BioNMR in Drug Research. Edited by Oliver Zerbe

ISBNs: 3-527-30465-7 (Hardback); 3-527-60066-3 (Electronic)

1
Modern Methods for the Expression of Proteins in Isotopically Enriched Form

Heiko Patzelt, Natalie Goto, Hideo Iwai, Kenneth Lundstrom, and Erhard Fernholz

1.1
Introduction

The introduction of stable isotopes into proteins has significantly reduced the time requirements for structure elucidation of biomolecules. Moreover, structural studies of proteins with molecular weights exceeding the 10 kDa limit are usually not possible without uniform isotope labeling because of severe resonance overlap and inefficient coherence transfer along the rather small ^3J ^{1}H-^{1}H couplings. Nowadays, efficient expression of recombinant proteins is a prerequisite for many techniques used in structural biology, but the requirement for isotope labeling in particular often precludes NMR (nuclear magnetic resonance) structure determination of proteins isolated from natural sources. Specifically, proteins that have been uniformly labeled with ^{15}N and/or ^{13}C are commonly required for NMR spectroscopy, especially for backbone chemical shift assignment procedures, which are greatly facilitated by the use of a series of rather sensitive multi-dimensional triple-resonance NMR experiments (see Chapt. 4) [1], in a process that can also be automated with good success [2]. Moreover, the random replacement of nonexchangeable protons by deuterons reduces ^{1}H-^{1}H dipolar interactions and scalar couplings, thereby reducing peak line widths considerably and allowing structure elucidation of proteins exceeding 30 kDa [3, 4]. Random fractionally deuterated protein samples also permit the use of longer mixing times in NOESY (nuclear Overhauser effect spectroscopy) experiments, since spin-diffusion pathways are largely eliminated. In addition, transverse-relaxation optimized spectroscopy (TROSY [5], see also Chapt. 10), which has been used for molecules larger than 100 kDa [6], benefits dramatically from deuteration. This stems from the fact that the TROSY component that is narrowed by the DD-CSA compensation is broadened by dipolar interactions with nearby protons.

Besides uniform labeling approaches, stable isotopes can also be introduced at specific sites in proteins in order to simplify the assignment process and to isolate spectral information from the region of interest. For example, biosynthetically-directed fractional ^{13}C labeling offers the possibility of making stereospecific assignments of all isopropyl methyl groups of Val and Leu residues [7]. In another approach often used for solid-state NMR termed *residue-specific* labeling, isotope labels are introduced at single sites in a protein, as described in another chapter of this volume (Chapt. 11). A related scheme, called amino acid-type labeling, is accomplished by expression in an amino acid-based medium

where only targeted amino acids contain isotope labels [8]. Preparing a series of samples with different isotopically enriched single amino acid types provides a useful approach for the assignment of large systems not accessible by traditional methods (for an example see Ref. [9]). However, labeling specificity and yield can suffer from isotope scrambling arising from metabolic conversion of amino acids. This problem can be circumvented through the use of specially engineered amino acid-type specific auxotrophic strains. An interesting alternative that can be used to achieve residue and/or amino acid type-specific labeling is presented by the *in vitro* cell-free expression systems. These systems are additionally advantageous when expression products display cell toxicity.

Another labeling strategy geared toward the study of large proteins combines the favorable relaxation properties conferred by extensive deuteration with site-specific strategies for introducing protons. For example, methyl groups and/or aromatic amino acids can be targeted for protonation in otherwise fully deuterated proteins. An alternative approach for the study of large proteins features segmental labeling methods based on protein splicing methodology. Consequently, longer stretches of protein are isotopically enriched, leaving the remainder unlabeled. Isotope editing will remove all signals from unlabeled segments of the proteins, thereby largely reducing resonance overlap and facilitating assignment. Protein splicing methods can also be used to introduce non-natural amino acids or chemical modifications into the sequence. Therein, a protein segment containing an unnatural residue can be chemically synthesized and then ligated to a recombinantly produced (isotopically enriched) segment. Protein splicing can also be used to produce proteins which have high potential for cytotoxicity [10] or to stabilize proteins through cyclization.

This chapter will mainly focus on recent developments in protein labeling methodology. For an introduction to methods involved in the generation and yield optimization of protein samples labeled with ^{13}C and/or ^{15}N by recombinant methods in *E. coli* the interested reader is referred to excellent reviews published in the literature [8, 11–18]. Here we first describe the use of labeled algal hydrolyzates for the production of labeled proteins in *E. coli* or other organisms. We then review methods used for the introduction of isotope labels into specific sites, and this is followed by a section on segmental labeling approaches. Subsequently, we will summarize recombinant protein expression methods in hosts other than *E. coli* that have proven to be especially suitable for post-translationally modified proteins and membrane proteins, before concluding with an introduction to cell-free expression systems.

1.2 Isotope-Labeled Proteins from Hydrolyzates of the Green Alga *Scenedesmus obliquus*

Although alternative expression systems have been successfully adapted for the production of isotope-labeled proteins (see Sect. 1.5), heterologous expression in *E. coli* often remains the method of choice for NMR sample preparation. There is a fundamental difference, however, with respect to the kind of medium in which the cells are cultivated. In a so-called "chemically defined" or "minimal" medium only one or a very limited number of carbon sources is provided, e.g. glucose or glycerol. All bacterial metabolites have to be biosynthesized by the cells through the various, sometimes lengthy and energy-de-

manding, metabolic pathways. In a "complex" or "rich" medium, the cells grow, as the name suggests, on a complex mixture of amino acids and/or carbohydrates. Amino acid interconversions, and thus the potential for isotope scrambling in selectively labeled samples, are here reduced to a minimum. Unlabeled fermentations are usually performed in complex media (i.e. yeast extract-containing LB for *E. coli*), since protein yield and cell density are here considerably higher than in minimal media. The same is desirable for isotope-labeled fermentations, but the limited availability and/or high price of commercial amino acid/sugar mixtures in the required isotope composition often impose fermentations on a single carbon source.

