CHAPTER FOUR

Preparing a Purification Summary Table

Richard R. Burgess

Contents

Abstract

Once a protein purification scheme has been developed, the purification, characterization, and use/structure of a target protein are usually published. It is highly desirable to present the major steps in the purification and the corresponding features of the protein at each step summarized in the form of a purification summary table. In considering whether to repeat a published protein purification, a reader needs this information to evaluate the purification, and to decide if it is worth following or if it needs major modifications. In this chapter, I discuss the main characteristics of a useful purification summary table and point out common mistakes and problems I see in many such tables.

1. Introduction

As an executive editor and editor-in-chief of the journal, *Protein Expression and Purification*, I have reviewed on the order of 100 protein purification papers a year for over 18 years. It is remarkable how many manuscripts I receive where there is either no purification summary or one that is severely lacking in necessary information and accuracy. The essentials of a reasonable purification table are illustrated by the example below.

Suppose one set out to purify an enzyme from the bacterium *E. coli* starting with 10 g of wet weight cell pellet from a 4-l culture (10 g of wet weight cells typically would contain about 2 g of dry weight and about

McArdle Laboratory for Cancer Research, University of Wisconsin-Madison, Madison, Wisconsin, USA

Methods in Enzymology, Volume 463
ISSN 0076-6879, DOI: 10.1016/S0076-6879(09)63004-4

1200 mg of total protein). The cells are lysed by sonication to give a crude lysate and the debris is removed by centrifugation to give a crude extract. A 45–50% saturated ammonium sulfate cut was prepared. The 50% saturated ammonium sulfate pellet was dissolved in buffer and diluted to low salt and applied to a DEAE anion exchange column. The column was washed at low salt and then eluted with a linear salt gradient from 0.1 to 0.6 *M* NaCl, the peak of activity eluting at about 0.25 *M* NaCl. The peak was pooled and applied to a Sephacryl S-300 gel filtration column and eluted with a buffer at constant salt (isocratically). The fractions of peak activity were pooled and shown by SDS–PAGE with Coomassie blue staining to be a single band. The specific activity of the final material is the same as that of a known pure reference sample. The main fractions were all assayed for enzyme activity and protein determinations were carried out. The resulting data are given in Table 4.1.

A purification summary table should allow a reader to evaluate easily the procedure and readily detect particularly effective and ineffective purification steps. It should be easy to see if large losses occurred at a particular step.

A suitable table will contain the following columns:

1. *Major steps in the purification*. These typically include steps like:
   Crude lysate (the result of cell or tissue disruption). This step is often omitted since assays may be difficult but it is useful and even essential when much of the expressed recombinant target protein is in an insoluble inclusion body.
   Crude extract (the lysate after any insoluble material has been removed by centrifugation)

Table 4.1 A typical purification summary table

Step	Total protein (mg)[b]	Total activity (units)[c]	Specific activity (units/mg)	Yield (%)	Purity (%)
Crude lysate[a]	1200	120	0.10	100	0.8
Crude extract	1000	110	0.11	92	0.9
Ammonium sulfate 45–55% cut[d]	180	75	0.42	62	3.4
DEAE column (pooled peak)	24	60	2.5	50	20
Sephacryl column (pooled peak)	3.6	46	12.5	38	100

[a] From 10 g of wet weight *E. coli* cell pellet (from 4 l of bacterial culture).
[b] Protein concentration determined by Bradford assay using BSA as a standard protein.
[c] Enzyme activity measured as described in the methods section.
[d] Crude extract material that is soluble at 45% but precipitates at 55% saturated ammonium sulfate.

Ammonium sulfate cut
Pooled peak from an ion exchange column
Pooled peak from a gel filtration column
Pooled peak from an affinity column
Concentrated and dialyzed final product
Solubilized inclusion bodies. This step and the following two are often used in the case where an expressed recombinant protein is produced as insoluble inclusion bodies; see Chapter 17.
Washed inclusion bodies
Refolded, centrifuged, and concentrated material

2. *Amount of total protein (mg).* This is usually determined by a standard protein assay. Most commonly these days a Bradford dye-binding assay or a bicinchoninic acid (BCA) assay is used (see Chapter 8). It is important to indicate in the methods section what protein is used as the protein standard (typically BSA). Once the protein is purified it can also sometimes be quantified by measuring its absorbance at 280 nm and the use of an appropriate molar extinction coefficient.
3. *Amount of target protein or total activity (mg or units).* If there is a suitable enzyme assay for the target protein, it should be carried out on material from each major step. If the protein is not an enzyme or there is no quantitative assay, and if the protein is visible on a Coomassie blue stained SDS–PAGE, then often the stained gel is scanned and the amount of protein in the target protein band is determined. In other words, purity is determined and multiplied by the total protein to give an estimate of the total target protein.
4. *Specific activity (units/mg).* If enzyme activity assays are possible, then the total activity (units) is divided by the total protein (mg) to give specific activity as units/mg.
5. *Overall yield (%).* The yield at a step in the procedure is the amount of target (either total target protein or total activity) at that step divided by the amount of target in the first step (defined as 100%).
6. *Purity of target protein (%).* Purity is often determined by scanning a stained SDS–PAGE and measuring the amount of the stain associated with the target band as a fraction of the stain associated with all the bands on the gel. If one has a reliable assay, then if the final material is pure, its specific activity can be used to define purity. For example, if an earlier step has a specific activity 10% of the final pure material, then the purity at that step would be 10%.
7. *Relative or fold purification.* This is not essential since it can be calculated from the other values above, but it is often useful. This is merely setting the initial purity at a value of one and then giving the purity at each step relative to that of the first step. For example, in Table 4.1, the final step represents an overall fold purification of 125.

