Protein Expression and Purification 82 (2012) 352–359

Contents lists available at SciVerse ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Single pH buffer refolding screen for protein from inclusion bodies

Bruno Coutard [a], Etienne G.J. Danchin [b], Rachid Oubelaid [c], Bruno Canard [a], Christophe Bignon [a,*]

[a] Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257 CNRS and Aix-Marseille University, 163, Avenue de Luminy, Case 932, 13288 Marseille, Cedex 09, France
[b] INRA UMR 1301, CNRS UMR 6243, UNSA, 400 route des Chappes, F-06903 Sophia-Antipolis, France
[c] Sanofi, 9 quai Jules Guesde 94400 Vitry-sur-Seine, France

ARTICLE INFO

Article history:
Received 7 November 2011
and in revised form 21 January 2012
Available online 4 February 2012

Keywords:
Recombinant protein
Refolding
Isoelectric point
Proteome
Screen

ABSTRACT

We previously reported the set up of an automated test for screening the refolding of recombinant proteins expressed as inclusion bodies in *Escherichia coli* [1]. The screen used 96 refolding buffers and was validated with 24 proteins, 70% of which remained soluble in at least one buffer. In the present paper, we have analyzed in more detail these experimental data to see if the refolding process can be driven by general rules. Notably, we found that proteins with an acidic isoelectric point (pI) refolded in buffers the average pH of which was alkaline and conversely. In addition, the number of refolding buffers wherein a protein remained soluble increased with the difference between its pI and the average pH of the buffers in which it refolded. A trend analysis of the other variables (ionic strength, detergents, etc.) was also performed. On the basis of this analysis, we devised and validated a new refolding screen made of a single buffer for acidic proteins and a single buffer for alkaline proteins.

Introduction

Escherichia coli (*E. coli*) is widely used for producing recombinant proteins. Unfortunately, the processing of heterologous proteins by the bacterium can be weak [2], and over-expression often leads to incorrectly or incompletely folded proteins. This folding incompleteness favors intermolecular hydrophobic interactions which result in protein precipitation in the form of inclusion bodies (IB)[1].

During the last two decades, techniques have been devised to refold IB (reviewed in [3,4]).

More recently, screening tests have been set-up for rapidly evaluating the efficiency of refolding buffers and refolding techniques using only a small aliquot of protein [1,5–10]. Commercial screens have also been made available (QuickFold [AthenaES], Pro-Matrix [Perbio], FoldIt [Hampton research], Protein Refolding Screen Kit [Molecular Dimensions Limited], iFOLD [Novagen], BioAssay Protein Refolding Kit [BIOMOL]). A symmetrical approach called "reverse screen" has also been proposed to screen for refolding additives [11].

Finally, a data base intended for recording successful refolding conditions has been created which will undoubtedly prove valuable for setting up refolding screens based on a sparse matrix approach [12].

Since refolding screens are generally composed of numerous buffers, proteins often refold in more than one buffer. In that case, it is necessary to perform a trend analysis of the experimental data to find out the best buffer composition for refolding the protein [13].

In a previous report, we described an automated 96-well IB refolding screen [1]. Using this screen, 70% of tested proteins remained soluble in at least one refolding buffer. Since the aim of the paper was only to show the usefulness of refolding screens in structural genomics, no trend analysis was performed to assess the influence of each variable or variable state of the screen. To rationalize the screen efficiency and potentially further improve it we performed a trend analysis of the experimental data of the original paper. Following the indications provided by the trend analysis, the number of refolding buffers was reduced from 96 to 1. This paper describes the rational of this evolution.

Materials and methods

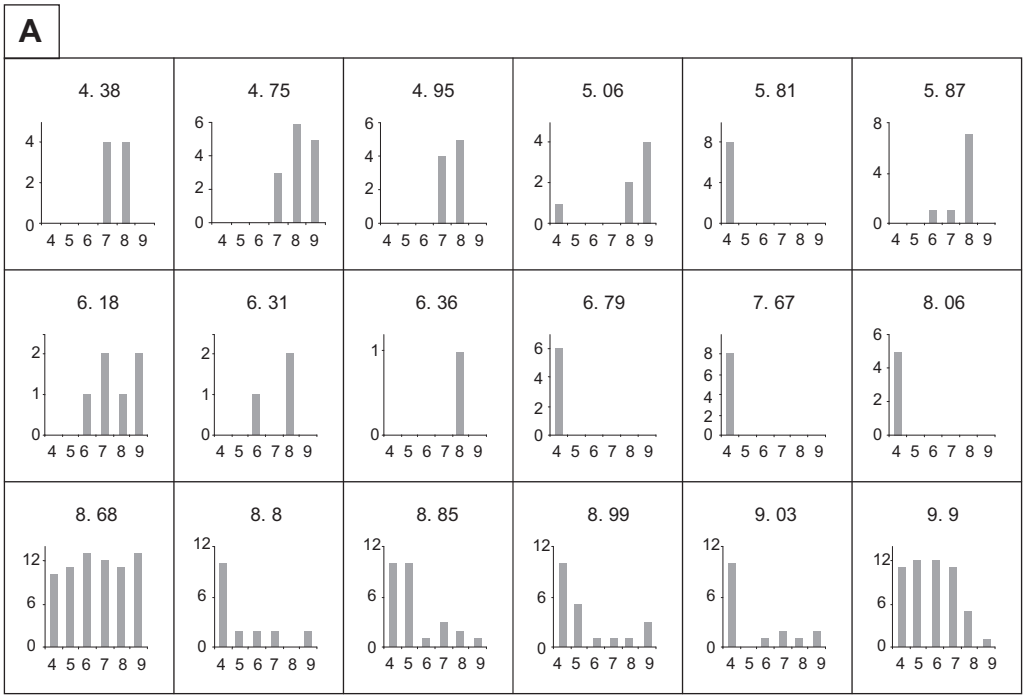
Calculation of the mean value of refolding pH (Fig. 1B)

Throughout, a refolding buffer in which a protein remains soluble is called "positive buffer". The number of positive buffers at a given pH (Fig. 1A) was multiplied by the corresponding pH value. The resulting values obtained for each pH were summed, and this sum was divided by the total number of positive buffers for the considered protein. Since each pH value was equally represented in the screen, an additional correction factor taking into account

* Corresponding author.
E-mail address: bignon@afmb.univ-mrs.fr (C. Bignon).

[1] *Abbreviations used:* IB, inclusion bodies; MES, 2-morpholino ethane sulfonic acid; MOPS, 4-morpholino propane sulfonic acid; CHES, 2-(cyclohexylamino) ethane sulfonic acid; CAPS, 3-(cyclohexylamino)-1-propane sulfonic acid; TMAO, trimethylamine *N*-oxide; NDSB, non detergent sulfobetaine; FAEA, feruloyl esterase A; DTT, dithiothreitol.