Several companies are meanwhile supplying labeled amino acid/sugar mixtures of acceptable quality. However, especially if larger-scale or repeated preparations of labeled proteins are envisaged, investment of some extra time for the in-house production of labeled complex growth media for *E. coli* or other host cells clearly becomes advantageous, especially from the point of view of financial considerations.

The most frequently employed source for complex amino acid/sugar mixtures, labeled in any combination of ^{2}H,^{13}C and/or ^{15}N in *E. coli* continues to be a phototrophic green alga. *Scenedesmus obliquus* was introduced for that purpose in 1972 by Crespi and Katz [19]. In recent years, the original protocols have been modified by several groups, leading to improvements in the yield and purity of the algal amino acid mixtures, thereby enhancing protein labeling efficiency and expression levels in the hosts [20–22].

Depending on the desired labeling pattern, $^{2}H_2O$, $^{13}CO_2$ and/or $^{15}NH_4Cl$ or $Na^{15}NO_3$ are used as exclusive isotope sources during the algal fermentation. All of the above have become commercially available at affordable rates. For the preparation of random partially isotope-labeled amino acid/sugar mixtures, unlabeled water, carbon dioxide or nitrogen salts are simply admixed to the labeled starting material in the appropriate proportions.

Thus, the three basic steps required for the preparation of a uniformly labeled protein for NMR experiments are

1. production of isotope-labeled algal hydrolyzates,
2. adaptation of the protein overproducing organism (usually, but not always, *E. coli*) to growth on the algal medium, and
3. preparation and purification of the isotope-labeled protein on a preparative scale.

If specific amino acid-type labeling is required, the labeled amino acid is added to the fermentation of the expression host (topic 1 above, see Sect. 1.2.3). In this case, a thorough isotope analysis of the expressed protein is advisable prior to NMR spectroscopic investigations. This is preferentially achieved by GC-MS analyses of the hydrolyzed amino acids from the protein product.

1.2.1
Production of Isotope-Labeled Algal Hydrolyzates

Cultures of *S. obliquus* can easily be grown photoautotrophically in two-tier flasks in an inorganic medium. A stepwise replacement of H_2O by $^{2}H_2O$ leads to deuterated cultures, and a replacement of CO_2 (the sole carbon source) by $^{13}CO_2$ and/or the replacement of

the nitrogen-containing salts by their ^{15}N-isotopomers produces ^{13}C- and/or ^{15}N-labeled cultures. These are used as inoculi for larger fermentations in Fernbach flasks or stirred or airlifting fermenters. All fermentations are continuously illuminated with standard plant light bulbs or fluorescent tubes. In the last stage of $^{2}H_2O$ fermentations, all salts containing crystal water are repeatedly dissolved in small amounts of $^{2}H_2O$ and lyophilized before addition to the medium. If the algal cells are harvested under sterile conditions, the recovered medium can be re-inoculated for up to four further fermentations, and only phosphate is supplemented when the growth rate declines [22]. The recycling of the media cuts isotope costs by about 90%.

The algal cell mass is then purified from low-molecular-weight metabolites and hydrolyzed in HCl (^{2}HCl in the case of a deuterated fermentation). After neutralization and lyophilization, a white powder (typically 2–2.5 g L^{-1} of medium and fermentation cycle) is obtained, containing around 50% amino acids (for composition see Ref. [23]), 30% sugars (composition in Ref. [24]) and 20% NaCl. This amino acid/sugar mixture for complex microbial growth media can be produced with any combination of ^{2}H, ^{13}C and/or ^{15}N, including random fractional label distributions. Used as carbon source it enables the simple and quick preparation of isotope-labeled, complex microbial growth media for the production of labeled proteins.

1.2.2 Adaptation of the Protein Overproducer to the Algal Medium

The described algal hydrolyzate contains amino acids and sugars in a physiological composition, i.e. in a relative composition similar to that required by most host cells. Amino acid biosynthesis, interconversion, and thus the potential for isotope scrambling, are minimized. When all potentially inhibitory low-molecular-weight compounds are removed by extraction prior to hydrolysis, most organisms grow well in media containing their typical salt and trace element composition, with the exception of NaCl, which is introduced as part of the algal hydrolyzate. Only the carbon sources are substituted by the algal amino acid/sugar mixture (e.g. yeast extract is replaced by algal hydrolyzate with the required isotope composition). Examples of the production of isotope-labeled proteins in Bacteria [25], Archaea [23, 26] and Eucarya [27] can be found in the recent literature.

Since changing the carbon source may influence bacterial growth and expression characteristics, a series of unlabeled test experiments is recommended in order to establish the minimum hydrolyzate concentration required, as well as the reproducibility of protein expression levels.

While expression in ^{13}C- and/or ^{15}N-labeled media is usually straightforward, most organisms need to be adapted in 3 to 4 steps to growth in $^{2}H_2O$ (e.g. 50, 75, 90, 100%). In addition, while formulating any deuterated medium, it is important to recall that the reading on pH meters equipped with normal glass electrodes is about 0.4 units lower in $^{2}H_2O$ than in H_2O with the same hydrogen/deuterium ion concentration [28]. Because of the different physical properties of $^{2}H_2O$, growth may be slower than usual, and the timing for induction of protein expression may require adjustment. The extent of deuteration depends on the type of experiments that will be performed. For investigations of internal dynamics using ^{15}N relaxation or for backbone assignment with triple-resonance

spectroscopy, 100% deuteration at the nonexchangable carbon sites maximizes signal sensitivity and resolution. On the other hand, structural information traditionally relies on distance restraints derived from $^{1}H,^{1}H$ NOEs. Nietlispach *et al.* have calculated and measured the effects of various levels of random fractional deuteration and found 50–70% deuteration most useful for larger proteins [29].