2. The Importance of Footnotes

Every protein and purification is different and footnotes are needed to help the reader understand what has been done. There should be a footnote that indicates the amount of raw material used in the preparation being summarized. For recombinant protein expressed in bacteria, for example, one should always give the number of grams of wet weight cell pellet used in the preparation. It is also useful to know the volume of the bacterial culture used, but that in itself is not enough since, depending on the growth media and conditions, the yield of wet weight cell pellet can range from 1 to 80 g/L. Another useful footnote indicates how the protein amount was determined (e.g., sometimes a Bradford assay is used on the early steps, but absorbance and extinction coefficient is used on the final product).

3. The Value of an SDS–Polyacrylamide Gel Analysis on Main Protein Fractions

I find that an SDS–polyacrylamide gel image is a very valuable complement to the purification summary table. If the same samples that represent the various steps in the purification table are also shown in a gel photo, then it is particularly easy to see the progress of the various fractionation steps toward production of a purified final target protein. The most useful gels are ones on which an equal proportion of the material at each step is loaded on the gel lanes.

4. Some Common Mistakes and Problems

1. *Use too many significant figures*. This is one of my pet peeves. I suspect that the concept of significant figures is no longer taught, because I find that a good 75% of the purification papers I review give protein amounts like 235.052 mg and yields like 46.72%. Just because a calculator or computer can divide two numbers and give one the result to eight figures does not mean that value is what one should enter into the table. Just remember that most protein quantification methods or enzyme assays are not accurate to better than 5–10%. When one writes 23.47 mg, one implies that it is not 23.46 or 23.48, but 23.47. In other words, one is implying that it is accurate to one part in over 2000 when it is not even

clear that it is accurate to one part in 20 (is it 22, 23, or 24?). No numbers should be given to more than three significant figures and in general most percentages can be given to two significant figures. Also, remember that any number resulting from the division of two numbers each accurate to $\pm 10\%$ will only be accurate to $\pm 20\%$.

2. *Calculate values erroneously*. Remarkably many tables contain simple arithmetic errors. All numbers should be checked and rechecked before submission of a manuscript.
3. *Use step yields instead of overall yields*. A step yield is the yield from a single step in the purification procedure, that is, the amount of target protein or activity after that step compared to that in the previous step of the procedure. A series of four fractionation steps might all give 60% step yields, but the overall yield is $(0.6)^4 = 0.6 \times 0.6 \times 0.6 \times 0.6 = 0.13$ or 13%. Overall yield is more useful. A procedure that gives an overall yield of a few percent may be due to a very lengthy and difficult purification of a rare or unstable protein, but more likely it indicates that the procedure has not been optimized very well.
4. *Calculate yield as yield of total protein*. Often I see a table in which the yield given is yield of total protein, rather than target protein. This is relatively useless information. Yield is always recovery of total target protein or activity.
5. *Write up and try to publish the first purification that gives any product*. There is a tendency for readers, especially inexperienced readers, to assume that a published purification is the result of many cycles of improvement and optimization, and represents the best way to purify the protein. This is very often not true. Many times, it is just a series of steps, often chosen arbitrarily that happens to result in some product. One should look very carefully at a purification to assess if it is a procedure that is worth trying to follow if one wants to purify some of the target protein. This is why a proper purification summary table is so valuable. If huge losses occur at a particular fractionation step, if the overall yield is very low, if the final purity is not high, or if similar fractionation steps are used several times, then perhaps the procedure should merely be used as a beginning point in designing a better, more effective purification.
6. *What to do when purifying a protein where a fusion partner is cleaved off during the procedure?* Very often a recombinant protein is expressed as a fusion with another protein or tag that aids in its folding or purification (see Chapter 16). Since most often the desired final product is the target protein without the fusion partner or tag, especially for structural studies, the fusion partner must be cleaved off by one of several specific proteases. Let us say that the target is 20 kDa and the fusion partner is 40 kDa, so the fusion protein expressed is 60 kDa. At the step where the fusion protein is cleaved, the yield of target protein seems to decrease by 67% even if all of it is recovered. How is this indicated in the summary table?

I suggest that the column on target protein amount contain two numbers; the mg of fusion protein and the calculated mg of the final target protein in parenthesis. That way at the step where the cleavage has occurred, one can continue giving just the mg of the cleaved product and the theoretical amount of cleaved product in the first step can be used to calculate the overall yield.