1046-5928/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2012.01.014



B

Target	MW	pI	Mean of refolding pH*	|pI – *|	Number of buffers
Rv1399c	36	4.38	7.5	3.12	8
Rv1208	37	4.75	8.14	3.39	14
Rv1564c	84	4.95	7.56	2.61	9
SPINE 5	23	5.06	8	2.94	7
Rv323c	27	5.81	4	1.81	8
Rv2392	30	5.87	7.67	1.80	9
αfTM	55	6.18	7.67	1.49	6
Rv2391	66	6.31	7.33	1.02	3
Rv1373	40	6.36	8	1.64	1
Rv1515c	36	6.79	4	2.79	6
Rv2045c	59	7.67	4	3.67	8
Rv1523	40	8.06	4	4.06	5
SPINE23	23	8.68	6.6	2.08	80
SPINE 21	52	8.8	5.22	3.58	27
Rv3487c	29	8.85	5.26	3.59	21
SPINE 10	23	8.99	5.4	3.59	18
SPINE 22	53	9.03	5.38	3.66	16
SPINE 63	23	9.9	5.81	4.09	52

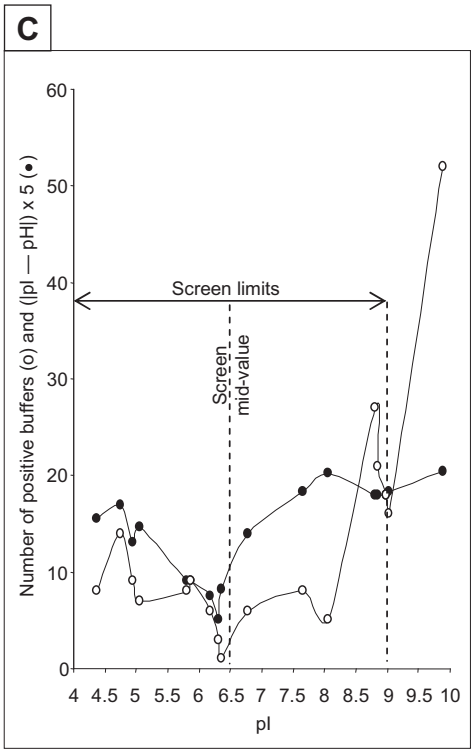


Fig. 1. Number of positive buffers as a function of refolding pH. (A), results are ordered from the most acidic (pI4.38, top left) to the most alkaline (pI9.9, bottom right) protein. Each histogram is the number of positive buffers (*Y* axis) at a given pH. The *X* axis is the pH scale (from 4 to 9) of the refolding screen. (B), Target proteins are ordered by increasing pI from top to bottom. MW, molecular weight of target proteins (in kDa). Mean of pH: see Materials and Methods for the calculation method. |pI − ∗| is the absolute value of the difference between pI and mean of pH. Number of buffers is the total number of positive buffers. (C), graphic presentation of B.

the bias introduced by the relative over- or under-representation of one pH value with regards to the others was not required. For example, Rv1399c (pI 4.36) remained soluble in 4 pH 7 buffers and 4 pH 8 buffers (Fig. 1A). The mean refolding pH value was: $(4 \times 7 + 4 \times 8)/(4 + 4) = 7.5$. The same calculation was done for all proteins, and the results were reported as a function of protein pI in Fig. 1B columns 3 and 4.

Statistical analysis of refolding data (Fig. 4)

For each variable (for example, the variable "pH"), the total number of positive buffers was divided by the number of occurrences of the considered variable state (for example, the variable state "pH 4") in the refolding plate to compensate for the bias eventually introduced by the over- or under-representation of this

variable state in the refolding plate of this particular state with regard to the others. For example,

- in the variable "pH", the variable state "pH 4" is used in 12 out of the 96 wells of the refolding plate (12 occurrences).
- The total number of positive buffers for the variable state "pH 4" is calculated as follows: protein with pI 5.06 (Fig. 1A) remains soluble in 1 pH 4 buffer (Fig. 1A), protein with pI 5.81 remains soluble in 8 pH 4 buffers (Fig. 1A), etc. In total, pH 4 is positive 1 time for protein with pI 5.06, + 8 times for protein with pI 5.81, + etc. = 89 times for all the proteins tested (Fig. 1A).
- The result reported in Fig. 4 for the variable state "pH 4" of the variable "pH" is 89 (total number of positive buffer for pH 4)/12 (number of occurrences of pH 4 in the refolding plate) = 7.42.

Expression and purification of recombinant proteins

Coding sequences were PCR amplified using the following primers: GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATGCCACCATGAAACATCACCATCACCATCAC-[22 first bases of coding sequence], and GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTA-[22 last bases of coding sequence]. PCR products were sub-cloned by recombination into prokaryotic expression vector pDEST14 (Invitrogen). Rosetta(DE3)pLysS cells (Novagen) were transformed with expression constructs, and transformants were selected on ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) (AC) plates. Isolated colonies were used to individually seed 1 ml AC-LB medium in 96-wells deep-wells. Cells were grown overnight at 37 °C in a shaking incubator (750 rpm). The next day, 4 ml AC-LB medium were seeded in 24-well deep-wells with 150 μl preculture. At OD_{600} = 0.5, 0.5 mM IPTG was added and recombinant proteins were allowed to express for 3 h at 37 °C with shaking. The deep-wells were centrifuged for 5 min at 2800*g*, and the culture medium discarded. Cell pellets were re-suspended in 750 μl of 50 mM Tris pH 8, 0.3 M NaCl, 0.1% Triton X100, 5% glycerol, 1 mM EDTA, 0.25 mg/ml lysozyme, and then frozen at −80 °C for 30 min. After thawing, DNAse (10 μg/ml) and $MgSO_4$ (20 mM) were added, and lysates were incubated for 30 min at 25 °C under 200 rpm shaking. After sonicating twice for 1 min on ice, deep-wells were centrifuged for 30 min at 2800*g* at 4 °C. Supernatants were discarded, 1 ml of 50 mM Tris pH 8, 25 mM imidazole, 0.3 M NaCl, 1 M urea, 2% Triton X100 was added to each well, and pellets were resuspended by vortexing. After spinning for 30 min at 2800*g*, supernatants were discarded and 200 μl of 50 mM Tris pH 8, 0.3 M NaCl, 8 M guanidinium chloride were added to each well. Pellets were redissolved by pipetting. Solubilized IB were mixed by shaking at 400 rpm for 5 min at room temperature with 100 μl of a 50% Chelatin Sepharose Fast Flow (GE Healthcare) suspension pre-equilibrated in the same buffer. Unbound material was removed by filtering under vacuum through 96-well filter plates (Novagen). Using on-line vacuum aspiration, Ni-beads were washed in place twice with 1 ml/well of 50 mM Tris pH 8, 25 mM imidazole, 0.3 M NaCl, 8 M urea. Recombinant proteins were then eluted in a 96-well plate located below the filter plate with 200 μl/well of 50 mM Tris pH 8, 500 mM imidazole, 0.3 M NaCl, 8 M urea. Purity of eluted proteins was checked by SDS–PAGE.