1.2.3
Preparation of Homogenously Isotope-Labeled Protein by Fermentation on Algal Media

Production of isotope-labeled proteins on a larger scale from the optimized test conditions is typically routine as long as the physical growth parameters (reactor type, aeration, etc.) are not changed significantly. However, the purification of deuterated proteins may require some adjustments, depending on the techniques utilized. For example, because of their considerably higher density, centrifugation gradients must be adapted. Also, the chromatographic properties of deuterated proteins may display differences relative to their unlabeled (^{1}H at natural abundance) counterparts, reflecting potential shifts in isoelectric point, stronger intramolecular hydrogen bonds and weaker van der Waals interactions. A final consideration is the re-introduction of protons at the exchangeable amide sites. Since the quantitative exchange of amide protons from the protein core can be extremely slow, it is sometimes necessary to expose the sample to denaturing conditions, followed by refolding in H_2O, if possible.

1.2.4
Amino Acid-Type Specific Labeling

The principal difficulty associated with the preparation of amino acid-type specific labeled proteins is the suppression of metabolic scrambling of the label into other amino acid types through the common metabolic pathways in the host cell. The use of a complex amino acid/sugar mixture, such as the one present in the algal hydrolyzates, reduces this danger greatly compared to fermentations on a single carbon source. In fact, the activity of many enzymes responsible for amino acid interconversions appears to be low or absent in bacteria grown under these conditions. For example, for the production of a fully deuterium-labeled protein containing ^{1}H-Trp, the host cell is grown on fully ^{2}H-labeled algal hydrolyzate in $^{2}H_2O$ to which unlabeled Trp is admixed. The individual amount of the differentially labeled amino acid to be added to the fermentation may vary for the different residues and depends on the biosynthetic origin of the amino acid [30] as well as on its background concentration in the algal hydrolyzate [22]. Labels in the biosynthetically central amino acids Asx and Glx show a pronounced tendency for biosynthetically-directed isotope relocation, since these molecules may be used as metabolic precursors for a number of different downstream amino acids (e.g. Met, Lys, Thr, and Ile or Pro and Arg, for Asx and Glx, respectively). The metabolically peripheral amino acids can usually be labeled specifically with much higher isotope purity.

For every new amino acid, a small series of test experiments is normally sufficient to establish a compromise between the minimum concentration required for high specific labeling (usually at least ten times the amount introduced with the algal hydrolyzate) and the

toxicity limit for the respective amino acid. Before a large-scale fermentation is attempted, the isotope composition downstream from the labeled amino acid should be analyzed. The most probable sites of undesired isotope incorporation are found in the same biosynthetic group [30]. Labeled Tyr, for instance, may be found after media supplementation with labeled Phe, whereas Cys and Gly labeling may result from the addition of labeled Ser.

1.2.5
Mass Spectrometric Analysis of the Labeled Amino Acids

For amino acid analysis the labeled protein needs to be hydrolyzed and derivatized. Most commonly the hydrolysis is performed in 6 M HCl, and the amino acids are converted into their isopropyl ester and pentafluoropropanamide derivatives (Fig. 1.1) before GC/MS analysis. The molecular ion is not always visible after standard electron impact (EI) ionization, and the fragment after loss of the carboisopropoxy group is the highest observable peak. This leaves m/e=175 plus the mass of the amino acid side chain, from which the degree of labeling can be directly deduced.

Fig. 1.2 shows illustrative mass spectra for derivatized Phe with various isotope patterns. In unlabeled Phe (Fig. 1.2a), the highest observable peak at m/e=266 accounts for the aforementioned fragment of 175 plus the mass of the benzyl group (C_7H_7=91). The strongest peak is due to the tropylium cation (C_7H_7) at m/e=91. Complete deuteration (Fig. 1.2d) takes molecular ion to m/e=274 (175+$C_7{}^2H_7$), and the tropylium signal ($C_7{}^2H_7$) accordingly to m/e 98. As expected, the spectrum of $^{13}C,^{15}N$-labeled Phe (Fig. 1.2e) shows the corresponding signals at m/e=275 (175+2+$^{13}C_7H_7$) and 98 ($^{13}C_7H_7$).

The signals in the spectra of the partially deuterated amino acid (Figs. 1.2b–d and 1.2f) show a statistical distribution of the masses around the calculated values for all fragments and are thus an unambiguous proof of a complete random distribution of the labels.

1.3
Selective Labeling Schemes

While general labeling strategies relying on expression of proteins using hydrolyzates of algae are useful for uniform or amino acid-specific labeling, expression can also be performed in minimal media in a cost-effective manner. These types of expression media

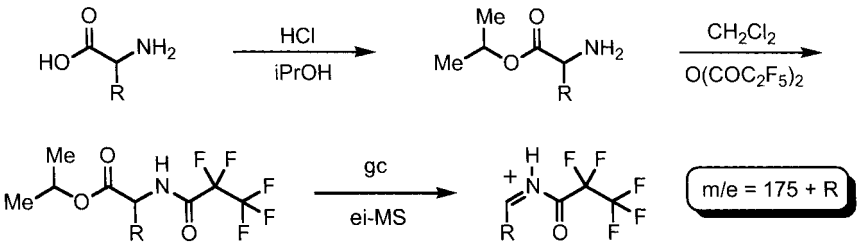


Fig. 1.1 Most common derivatization of amino acids for GC-MS analysis. The fragment without the carboisopropoxy group normally produces the highest observable peak, which is used for the determination of the isotope composition.

