether and how precipitation in hanging drops is related to the solubility and stability of the protein is a debatable question. In case of PRD1 for instance, the absence of visible precipitates was not indicative of the presence of aggregates in the samples, precluding further NMR experiments. Protein behaviour could also be investigated by dynamic light scattering (DLS) using the incomplete factorial approach proposed here.

The systematic search of SAS conditions can provide valuable information on protein behaviour regarding physico-chemical parameters. CAT-PRD1 was found to remain fully soluble at pH above its theoretical pI value (6.04) regardless of salt nature and concentration. Below pH 6, the

effect of the different salts tested follows the Hofmeister series: the most lyotropic anion (sulfate) is the most stabilising. By contrast, the range of pH where PRD1 (p$I = 6.07$) remains soluble is very narrow (pH 7–7.5) and the effect of salts cannot be related to the Hofmeister series. Moreover, the protein behaved differently during the extraction procedure than once purified: a high concentration of the chaotropic anion SCN^- was the most efficient for PRD1 solubilisation from *E. coli* extracts whereas it was destabilising on the purified protein. This illustrates the difficulty in establishing general rules regarding protein SAS conditions and the need of systematic approaches such as the one we propose here for their fast determination.

Concluding remarks

With the advent of high throughput methodologies, it has become urgent to develop rapid and efficient methods for preparing samples suitable for structural studies. The present paper is aimed to call the attention of NMR spectroscopists to the possibility of adapting methods already successfully used by crystallographers for this purpose. Here we developed an experimental procedure based on IFD for determining the best extraction and SAS conditions for a given protein. We believe that the systematic search of optimised conditions by such methods could considerably reduce the waste of time and materials required for obtaining good quality spectra and assessing the feasibility of a structural study by NMR. For both CAT-PRD1 and PRD1, we succeeded in finding SAS conditions that enabled us to prepare samples sufficiently concentrated and stable for multidimensional NMR experiments. However, whereas high quality spectra were obtained for CAT-PRD1 in spite of its size (169 residues forming a 40 kDa homodimer), the NMR spectra recorded for PRD1, although smaller, were always of poor quality because of protein aggregation. Obtaining a concentrated soluble protein sample is just the first limiting step for NMR structural studies but it is not a guarantee of success.

Electronic Supplementary Material is available at http://dx.doi.org/10.1007/s10858-006-0003-0

Acknowledgements

We would like to thank Stéphane Aymerich for providing us CAT-PRD1 and PRD1 encoding plasmids. We also thank Jean-Baptiste Reizer for stimulating discussions. The Fondation pour 1a Recherche Médicale and the French Ministère de la Recherche are greatly acknowledged for financial support.

References

Abergel, C., Coutard, B., Byrne, D., Chenivesse, S., Claude, J.B., Deregnaucourt, C., Fricaux, T., Gianesini-Boutreux, C., Jeudy, S., Lebrun, R., Maza, C., Notredame, C., Poirot, O., Suhre, K., Varagnol, M. and Claverie, J.M. (2003) *J. Struct. Funct. Genomics* **4**, 141–157.

Armstrong, N., De Lencastre, A. and Gouaux, E. (1999) *Prot. Sci.* **8**, 1475–1483.

Audic, S., Lopez, F., Claverie, J.M., Poirot, O. and Abergel, C. (1997) *Proteins* **29**, 252–257.

Bagby, S., Tong, K.I., Liu, D., Alattia, J.R. and Ikura, M. (1997) *J. Biomol. NMR* **10**, 279–282.

Bagby, S., Tong, K.I. and Ikura, M. (2001) *Methods Enzymol.* **339**, 20–41.

Carter, C.W. Jr. and Carter, C.W. (1979) *J. Biol. Chem.* **254**, 12219–12223.

Carter, C.W. Jr. (1999) In *Crystallisation of Nucleic Acids and Proteins: A Practical Approach*, Ducruix, A. and Giegé, R. (Eds.), Oxford University Press, New York, pp. 75–120.

Chen, G.Q. and Gouaux, E. (1997) *Proc. Natl. Acad. Sci.* **94**, 13431–13436.

Collins, K.D. (2004) Methods, **34**, 300–311.

Declerck, N., Vincent, F., Hoh, F., Aymerich, S. and van Tilbeurgh, H. (1999) *J. Mol. Biol.* **294**, 389–402.