Protein refolding

For testing the efficiency of acidic and alkaline refolding buffers (Table 1), a 100 μl aliquot of each affinity-purified denatured recombinant protein was individually loaded in a dialysis button (Hampton Research). The button was closed with a 5 kDa cut-off dialysis membrane, and then incubated in 50 ml refolding buffer overnight at 4 °C under continuous agitation.

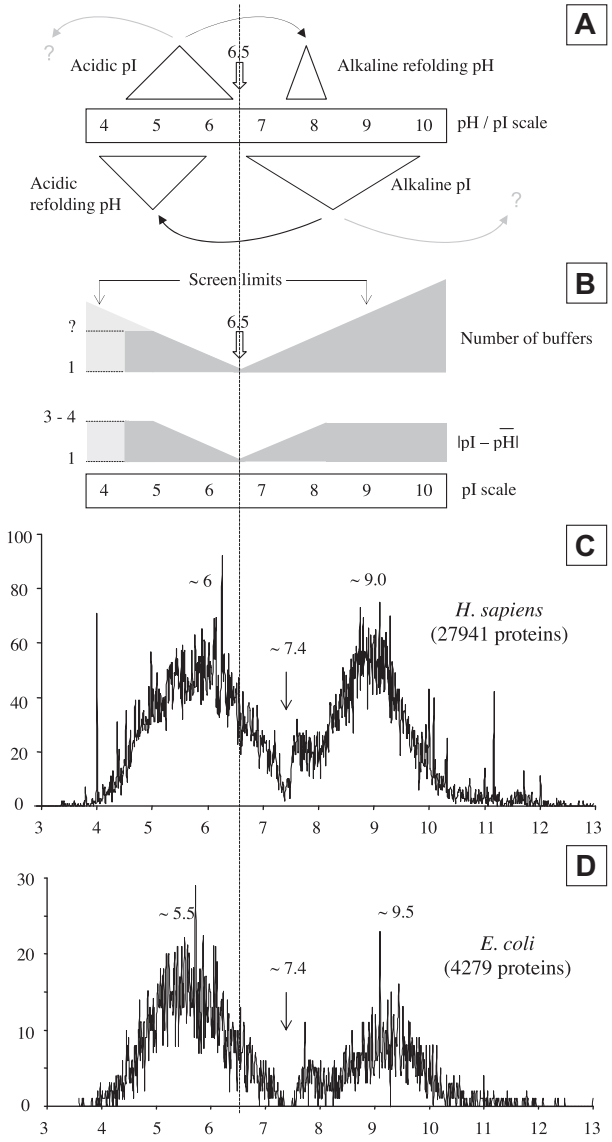


Fig. 2. Schematic representation of the relation between mean refolding pH, number of positive buffers, and protein pI in our set of test proteins and in two proteomes. (A), Acidic proteins refold in alkaline buffers and conversely (black arrows). Because of their odd behaviors, proteins Rv323c and SPINE23 were omitted. The possibility for acidic and alkaline proteins to respectively refold in acidic and alkaline buffers is indicated by gray arrows pointing to a question mark. (B), Schematic and superimposed drawing of the number of positive buffers and of |pI − mean of refolding pH| pointing out their common lowest point. The pale gray area on the left with a question mark indicates a speculated increase of the number of positive buffers for proteins with pI at and beyond the acidic limit of the refolding screen. Values on the left were deduced from Fig. 1B. Note that in contrast with |pI − mean of refolding pH|, the number of positive buffers cannot be described as a plateau, and the higher value is therefore assigned a question mark. (C), Distribution of the number of proteins in *H. sapiens* proteome as a function of their pI. (D), Distribution of the number of proteins in *E. coli* proteome as a function of their pI.

To calculate the percentage of refolded protein, 20 μl of dialyzed protein was saved. The remaining volume of dialyzed protein was filtered under vacuum through 96-well 0.22 μm filter plates (Millipore, MultiScreen GV 0.22 μm), and 20 μl of the filtrate was up taken for protein assay. The amount of protein present in the same volume (20 μl) of dialyzed protein before and after filtration was measured by adding 60 μl of water and 20 μl of Bio-Rad protein assay reagent, and then measuring the OD_{590}. The refolding

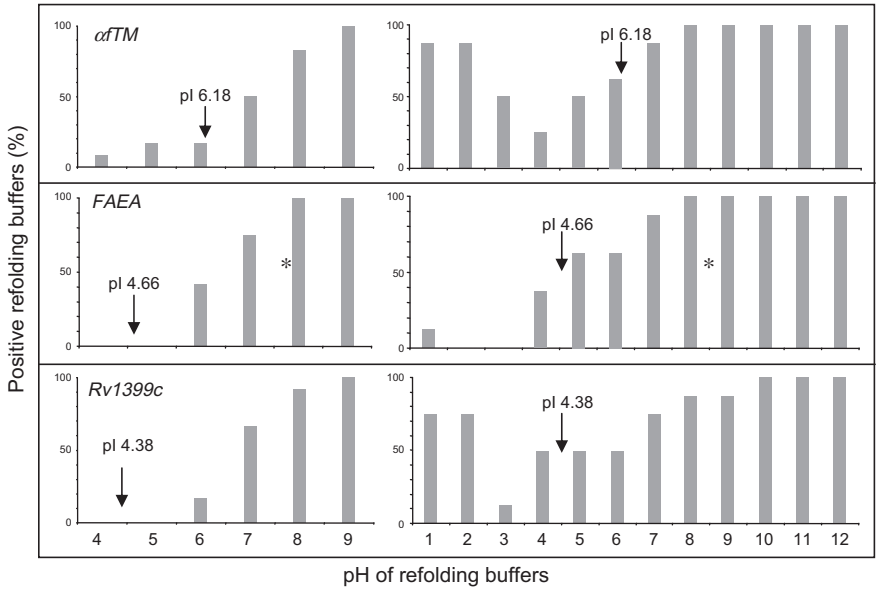


Fig. 3. Comparison of two 96-well refolding plates with different pH ranges. Each histogram represents the percentage of positive buffers (*Y* axis) at a given pH. The results obtained using refolding plates with pH range 4 to 9 [1] (left) and 1 to 12 (right) are reported. The detailed composition of the buffers using pH 1 to 12 is in Table S1. Vertical arrows locate the protein pI on the pH scale. *mean value of refolding pH.