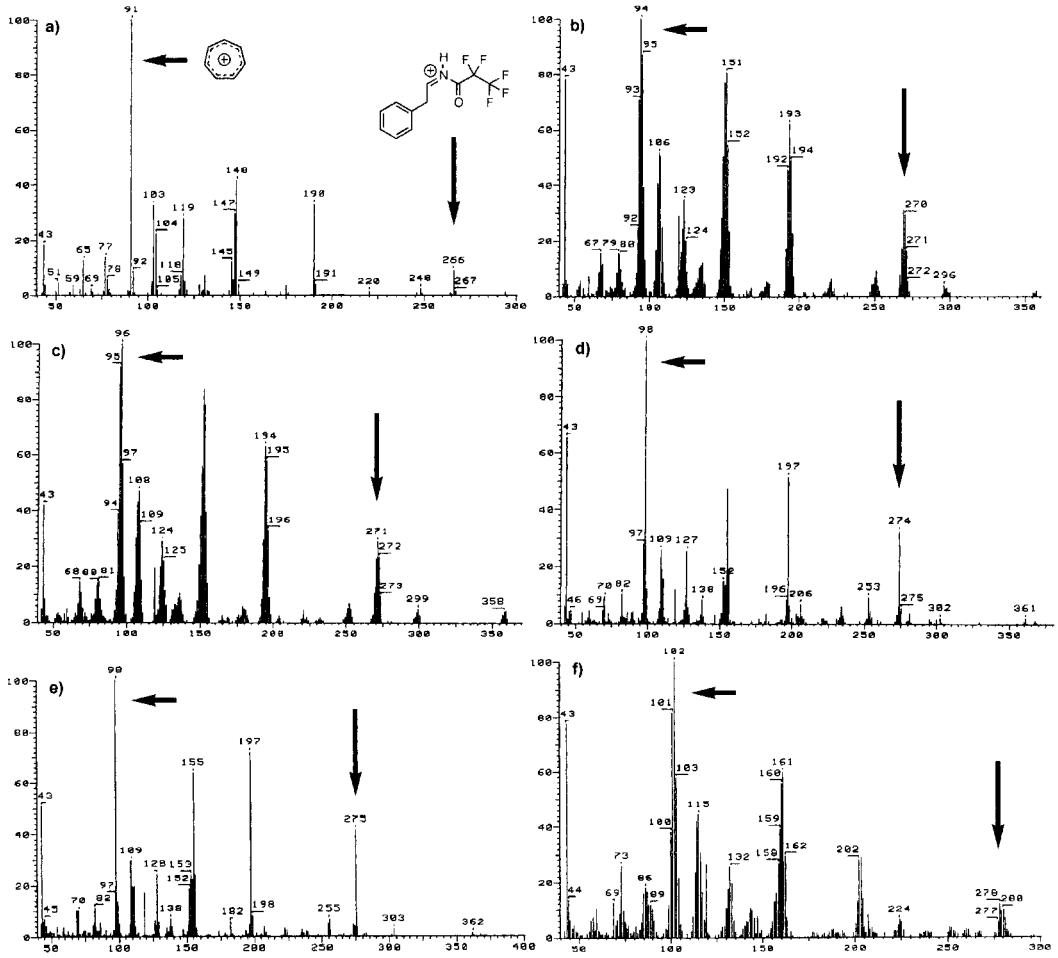
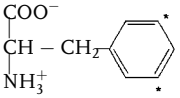


Fig. 1.2 EI mass spectra of derivatized phenylalanine, isolated from *S. obliquus*; **a** unlabeled, **b** random 50% ^{2}H-, **c** random 70% ^{2}H-, **d** 100% ^{2}H-, **e** ^{13}C,^{15}N-, and **f** ^{13}C,^{15}N/random 75% ^{2}H-labeled.

typically use glucose and ammonium salts as the sole carbon and nitrogen sources, respectively, and include D_2O as the medium base when deuteration is required. These growth conditions are easily adapted for uniform or random fractional isotope labeling. In addition, modifications to this general formula can be made to introduce isotope labels into specific sites, for example, by supplemention of the expression media with specifically labeled amino acids or amino acid precursors. In the following section we will review a number of methods used for different patterns of selective labeling in *E. coli*. Reagents used in these different strategies, with corresponding literature references, are summarized in Tab. 1.1.

Tab. 1.1 Chemical structures of metabolites involved in selective isotope labeling strategies (from Ref. [14] with kind permission)

Labeling agent	*Chemical structure*	*Incorporated as*	*Reference*
[3-^{2}H] α-keto-*iso*valerate	$^{*}CH_3 - {}^{*}CD(^{*}CH_3) - CO - {}^{*}COO^-$	(^{1}H-δ methyl)-Leu (^{1}H-γ methyl)-Val	43
[3,3-$^{2}H_2$] α-keto-butyrate	$^{*}CH_3 - {}^{*}CD_2 - {}^{*}CO - {}^{*}COO^-$	(^{1}H-δ1 methyl)-Ile	42
[ε-^{13}C]-L-phenylalanine		[^{13}C] Phe	52
[2-^{13}C] or [1,3-$^{13}C_2$]-glycerol	$CH_2(OH) - CH(OH) - CH_2(OH)$ or $^{*}CH_2(OH) - CH(OH) - {}^{*}CH_2(OH)$	^{12}C-^{13}C-^{12}C pattern	162
^{13}C pyruvate	$^{*}CH_3 - {}^{*}CO - {}^{*}COO^{-\P}$	(^{1}H-δ methyl)-Leu (^{1}H-γ methyl)-Val (^{1}H-γ2 methyl)-Ile (^{1}H-β methyl)-Ala	35
[3-^{13}C] pyruvate	$^{*}CH_3 - {}^{*}CO - {}^{*}COO^{-\Pi}$	(^{1}H-δ methyl)-Leu (^{1}H-γ methyl)-Val (^{1}H-γ2 methyl)-Ile (^{1}H-β methyl)-Ala	39, 163

* Indicates positions labeled by ^{13}C. Growth conditions optimized to maximize yields of CH_3 isotopomers in target methyl groups. Π Protocol for expression leads to production of all possible methyl isotopomers, with desired species CHD_2 for ^{13}C relaxation measurements.

1.3.1 Reverse-Labeling Schemes

1.3.1.1 Selective Protonation of Methyl Groups in ^{2}H-Labeled Proteins

As described in the introduction, deuteration is routinely used to reduce the rapid transverse relaxation rates characteristic of larger proteins, leading to improvements in peak line widths and experimental sensitivity. Deuteration of all the carbon-bound protons maximizes the sensitivity gains that can be obtained from this labeling strategy, and has thus proven useful for in the assignment of backbone ^{1}H, ^{15}N, ^{13}C and side-chain ^{13}C chemical shifts of large proteins. However, the elimination of all but the exchangeable protons significantly impedes structural studies that rely on conventional NOE approaches. Although in some cases it is possible to use only backbone ^{1}HN-^{1}HN NOEs to obtain a protein global fold, the accuracy of these structures is very low because of the small proportion of distance restraints between protons from nonsequential residues. These global folds tend to be less compact than high-resolution structures, with the backbone pairwise root-mean-squared deviation (rmsd) to the high-resolution structure ranging from 5 to 8 Å [31, 32].