Doan, T. and Aymerich, S. (2003) *Mol. Microbiol.* **47**, 1709–1721.

Ducat, T., Declerck, N., Kochoyan, M. and Déméné, H. (2002) *J. Biomol. NMR* **23**, 325–326.

Ducruix, A. and Giegé, R. (1999) In *Crystallisation of Nucleic Acids and Proteins: A Practical Approach*, Ducruix, A. and Giegé, R. (Eds.), Oxford University Press, New York, pp. 1–16.

Folkers, G.E., Buuren, B.N.M. and Kaptein, R. (2004) *J. Struct. Funct. Genomics* **5**, 119–131.

Frydman, L., Scherf, T. and Lupulescu, A. (2002) *Proc. Natl. Acad. Sci.* **99**, 15858–15862.

Golovanov, A.P., Hautbergue, G.M., Wilson, S.A. and Lian, L.-Y. (2004) *J. Am. Chem. Soc.* **126**, 8933–8939.

von Hippel, P.H. and Schleich, T.W. (1969) In *Structure and Stability of Biological Macromolecules*, Fashmann, G. and Timasheff, S. (Eds.), Marcel Dekker, New York, pp. 417–574.

Hebrard, E., Pinel-Galzi, A., Catherinot, V., Labesse, G., Brugidou, C. and Fargette, D. (2005) *J. Virol.* **79**, 4409–4414.

Hofmeister, F. (1888) *Arch. Exp. Pathol. Pharmackol.* **24**, 247–260.

Jancarick, J. and Kim, S.H. (1991) *J. Appl. Cryst.* **24**, 409–411.

Kim, S. and Szyperski, T. (2003) *J. Am. Chem. Soc.* **125**, 1385–1393.

Kupce, E. and Freeman, R. (2003) *J. Am. Chem. Soc.* **125**, 13958–13959.
Leberman, R. (1991) *FEBS Lett.* **284**, 293–294.
Lepre, C.A. and Moore, J.M. (1998) *J. Biomol. NMR* **12**, 493–499.
Lindwall, G., Chau, M., Gardner, S.R. and Kohlstaedt, L.A. (2000) *Protein Eng.* **13**, 67–71.
Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 12366–12371.
Press, W.H., Flannery, B.P., Teukolski, S.A. and Vetterling, W.T. (1987) In *Numerical Recipes in Fortran: The Art of Scientific Computing*, Cambridge University Press, Cambridge, pp. 498–539.
Schnetz, K., Stülke, J., Gertz, S., Krüger, S., Krieg, M., Hecker, M. and Rak, B. (1996) *J. Bacteriol.* **178**, 1971–1979.
Servant, P., Le Coq, D. and Aymerich, S. (2005) *Mol. Microbiol.* **55**, 1435–1451.
Slamti, L. and Lereclus, D. (2002) *Embo J.* **21**, 4550–4559.
Shih, Y.P., Kung, W.M., Chen, J.H., Yeh, C.H., Wang, A.J.H. and Wang, T.F. (2002) *Protein Sci.* **11**, 1714–1719.
van Tilbeurgh, H. and Declerck, N. (2001) *Curr. Opin. Struct. Biol.* **11**, 685–693.
Tjandra, N. and Bax, A. (1997) *Science* **278**, 1111–1114.
Trésaugues, L., Collinet, B., Minard, P., Henckes, G., Aufrère, R., Blondeau, K., Liger, D., Zhou, C.-Z., Janin, J., van Tilbeurgh, H. and Quevillon-Cheruel, S. (2004) *J. Struct. Funct. Genomics* **5**, 195–204.
Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C and Bignon, C. (2004) *Prot. Sci.* **13**, 2782–2792.
Wu, H., Gao, J., Sharif, W.D., Davidson, M.K. and Wahls, W.P. (2004) *Protein Expr. Purif.* **38**, 136–144.
Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharyya, S., Gutierez, P., Denisov, A., Lee, C.H., Cort, J.R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., MacIntosh, L.P., Gehring, K., Kennedy, M.A., Edwards, A.M. and Arrowsmith, C.H. (2002) *Proc. Natl. Acad. Sci.* **99**, 1825–1830.
Zhou, P., Lugovskoy, A.A. and Wagner, G. (2001) *J. Biomol. NMR* **20**, 11–14.