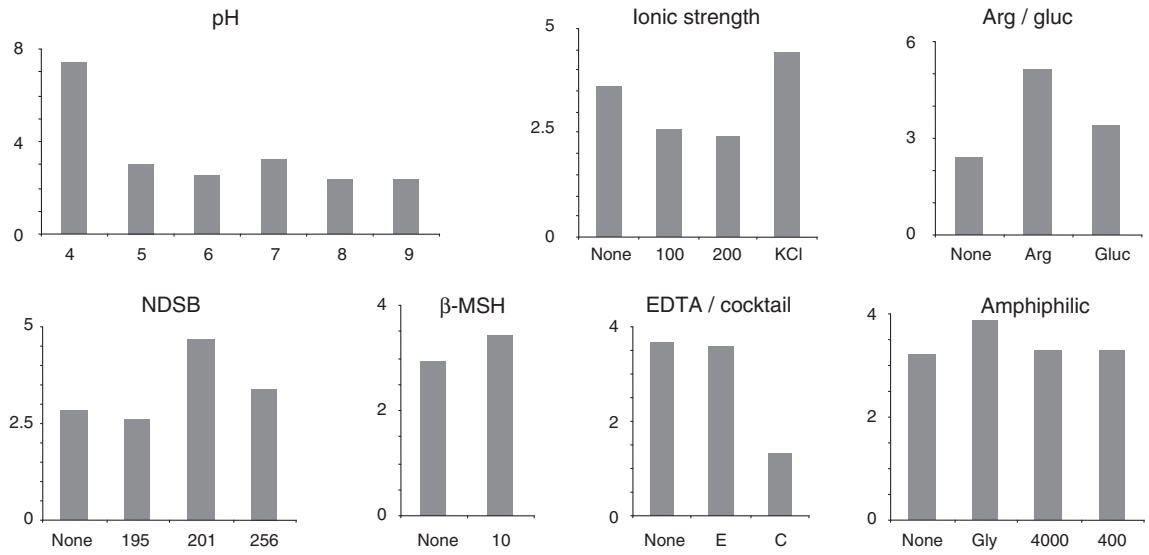


Fig. 4. Trend analysis of refolding data. The weight of each variable used in the original refolding plate was evaluated as described in Materials and Methods using the data provided by the 24 proteins refolded in Ref. [1]. The *Y* axis indicates the total number of positive buffers divided by the number of occurrences in the refolding plate of the considered variable state as described in Materials and Methods. For pH, the *X* axis is the pH scale (from 4 to 9) of the refolding screen. Ionic strength: 100, 100 mM NaCl; 200, 200 mM NaCl; KCl, 100 mM KCl. Arg/gluc: Arg, 800 mM arginine; Gluc, 500 mM glucose. NDSB: 195, NDSB 195; 201, NDSB 201; 256, NDSB 256. β-MSH, beta mercaptoethanol; 10, 10 mM. EDTA/cocktail: E, 1 mM EDTA (ethylene-diamine-tetra-acetic acid); C ("cocktail"), 50 μM of each of the following: NADH, Thiamine HCl, biotin, $CaCl_2$, $MgCl_2$, $CuSO_4$, $ZnCl_2$, $CoSO_4$, ADP, $NiCl_2$. Amphiphilic: Gly, 20% glycerol; 4000, polyethylene glycol 4000; 400, polyethylene glycol 400.

efficiency was defined as the ratio OD_{590} (filtered)/OD_{590} (not filtered), and expressed as percentage.

For solubility assay (Fig. 3), Rv1399c, FAEA and αfTM were processed as described [14].

For activity assay (Fig. S1), Rv1399c and αfTM were refolded by 1/20 dilution of the denatured protein in the alkaline refolding buffer (see Results section for details). After refolding, protein concentrations were determined by measuring the OD_{280} against the refolding buffer supplemented with 1/20 volume of denaturing buffer. Known amounts of refolded Rv1399c or αfTM were incubated in 500 μl of 10 mM Tris pH 8, 100 mM NaCl, 1 mM para-nitrophenyl butyrate (SIGMA N-9876) (Rv1399c) or 4-nitrophenyl α-L-fucopyranoside (SIGMA N3628) (αfTM) for 10 min at 37 °C. The same mixture was incubated in parallel in the absence of enzyme. The amount of product was estimated by measuring the OD_{410} obtained in the presence of enzyme and then subtracting from this value the OD_{410} obtained in the absence of enzyme. This latter result was then divided by the amount of added enzyme.

The refolding plate using pH 1–12 buffers (Fig. 3 and Table S1) was devised using an incomplete factorial approach. Parameters and parameter values were the followings. (1) Buffers (50 mM): sodium maleate pH 1, Glycine–HCl pH 2, sodium citrate pH 3, sodium acetate pH 4, sodium acetate pH 5, 2-morpholino ethane sulfonic acid (MES) pH 6, 4-morpholino propane sulfonic acid (MOPS) pH

Table 1
Protein solubility after refolding using the one buffer screen.

Target[a]	MW	pI	Refolding buffer	Solubility (%)[b]
228	29	5.02	alkaline	47
232	29.4	5.29		2
229	29.1	5.63		56
85	72.8	6.6		21
114	30.5	7.27	acidic	81
129	48.6	7.34		49
146	28.7	7.99		45
131	48.6	8.43		39
130	49	8.45		79
132	49.5	8.61		56
234	18	9.21		49

MW, protein molecular weight (in kDa).
[a] Target numbering refers to a nomenclature internal to VIZIER project (http://www.vizier-europe.org/).
[b] See Materials and Methods for details.

7, Tris pH 8, 2-(cyclohexylamino) ethane sulfonic acid (CHES) pH 9, Glycine–NaOH pH 10, 3-(cyclohexylamino)-1-propane sulfonic acid (CAPS) pH 11, ortho boric acid pH 12. (2) Folding enhancers (0 or 1 M [except trehalose: 0.5 M]): sorbitol, trehalose, glucose, sucrose, glycerol, glycine, trimethylamine *N*-oxide (TMAO), sarcosine, betaine, proline. (3) Aggregation suppressors (0 or 0.1 M [except: arginine: 1 M]): arginine, Tween 20, non detergent sulfobetaine (NDSB) 201, NDSB 256, Nonidet P40, Triton X100. (4) Reducing agent (0 or 10 mM): dithiothreitol (DTT). (5) Ionic strength (0 or 0.1 M): KCl. To fit with a microplate format, the 3696 combinations of the full factorial were reduced to 96 using JMP6 design of experiment software (SAS Statistical Discovery).