For the purpose of increasing the number of protons in the protein core while maintaining the benefits of extensive deuteration, it is possible to re-introduce protons using a "reverse isotope" labeling approach. In some of the original approaches, side chains of target amino acid types were selectively protonated in deuterated proteins by adding protonated forms of these amino acids to the D_2O growth medium (see, for example, Refs. [33, 34]). In a variation of this theme, Rosen and coworkers developed a protocol to selectively incorporate protons at the methyl positions of Ala, Val, Leu and Ile $\gamma 2$ [35]. Methyl groups are enriched in protein hydrophobic cores and hence are attractive targets for selective protonation [36]. In addition, NMR spectroscopic properties of methyl groups are favorable owing to reasonably well-resolved ^{13}C-^{1}H correlations and rapid rotation about the methyl symmetry axis that reduces peak line widths [37].

The original protocol for the production of ^{15}N, ^{13}C, ^{2}H, ^{1}Me (protonated methyl)-proteins utilized ^{13}C,^{1}H-pyruvate as the exclusive source of carbon in 100% D_2O minimal media [35]. Pyruvate can be diverted into the tricarboxylic acid (TCA) cycle to produce many of the intermediates used in the biosynthesis of amino acids [38], but can also be directly incorporated into amino acids either by transamination (Ala), or reactions with threonine (Ile), pyruvate (Val) or both pyruvate and acetyl CoA (Leu). As was observed, pyruvate that is directly incorporated into these amino acids will largely retain the methyl protons, while those amino acids synthesized indirectly via intermediates will be highly deuterated. However, the incorporation of protons at each methyl site tends to be variable, with the result that methyl isotopomers such as $^{13}CHD_2$ and $^{13}CH_2D$ are also produced [35]. The additive deuterium isotope effect on both carbon and proton chemical shifts produces upfield shifts relative to the $^{13}CH_3$ peak by 0.02 and 0.3 ppm per ^{2}H atom in the ^{1}H and ^{13}C dimensions respectively. As a result, ^{13}C-^{1}H correlation spectroscopy of the methyl region shows three peaks for every methyl group labeled in this way, translating into resolution and sensitivity problems. Nonetheless, through the use of ^{2}H-purging pulse schemes during the acquisition of carbon chemical shift, it is possible to remove the peaks arising from methyl groups containing ^{2}H [32]. In addition, it should be noted that other pyruvate-based labeling schemes have also found great utility in the measurement of side-chain dynamics involving these methyl-containing side chains [39–41].

More recently, the yield and uniformity of methyl group protonation was enhanced through the use of α-ketoisovalerate in combination with α-ketobutyrate to produce ^{15}N, ^{13}C, ^{2}H-labeled proteins with protons introduced at the methyl positions of Leu, Val and Ile ($\delta 1$) [42, 43]. Proteins expressed in D_2O/^{13}C, ^{2}H-glucose/$^{15}NH_4Cl$ minimal media can be supplemented with ^{13}C, [3, ^{2}H] α-ketoisovalerate for the selective protonation of the Val and Leu methyl groups and [3,3-$^{2}H_2$], ^{13}C α-ketobutyrate for Ile $\delta 1$ methyl group labeling. Using this strategy, labeling efficiency of the target methyl groups was shown to exceed 90%, while high levels of deuteration were maintained at other sites without production of methyl group isotopomers containing deuterium. Since the selectively deuterated form of these amino acid precursors can be obtained by base-catalyzed proton exchange in aqueous buffer, the protonated, commercially available forms of these precursors are straightforwardly adapted to this labeling scheme.

1.3.1.2 Structure Determination of Selectively Methyl Protonated Proteins

High levels of deuteration combined with selective methyl protonation using one of the schemes outlined above permits the measurement of $^{1}H^{N}$-$^{1}H^{N}$, $^{1}H^{N}$-methyl and methyl-methyl NOEs. Global folds can be determined using this subset of NOEs, where the quality of these structures is a reflection of protein topology, secondary structure content, and the location and distribution of methyl groups in the protein [32]. In the case of a 30 kDa cell adhesion fragment from intimin, for example, intradomain backbone rmsd values of the ensemble of structures ranged between 1.5 and 1.8 Å from the mean [44]. In contrast, MBP structure quality was lower, with intradomain backbone rmsds between NMR and crystal structures of 3.1–3.8 Å [45]. Although structures produced by this methodology are often of a preliminary quality, they can nonetheless be useful in the localization of ligand or protein interaction sites and the identification of homologous proteins (e. g. Ref. [46]). Global folds can also be used as a structural stepping stone in the generation of high-resolution structures, since the assignment of additional NOEs from random fractionally deuterated samples or fully protonated molecules is facilitated by the use of a preliminary structure [47, 48]. Further improvements in the quality of structures can also be obtained through the incorporation of additional restraints such as dipolar couplings ([45]) or homology modeling (e. g. Ref. [49]).

1.3.1.3 Introducing $^{1}H,^{12}C$ Aromatic Residues into Otherwise ^{13}C Uniformly Labeled Proteins

Alternative schemes involving selective protonation have also been developed to increase the number of side-chain distance restraints that can be obtained in highly deuterated proteins. For example, ^{1}H, ^{12}C Phe and Tyr can be directly incorporated into an otherwise uniformly ^{13}C-labeled protein expressed in minimal media [50]. Since these amino acids are also preferentially located in the hydrophobic cores of proteins, as well as at ligand binding interfaces, distance restraints involving these residues can be very valuable. This labeling strategy was shown to be useful for proteins under 30 kDa with relatively few Phe and Tyr such as a 24 kDa Dbl homology domain [48] and the 25 kDa antiapoptotic protein Bcl-xL [51]. In cases where overlap in the aromatic spectrum becomes problematic, a synthetic strategy has been introduced to produce Phe that is ^{13}C labeled only at the epsilon position [52]. An illustration of the utlility of this approach is provided by the structure determination of a 21 kDa Dbl homology domain containing seven phenylalanine residues [47].

1.3.1.4 Backbone-Labeled Proteins

Protocols for selective isotope labeling of protein backbone atoms are also being developed, since the prevention of ^{13}C incorporation at the C^{β} site circumvents resolution problems associated with homonuclear $^{1}J_{C\alpha C\beta}$ coupling. Toward this end, syntheses of backbone-labeled amino acids have been described for ten different amino acids starting from ^{15}N, $^{13}C_2$-glycine [53, 54]. While original demonstrations of backbone labeling utilized a CHO cell expression system to prevent isotope scrambling [53], bacterial cell expression systems have also proven amenable to this strategy [54]. In this case, the expression medium must contain the full complement of amino acids, which are then replaced with those that are $^{13}C\alpha$, ^{13}CO, ^{15}N, and $^{1}H\alpha$ (or 50% $^{2}H\alpha$) labeled just prior to induction of protein expression. As was demonstrated for ubiquitin backbone-labeled with a subset of amino acids, sensitivity and

resolution in HNCA-type experiments is enhanced, and couplings can be readily measured from IPAP ^{1}H-^{13}C HSQC (heteronuclear single-quantum correlation) spectra [54, 55].

1.3.2 Selective ^{13}C Methyl Group Labeling

To reduce the expense of producing selectively methyl-labeled proteins, it is possible to use ^{13}C-methyl iodide to synthesize α-ketobutyrate and α-ketoisovalerate containing ^{13}C only at the methyl sites [56]. A larger than 20-fold reduction in the cost of precursor molecules can be achieved using this synthetic strategy in place of the commercially available uniformly ^{13}C-labeled isotopomers. Although these compounds can be adapted to the selective methyl protonation scheme described above, they can also be used to produce proteins that only contain ^{13}C at the methyl positions of Val, Leu, and Ile $\delta 1$, with ^{12}C at all other sites. The reduced cost, spectral simplification and sensitivity enhancement of ^{13}C methyl-labeled proteins over uniformly ^{13}C-labeled samples facilitates the use of chemical shift mapping in the search for potential lead compounds in the drug discovery process. If, on the other hand, selective methyl protonation is required, the ^{1}H, ^{13}C methyl-labeled α-ketoisovalerate and α-ketobutyrate would be added to D_2O expression media containing ^{2}H, ^{13}C-labeled glucose and ^{15}N-labeled ammonium salt. However, compared to the uniformly ^{13}C-labeled selectively protonated samples described previously, structure determination is less straightforward since backbone carbon atoms for isoleucine and valine are derived from the nonmethyl portion of the α-ketobutryate and α-ketoisovalerate, respectively, and would therefore contain the ^{12}C isotope. Nonetheless, once assignments are made, NOE measurements involving these methyl groups benefit from elimination of one-bond ^{13}C-^{13}C couplings leading to narrow carbon line widths without the requirement for constant-time evolution periods [51].

1.4 Intein-Based Protein Engineering for NMR Spectroscopy

Recently, new advances in biochemistry have opened up a novel approach for protein engineering, which utilizes a protein-splicing domain. Protein splicing is a post-translational chemical modification discovered in nature, which catalyzes the excision of an intervening polypeptide (*inte*rnal prote*in*, intein) while simultaneously ligating both the N- and C-terminal flanking polypeptide chains (Fig. 1.3, 1.5 A), analogous to RNA splicing [57, 58]. This unique enzymatic process has been used in various biochemical and biotechnological applications such as protein purification, protein ligation, backbone cyclization and C-terminal modifications. In particular, protein ligation using inteins has opened up a new way for the production of segmentally labeled proteins, thereby reducing the complexity of NMR spectra. For larger proteins or multimeric proteins, it may be necessary to combine selective labeling and segmental isotope-labeling approaches. Moreover, it will be useful in cases where information is desired only for a small part of a large protein, e. g. to characterize interactions with known binding sites or to detect conformational changes in a specific region. Another potentially useful application of inteins for NMR is backbone cyclization to enhance the stability of proteins.