Proteome pI

We analyzed the pI from the complete sets of predicted protein sequences encoded by the human genome and the genome of *E. coli* K12 in Swissprot (release 53.1 of June 12th 2007, http://www.expasy.ch/sprot/) [15,16]. Accession numbers corresponding to these two complete sets were retrieved via the Sequence Retrieval System (SRS 5.1.0, http://www.expasy.ch/srs5/) and saved as lists in raw text format. The human list consisted of 18,768 accession numbers while the one from *E. coli* contained 4404. The two lists were uploaded on the pI/Mw online interface (http://www.expasy.ch/tools/pi_tool.html) [17,18] for computation of the predicted isoelectric point (pI) and molecular weight (Mw) of each predicted protein sequence. The human compartment of Swissprot contains a total of 20,925 predicted protein-coding genes; the difference with the 18,768 accession numbers is due to genes that are processed into multiple predicted proteins (polyproteins). Similarly, 4414 proteins were predicted from the 4404 *E. coli* accession number. It should be noted here that all proteins predicted as fragmentary in Swissprot were removed from the analysis to avoid statistical analysis on incomplete or not correctly predicted proteins. This resulted in 4403 and 20,565 analyzed proteins for *E. coli* and *H. sapiens*, respectively. Results were exported in numerical format and imported into MS Excel software for calculation of graphs.

Results

To assess the respective importance of the different chemicals used in our published refolding screen [1], the refolding buffers in which proteins remained soluble ("positive buffers") were clustered as a function of the different states of each variable of the screen. Note that all the raw data used for the trend analysis are from [1].

This procedure was mainly applied to the variable "pH" for two reasons: it was expected to have the most prominent effects on refolding, and this variable was made of the largest number of values (six pH values ranging from 4 to 9 were used whereas the other variables were used under a binary mode (present or absent)).

For each protein, the number of positive buffers was plotted against the pH scale. The results are reported in Fig. 1A. Except for one protein (Rv323c, pI 5.81), the average pH of positive buffers for proteins with pI below 6.36 was alkaline. Conversely, the average pH of positive buffers for proteins with pI above 6.79 was acidic, although SPINE23 (pI 8.68) seemed to refold in any buffer.

In order to better quantify this first observation, the mean value of refolding pH was calculated for each protein as described in Materials and Methods. The result of this calculation was reported as a function of protein pI in Fig. 1B columns 3 and 4. The results confirmed that regardless of their size (column 2), proteins with a pI below 6.36 and above 6.79 remained soluble after refolding in buffers the average pH of which was, respectively alkaline and acidic. The results confirmed that SPINE23 (pI 8.68) remained equally soluble in all buffers, hence the coincidence of its mean refolding pH (pH 6.6) with the mid value of the refolding screen (pH 6.5). Finally, the results also confirmed that Rv323c (pI 5.81) did not comply with the general rule but remained soluble only in the utmost acidic buffers of the screen (pH 4).

To get more insight in the underlying mechanism, we next calculated for each protein the difference between a protein pI and the mean value of refolding pH defined above. In all cases, there was at least ~1 and at most ~4 pH unit difference (Fig. 1B, column 5). Combined with the previous observation, this suggested that a protein with a pI above 6.79 did not need to be tested for refolding in a pH range exceeding 4 pH units in the acidic direction, and a protein with a pI below 6.36 did not need to be tested for refolding in a pH range exceeding 4 pH units in the alkaline direction. The absolute value of this difference was minimal (~1) at pI 6.31 and increased on both sides to roughly reach a plateau (~3 to 4) at pI ~5 and ~8 in the acidic and alkaline directions, respectively. Again, SPINE23 (pI 8.68) behaved differently with a lower |pI − mean of refolding pH| (2 instead of ~3 to 4), a direct consequence of its capacity to remain soluble in any buffer.

Finally, the number of positive buffers was reported for each protein (Fig. 1B, column 6) and proved to be minimal for proteins with pI 6.36. By comparing the data in columns 5 and 6, we observed some degree of correlation between |pI − mean of refolding pH| (column 5) and the number of positive buffers (column 6). Whether this correlation supported the idea that |pI − mean of refolding pH| was the cause of the number of positive buffers can be proposed, but remains speculative at this stage. The correlation was more visible using a graphic layout (Fig. 1C). Fig. 1C also revealed that proteins with pI below 5 or above 8 remained soluble in a larger number of buffers than those with pI within these values, and particularly those with pI around 6.36. Conversely, the number of positive buffers increased dramatically for proteins with pI located at (pH 9) and beyond (pH 10) the alkaline limit of the screen. The lack of proteins with pI below 4.38 did not allow the symmetrical situation to be described at or below the acidic limit of the screen (pH 4), which could therefore be only speculated.

Three of the above observations could be summarized as follows. (i) The switch from alkaline to acidic refolding buffers was located between pI 6.36 and pI 6.79. (ii) The difference between a given pI and the mean refolding pH was minimal at pI 6.31. (iii) The lowest number of refolding buffers was found for a protein with pI 6.36. As indicated by gray boxes in Fig. 1B, columns 3–6, these three events were clustered between pI 6.31 and 6.79. Since our refolding screen spanned a 4–9 pH scale, the middle of this scale was pH 6.5 which superimposed well with this clustering. In Fig. 2A and B, we propose a schematic layout of this summary.

We have seen that the point where the proteins switched from acidic to alkaline buffers for refolding was not 7 but was located somewhere between pI 6.31 and pI 6.79 (Fig. 1B columns 3 and 4, and Fig. 2A). Since these values coincided with the midpoint of our refolding screen, we wondered whether the partition point could have been artificially created by the pH range chosen for the screen (pH 4–9). According to the scheme presented in Fig. 2B, the likelihood of protein solubility increased with the difference between the protein pI and the mean refolding pH with a maximal difference of 4 units. In other words, acidic proteins had no other "choice" but to refold in an alkaline buffer (and conversely for alkaline proteins), because refolding in the acidic direction was precluded by the lack of refolding buffers with pH below 4 (or above 9 for alkaline proteins). This "impossible choice" is represented as 2 gray arrows pointing to a question mark in Fig. 2A.

To test this hypothesis, a new refolding plate was made with a pH scale ranging from 1 to 12, and the solubility of an acidic protein with a pI close to the switch point (*Thermotoga maritima* αfTM, pI 6.18) was assayed. According to the refolding scheme of Fig. 2A, αfTM was expected to remain soluble in buffers with alkaline pH from 1 to 4 pH units higher than 6.18 and possibly beyond this value. And indeed, this predicted result was experimentally validated with both the original and the new plates (Fig. 3, upper panel). Interestingly, while αfTM solubility decreased in refolding buffers with pH 6 to 4 in both refolding plates, it started to increase again at pH 3 in the new refolding plate, and kept on increasing with pH drop to reach a maximum at the two lowest pH of the scale (pH 2 and 1). Incidentally, these data could explain the apparently odd behavior of Rv323c which remained soluble only in the most acidic buffers of the original refolding plate (pH 4, see Fig. 1A), and not in buffers with intermediate pH values. Thus, it seemed that the "choice" to refold in an alkaline buffer for an acidic protein was forced by the design of the refolding screen. It is noteworthy however that while 100% buffers with pH 8 to 12 promoted αfTM solubility, refolding buffers with pH 2 and 1 reached a lower maximum (85%).

To evaluate whether this observation would also apply to proteins with a more acidic pI, the same experiment was performed using *Aspergillus niger* feruloyl esterase A (FAEA) (pI 4.66) and *Mycobacterium tuberculosis* Rv1399c (pI 4.38). Again, the solubility of the two proteins followed in both the original and the new plates the predictions of the refolding scheme depicted in Fig. 2A and B, and increased when the buffer pH increased (Fig. 3, medium and lower panels). In the acidic direction the behavior of the proteins differed from each other. While FAEA exhibited a tiny solubility increase at pH 1, Rv1399c was found soluble at pH 2 and 1. As already observed with αfTM, the solubility maximum reached by Rv1399c in the acidic direction was lower (75%) than that reached in the alkaline direction (100%).

When using the original plate, the mean refolding pH of FAEA was 7.8. Because it had a larger pH range, the new plate shifted the mean refolding pH of FAEA from 7.8 to 8.6 (denoted by asterisks in Fig. 3, medium panel). Surprisingly, although the |pI − mean of refolding pH| of FAEA increased proportionally (from 3.2 in the original plate to almost 4 in the new plate), it remained within the limits determined using the original plate (Fig. 2B).

The results obtained with αfTM, FAEA and Rv1399c strengthened the hypothesis that acidic proteins refolded in alkaline buffers because of the design of our refolding screen rather than because of intrinsic properties of the refolded proteins.

Although we did not perform the experiment with alkaline proteins, we anticipate that similar results would have been obtained. However, we decided to investigate neither the refolding of acidic proteins in acidic buffers nor that of alkaline proteins in alkaline buffers any further because (i) the results obtained using αfTM and Rv1399c indicated that alkaline buffers were more efficient for refolding acidic proteins, and although we did not test it we anticipated that the same was true for alkaline proteins with regards to acidic buffers; (ii) our goal was to set up a practical screen, and we considered that refolding proteins at extreme pH was not a reasonable option as the observed solubility at extreme pH could be due to denatured soluble proteins.

In Fig. 2A, the base of triangles representing the range of refolding pH was smaller than the base of triangles representing the range of corresponding pI. This was true for acidic as well as alkaline proteins, and suggested that refolding a population of proteins spanning a wide range of pI values would not require that an as large refolding pH range be tested for obtaining acceptable refolding results. In light of the results reported in the first part of this study (Figs. 1 and 2), this observation prompted us to consider the possibility of shrinking the number of refolding buffers of the screening test from 96 to only 1 (one acidic buffer for refolding alkaline proteins, and one alkaline buffer for refolding acidic proteins). To test this hypothesis, a potentially "universal" refolding buffer was devised.

To that end, another trend analysis taking into account all the values of all the variables of the original refolding screen was performed using the same data [1]. The results are summarized in Fig. 4. Some chemicals which proved equally or less efficient when present than when absent were not considered for use in the new single buffer screen. This concerned NaCl, NDSB195, the "cocktail", EDTA, PEG 400 and 4000. Conversely, the presence of β-MSH, Arg, NDSB 201 and 256 and KCl improved protein solubility. The new refolding buffer core was therefore made of 100 mM KCl, 10 mM β-MSH, 200 mM Arg, 50 mM NDSB201, and 50 mM NDSB256. For making acidic or alkaline refolding buffer, 50 mM Na acetate pH 4 or Tris pH 8 were respectively added to the core buffer.

The capacity of the new buffer to promote protein solubility during refolding was evaluated with a set of 11 viral proteins expressed insoluble in *E. coli* under all tested conditions. Following the prediction of Fig. 2A, proteins were refolded in only one of the two buffers, and the choice between acidic and alkaline buffer was based on the protein pI only. The results are reported in Table 1. A protein solubility threshold of 20% after refolding was considered compatible with the amount of soluble protein required for subsequent crystallogenesis steps. According to this criterion, 90% (10 out of 11) of the tested proteins remained soluble after refolding in the acidic or in the alkaline buffer. This success rate validated the hypothesis that a single refolding buffer could be used provided it was correctly chosen with respects to the pI of the protein to refold. More precisely, 100% proteins (7 proteins) remained soluble in the acidic buffer, and two of them exhibited a refolding efficiency of ~80% (VIZIER114 and 130). By contrast, out of 4 proteins refolded in the alkaline buffer 1 did not refold (VIZIER232, 2% solubility) and 1 was barely above the threshold (VIZIER85, 21% solubility). Interestingly, VIZIER85 had a pI of 6.6, which we have seen corresponded to proteins associated with the lowest number of positive buffers when refolded in the original plate (Fig. 1B and C). In addition, pI 6.6 was also located at the switch-point between alkaline and acidic refolding (Fig. 1B, column 3 and 4). Whether such proteins would benefit from refolding trials in both acidic and alkaline buffers was not evaluated in the present study but could be a possible way of investigation. Although definitive conclusions could not be drawn from this limited set of proteins, it seemed that the acidic buffer had higher refolding capacities (100% of 7 proteins) than the alkaline buffer (⩽75% of 4 proteins). If this higher refolding efficiency of the acidic over the alkaline buffer was confirmed using larger sets of proteins, then it could be worth trying alkaline buffers with pH higher than 8, as suggested by Fig. 2B.

Although informative, the fact that a protein remains soluble in a given refolding buffer is not an indisputable evidence of folding. The latter can be evaluated by different means such as functional

and biophysical/biochemical methods. The latter comprise, for example, size exclusion chromatography which allows distinguishing among monomers and multimers and circular dichroism which allows asserting that secondary structures (α-helix, β-sheet) are present within the refolded protein. The former comprise, for example, biological tests such as binding assay or enzymatic activity measurement.