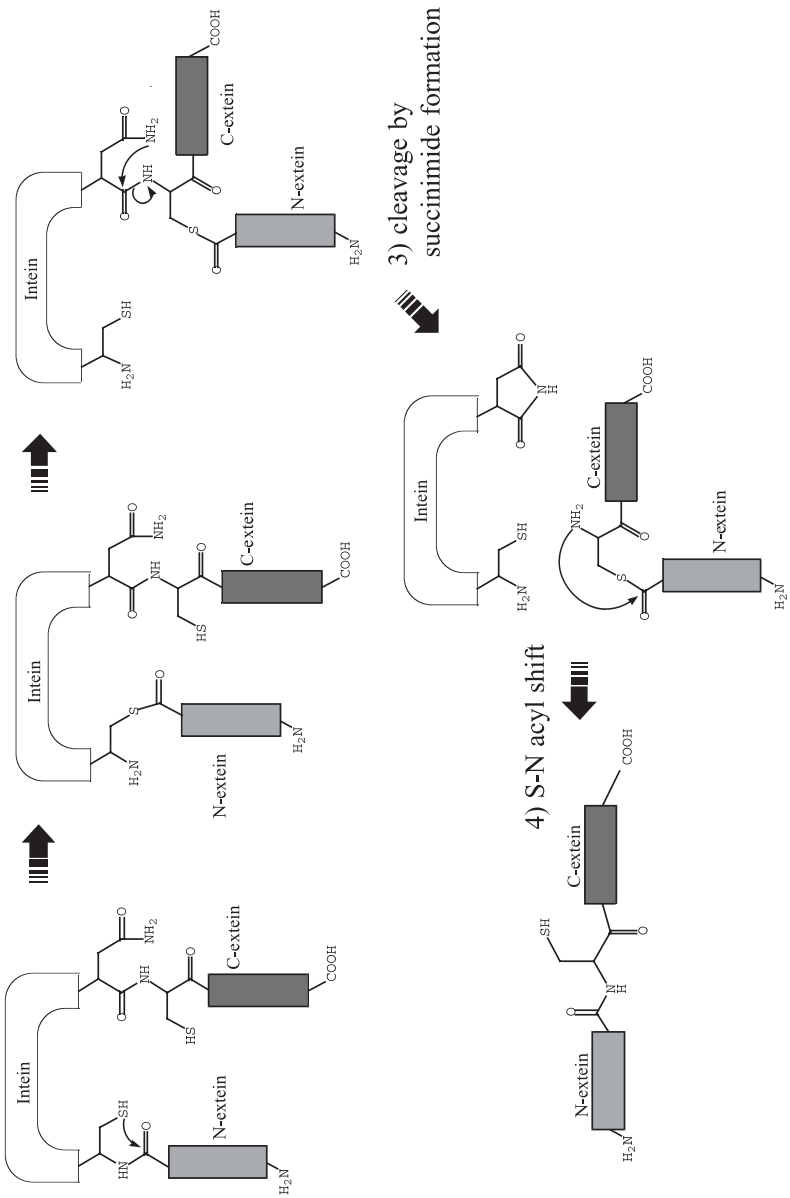


Fig. 1.3 The currently accepted chemical mechanism of protein splicing. **1** N-S(O) acyl shift, **2** transesterification, **3** cleavage by succinimide formation, **4** S(O)-N acyl shift.

1.4.1 Segmental Labeling of Proteins

Although more than 100 protein splicing domains have been found in nature [59], only a handful have been used for segmental labeling purposes, namely *Sce*VMA (PI-*Sce*I), PI-*Pfu*I and PI-*Puf*II. The currently accepted reaction mechanism for protein splicing consists of the following four steps, namely, (i) N-S(O) acyl shift, (ii) transesterification, (iii) succinimide formation, and (iv) S(O)-N acyl rearrangement (Fig. 1.3). Either a subset of these four steps or the entire reaction process can be used for protein ligation. For example, the intein-mediated protein ligation (IPL) approach utilizes a subset of these reactions by using a modified intein as described in Sect. 1.4.1.1. On the other hand, the split intein approach requires all four reaction steps. In this case the success of the reaction depends on refolding properties of the split intein (Sect. 1.4.1.2).

1.4.1.1 Intein-Mediated Protein Ligation (IPL)/Expressed Protein Ligation (EPL) using the IMPACT System

The IMPACT (Intein-Mediated Purification with an Affinity Chitin-binding Tag) system was originally developed as a novel purification method by New England Biolab. It makes use of a modified intein, *Sce*VMA, in which the active site was mutated from His-Asn-Cys to His-Ala-Cys, so that the usual cleavage due to succinimide formation involving the side-chain of Asn is no longer possible (Figs. 1.4 A and 1.6 A). Since a chitin-binding domain (CBD) is fused to the C-terminus of the intein, this protein can be immobilized to a chitin column, providing a convenient tool for purification [60]. The desired protein segment is fused at the N-terminus of the intein, which can be liberated from the intein-CBD portion of the fusion protein by addition of nucleophiles such as dithiolthreitol (DTT), ethanethiol, 2-mercaptoethane sulfonic acids (MESNA), hydroxylamine or cysteine. The IMPACT system provides a good opportunity to expand the application of native chemical ligation, which was originally developed by the Kent group, to a variety of protein targets, because the C-terminus of the N-terminal fusion polypeptide can be converted easily into a thioester group by the intein-mediated cleavage [61]. For native chemical ligation, an N-terminal peptide segment containing a C-terminal thioester is chemoselectively ligated to a C-terminal peptide segment that has an N-terminal cysteine in aqueous solution, without protecting any functional groups in the peptides. With the intein-based *E. coli* expression system, it is now possible to produce larger protein segments with a C-terminal thioester group easily, which can subsequently be used for native chemical ligation. Using this approach, Xu et al. have demonstrated the domain-selective ^{15}N labeling of the SH2 domain of the Abl-kinase SH domain [62]. In this experiment, the ^{15}N-labeled SH2 domain containing an N-terminal cysteine capped with a specific proteolytic sequence, which can be removed, was expressed and purified in ^{15}N-labeled media. The protective N-terminal sequence was subsequently removed by proteolysis in order to create an N-terminal thionucleophile (N-terminal cysteine). The N-terminal segment of the SH3 domain was separately produced as the intein-fusion protein in unlabeled media and eluted with ethanethiol to form the C-terminal thioester. The unlabeled SH3 domain and ^{15}N-labeled SH2 domain were ligated in aqueous solution at pH 7 in the presence of thiophenol and benzyl mercaptan, achieving a yield of 70%.