We already reported that both the solubility and enzymatic activity of refolded FAEA increased with pH and was maximal at pH 9 when a pH 4 to 9 screen was used [14]. The ability of the alkaline refolding buffer to provide soluble and folded proteins was checked here by refolding the other two acidic proteins used in Fig. 3 (Rv1399c and αfTM) in the alkaline buffer, and then assaying the enzymatic activity of the refolded proteins. The result, reported in Fig. S1, indicated that these refolded enzymes were active on their specific substrate. This assay was intentionally performed under identical conditions (i.e., not optimized for each enzyme in terms of pH, temperature, etc.) for the reasons discussed below.

Finally, we already reported the efficiency of the acidic buffer [19]. In that study, two out of three viral proteins were refolded in the acidic buffer (50 mM Na acetate pH 4, 100 mM KCl, 10 mM β-MSH, 200 mM Arg, 50 mM NDSB201, and 50 mM NDSB256), and their folding assessed by size exclusion chromatography and crystallization. Arg and the two NDSB were not used during the quantitative production in order to simplify the refolding process. Interestingly, no significant difference was observed in the refolding yield, suggesting that pH might be the most important variable in the refolding mechanism.

Discussion

Thanks to a trend analysis of our published data, we have devised what can be described as a "single buffer refolding screen". This formulation sounds self-contradictory if "screening" is understood as "buffer screening" but remains acceptable if it is understood as "protein screening" (i.e., screening for proteins that are prone to *in vitro* refolding) as it is the case of high throughput refolding screenings in structural genomics. However, although this new screen was primarily intended for screening the refolding of a large number of proteins, which is generally the case with post-genomic programs, it can also be used for testing the refolding of a single protein without the need for preparing a large number of refolding buffers.

A direct consequence of testing protein refolding in a single buffer is that, in contrast with 96-well refolding screens, smaller culture volumes produce enough protein for the test. Furthermore, the number of proteins to test can be increased. This can help solving a bottleneck in structural genomics programs in which proteins must be processed in parallel at high throughput, and opens new perspectives in post-genomics. For example, the aim of "Enzyme Genomics" is to discover enzymatic activities among proteins of unknown function at genome scale [20]. This approach is currently limited by the availability of soluble recombinant proteins expressed in *E. coli*. Because it requires small culture volumes and a single refolding buffer, the procedure described in the present study allows considering refolded proteins for Enzyme Genomics projects as well. In this regard, the crude results obtained with Rv1399c and αfTM (Fig. S1) indicate that enzymatic activities can be tested (1) with minute amounts of refolded protein, (2) directly after refolding without further purification, (3) using substrates (paranitrophenyl-derivatives) which can be processed in parallel allowing the results to be obtained in the same and possibly automated procedure. This possibility is strengthened by the description [21,22] and launch (Equilibrium Dialyzer-96™, Harvard Bioscience) of 96-well dialysis devices which could be used for fully automating the one buffer refolding screening in high throughput programs.

Unexpectedly, this screening approach relying on one acidic and one alkaline refolding buffer happens to be supported by whole genome sequencing data. Plotting the number of proteins of most proteomes as a function of their pI definitely reveals a bimodal distribution. For instance, when proteomes as distant as those of *H. sapiens* and of *E. coli* were analyzed, both revealed the same bimodal distribution (Fig. 2C and D), which could hence be considered a general feature of the living world.

In both cases, the number of proteins was close to 0 at pH 7.4 – presumably because 7.4 is the physiological pH of most living organisms – and increased to reach two maxima at 5.5–6 and 9–9.5 in the acidic and alkaline directions, respectively [23,24]. Practically, this means that having to refold a pro- or eukaryotic protein with a pI close to 7.4 is a low probability event, and this probability increases with the distance from this value to reach a maximum at pI 5.5–6 and 9–9.5.

This bimodal distribution was also found in the data reported in Figs. 1C and 2A and B, the pH difference between the two troughs (6.5 in our data (number of positive buffers and |pI − mean of refolding pH|, Figs. 1C and 2A and B) and 7.4 in proteomes (actual number of proteins, 2C and D) being most likely explained by the artifact introduced by the design of the screen and described in details in Fig. 3.

This property of proteomes has direct consequences on the general design of refolding screens. The trend analysis reported in Fig. 2A suggested that pro- and eukaryotic proteins with acidic pI (∼5.5 (*E. coli*) or ∼6 (*H. sapiens*), Fig. 2C and D) will refold in an alkaline buffer, and that proteins with alkaline pI (∼9.5 (*E. coli*) or ∼9 (*H. sapiens*), Fig. 2C and D) will refold in an acidic buffer. Now, according to Fig. 1C, proteins with "neutral" pI (around 7.4), although scarce, should be more tricky to refold. However, the results reported in Fig. 2A indicate that proteins in this problematic region will unambiguously be prone to refold in the acidic buffer. Thus, the midpoint mismatch between our refolding screen (6.5) and proteomes (7.4), although artefactually produced by the screen design, provides by chance a direct solution for refolding rarely encountered but difficult to refold neutral (pI 7.4) proteins. Therefore, in addition to providing an easier handling than a 96-wells screen, a refolding screen using at least an alkaline and an acidic buffer should also provide a suitable answer – at least regarding pH – for refolding most of the globular proteins whatever their pI.

Considering what we have learned from proteomes (Fig. 2C and D), it is not surprising to see that the pI distribution of recombinant proteins whose refolding has been deposited in the Refold Database (http://refold.med.monash.edu.au/analysis.php; in "Bar Graph" select "Isoelctric point") follows the same bimodal mode. Incidentally, this also indicates that refolded recombinant proteins are not a biased subset of natural proteomes.

By contrast, most of the refolding buffers reported in the Refold Database are alkaline (pH 8–8.5, (http://refold.med.monash.edu.au/analysis.php; in "Bar Graph" select "Refolding pH and Isoelectric point")). The data reported herein would suggest a completely different strategy taking into account the bimodal pI distribution, and making use of the alkaline buffer for acidic proteins only and of an acidic buffer for the alkaline proteins.

In conclusion, we would like to propose simple guidelines to best use this single buffer refolding screen. (1) Considering the relatively high number of different chemicals tested in our studies on protein refolding (see Table S1 of the present study and Table 1 of reference [1]), the core buffer used in our single buffer screen should be a convenient basis for a first trial, provided the refolding pH is chosen as described in the results section. (2) In a second step, it might be useful to redesign the core buffer with different

additives on a binary mode (absence or presence of the additive at a single concentration) so as to fit more closely with specific requirements of the protein of interest. For example, lipases might require the use of specific detergents, proteins from thermophilic organisms might require higher refolding temperatures, etc. (3) In a third step, the action of the same additives could be further tuned by using different concentrations of the best additive.

Conclusions

The trend analysis of experimental refolding data that we had previously obtained using a 96-wells refolding screen [1] led to the conclusion that refolding recombinant proteins could be easily performed using a single refolding buffer.

Acknowledgments

This work was funded by the European research program EVA (FP7 Capacities, GA 228292, http://www.european-virus-archive.com/), and by the French program "Vaincre La Mucoviscidose", http://www.vaincrelamuco.org/ewb_pages/p/projets-finances-detail.php?id=112506. We thank Drs Jacques Rohayem, Xavier De Lamballerie and Eric Snijder for providing the PCR products encoding proteins listed in Table 1, and Séverine Blanc for her excellent technical support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.pep.2012.01.014.

References

[1] R. Vincentelli, S. Canaan, V. Campanacci, C. Valencia, D. Maurin, F. Frassinetti, L. Scappucini-Calvo, Y. Bourne, C. Cambillau, C. Bignon, High-throughput automated refolding screening of inclusion bodies, Protein Sci. 13 (2004) 2782–2792.
[2] B. Coutard, B. Canard, The VIZIER project: overview; expectations; and achievements, Antiviral Res. 87 (2010) 85–94.
[3] K. Tsumoto, D. Ejima, I. Kumagai, T. Arakawa, Practical considerations in refolding proteins from inclusion bodies, Protein Expr. Purif. 28 (2003) 1–8.
[4] M. Li, Z.G. Su, J.C. Janson, In vitro protein refolding by chromatographic procedures, Protein Expr. Purif. 33 (2004) 1–10.
[5] G. Chen, E. Gouaux, Overexpression of a glutamate receptor (GluR2) ligand binding domain in *Escherichia coli*: application of a novel protein folding screen, Proc. Natl. Acad. Sci. USA 94 (1997) 13431–13436.
[6] N. Armstrong, A. de Lencastre, E. Gouaux, A new protein folding screen: application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase, Protein Sci. 8 (1999) 1475–1483.
[7] P. Sijwali, L. Brinen, P. Rosenthal, Systematic optimization of expression and refolding of the *Plasmodium falciparum* cysteine protease falcipain-2, Protein Expr. Purif. 22 (2001) 128–134.
[8] L. Trésaugues, B. Collinet, P. Minard, G. Henckes, R. Aufrere, K. Blondeau, D. Liger, C.Z. Zhou, J. Janin, H. Van Tilbeurgh, S. Quevillon-Cheruel, Refolding strategies from inclusion bodies in a structural genomics project, J. Struct. Funct. Genomics. 5 (2004) 195–204.
[9] C. Scheich, F. Niesen, R. Seckler, K. Bussow, An automated in vitro protein folding screen applied to a human dynactin subunit, Protein Sci. 13 (2004) 370–380.
[10] L. Lin, J. Seehra, M.L. Stahl, High-throughput identification of refolding conditions for LXRβ without a functional assay, Protein Expr. Purif. 47 (2006) 355–366.
[11] D. Ejima, K. Ono, K. Tsumoto, T. Arakawa, Y. Eto, A novel "reverse screening" to identify refolding additives for activin-A, Protein Expr. Purif. 47 (2006) 45–51.
[12] M.K.M. Chow, A.A. Amin, K.F. Fulton, T. Fernando, L. Kamau, C. Batty, M. Louca, S. Ho, J.C. Whisstock, S.P. Bottomley, A.M. Buckle, The REFOLD database: a tool for the optimization of protein expression and refolding, Nucleic Acids Res. 34 (2006) D207–D212.
[13] M.S. Willis, J.K. Hogan, P. Prabhakar, X. Liu, K. Tsai, Y. Wei, T. Fox, Investigation of protein refolding using a fractional factorial screen: a study of reagent effects and interactions, Protein Sci. 14 (2005) 1818–1826.
[14] I. Benoit, B. Coutard, R. Oubelaid, M. Asther, C. Bignon, Expression in *Escherichia coli*, refolding and crystallization of *Aspergillus niger* feruloyl esterase A using a serial factorial approach, Protein Expr. Purif. 55 (2007) 166–174.
[15] B. Boeckmann, A. Bairoch, R. Apweiler, M.C. Blatter, A. Estreicher, E. Gasteiger, M.J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, Nucleic Acids Res. 31 (2003) 365–370.
[16] E. Gasteiger, E. Jung, A. Bairoch, SWISS-PROT: connecting biomolecular knowledge via a protein database, Curr. Issues Mol. Biol. 3 (2001) 47–55.
[17] B. Bjellqvist, B. Basse, E. Olsen, J.E. Celis, Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions, Electrophoresis 15 (1994) 529–539.
[18] B. Bjellqvist, G.J. Hughes, C. Pasquali, N. Paquet, F. Ravier, J.C. Sanchez, S. Frutiger, D.F. Hochstrasser, The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences, Electrophoresis 14 (1993) 1023–1031.
[19] V. Lantez, K. Dalle, R. Charrel, C. Baronti, B. Canard, B. Coutard, Comparative production analysis of three Phlebovirus nucleoproteins under denaturing or non-denaturing conditions for crystallographic studies, PLOSntds 5 (2011) e936.
[20] E. Kuznetsova, M. Proudfoot, S.A. Sanders, J. Reinking, A. Savchenko, C.H. Arrowsmith, A.M. Edwards, A.F. Yakunin, Enzyme genomics: application of general enzymatic screens to discover new enzymes, FEMS Microbiol. Rev. 29 (2005) 263–279.
[21] I. Kariv, H. Cao, K.R. Oldenburg, Development of a high throughput equilibrium dialysis method, J. Pharmaceutical Sciences. 90 (2001) 580–587.
[22] M.J. Banker, T.H. Clark, J.A. Williams, Development and validation of a 96-well equilibrium dialysis apparatus for measuring plasma protein binding, J. Pharm. Sci. 92 (2003) 967–974.
[23] G.F. Weiller, G. Caraux, N. Sylvester, The modal distribution of protein isoelectric points reflects amino acid properties rather than sequence evolution, Proteomics 4 (2004) 943–949.
[24] S. Wu, P. Wan, J. Li, D. Li, Y. Zhu, F. He, Multi-modality of pI distribution in whole proteome, Proteomics 6 (2006) 449–455.