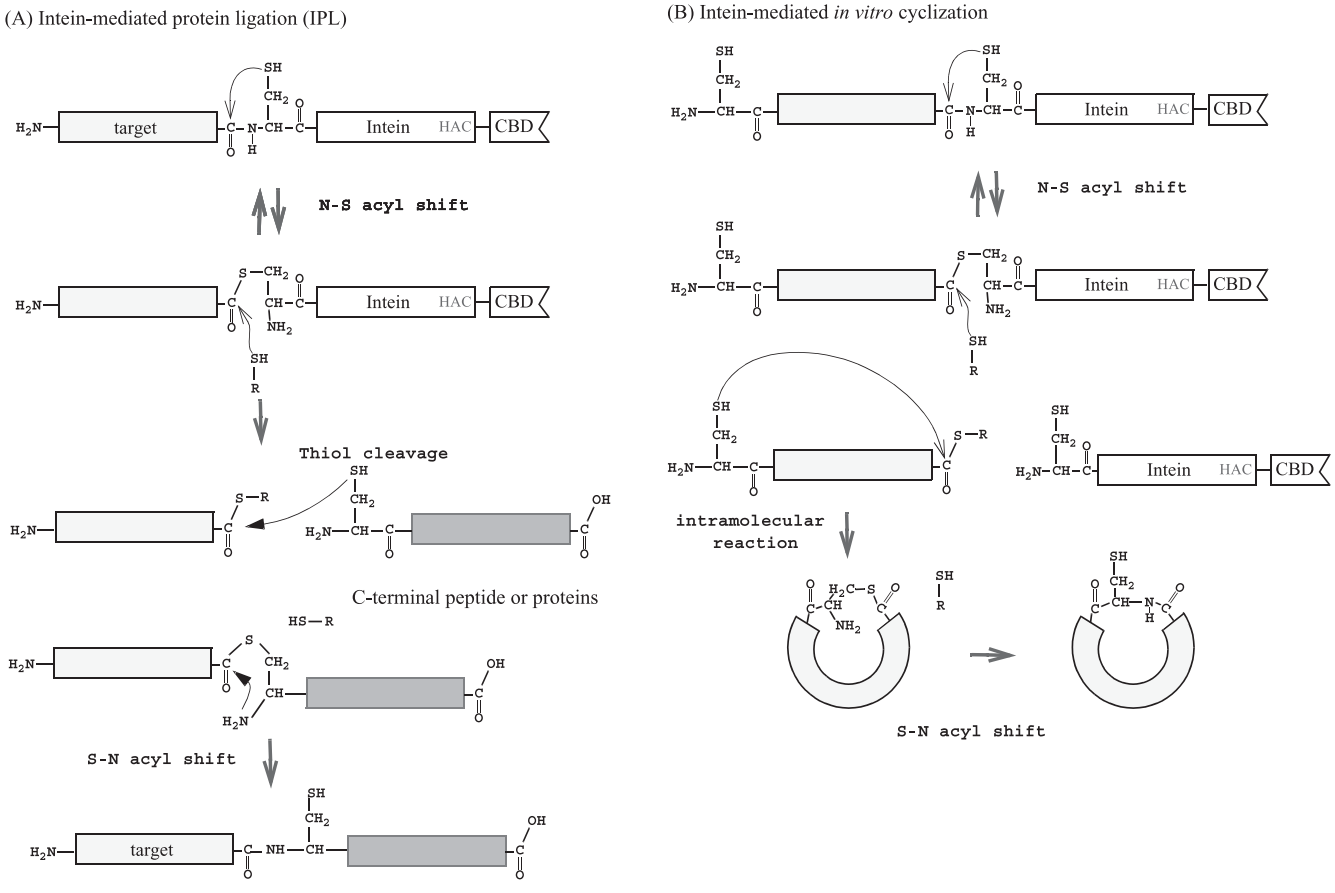


Fig. 1.4 **A** Intein-mediated protein ligation (IPL) (also called expressed protein ligation EPL). **B** The chemical mechanism of protein cyclization by the IPL/EPL approach. HAC denotes the sequence of the active site of the intein, e.g. His-Ala-Cys. CBD stands for chitin-binding domain.

It has also been demonstrated that multiple ligation steps can be performed with the IPL/EPL approach, thereby illustrating its potential use in central-segment labeling [63].

One of the intrinsic limits of the IPL/EPL system is the requirement for a cysteine residue at the site of protein ligation, which will replace the thioester group. Recently, it was shown that this requirement could be avoided by using a cleavable thiol-containing auxiliary group. Low et al. demonstrated protein ligation by introducing a cleavable thiol-containing auxiliary group, 1-phenyl-2-mercaptoethyl, at the alpha-amino group of a chemically synthesized peptide, which is removed upon protein ligation [64]. Unfortunately, this modification at the N-terminus could be difficult to introduce into proteins which are prepared from bacterial expression systems, and hence its use could be limited to situations where the C-terminal peptide can be chemically synthesized. Nevertheless, the removal auxiliary approach presents an opportunity for segmental isotope labeling regardless of the primary sequence.

A second approach that can be adopted to overcome the intrinsic requirement for cysteine at the N-terminus of C-terminal fragment utilizes the enzyme subtiligase, a double mutant of subtilisin, which is able to join two unprotected peptides. Thioester-modified proteins were shown to present good substrates of subtiligase [65]. However, although this approach could be potentially useful for general isotope labeling, the efficiency of this process remains to be proven.

It is noteworthy that there is another limiting factor in the choice of amino acid types at the junction sites which affect the enzymatic process of the intein. For example, in the case of *Sce*VMA (also called PI-*Sce*I) from the IMPACT system, proline, cysteine, asparagine, aspartic acid, and arginine cannot be at the C-terminus of the N-terminal target protein just before the intein sequence. The presence of these residues at this position would either slow down the N-S acyl shift dramatically or lead to immediate hydrolysis of the product from the N-S acyl shift [66]. The compatibility of amino acid types at the proximal sites depends on the specific inteins and needs to be carefully considered during the design of the required expression vectors. The specific amino acid requirements at a particular splicing site depends on the specific intein used and is thus a crucial point in this approach.

1.4.1.2 **Reconstitution of Split Inteins**

It has been demonstrated that an intein can be split into two fragments and reconstituted *in vitro* as well as *in vivo* to form an active intein capable of *trans* splicing [67–69]. This *trans*-splicing activity can be directly used for protein ligation as an alternative to the native chemical ligation step, which requires additional thionucleophile groups. Yamazaki et al. have applied *trans*-splicing to the segmental labeling of RNA polymerase subunit α by splitting an intein from *Pyrococcus furiosus*, PI-*Pfu*I (Fig. 1.5 B) [70]. Each half of the split intein fused to the N- (or C)-extein was produced separately, one in isotopically labeled and the other in unlabeled medium. The independently prepared protein fragments were expressed as inclusion bodies and purified under denaturing conditions. The two independently prepared fragments were reassembled and refolded in aqueous solution in order to form a functional intein, resulting in a ligated extein fragment and a spliced intein. The splicing reaction was found to be efficient at elevated temperature (70 °C), presumably because PI-*Pfu*I is a thermophilic enzyme. However, the general use

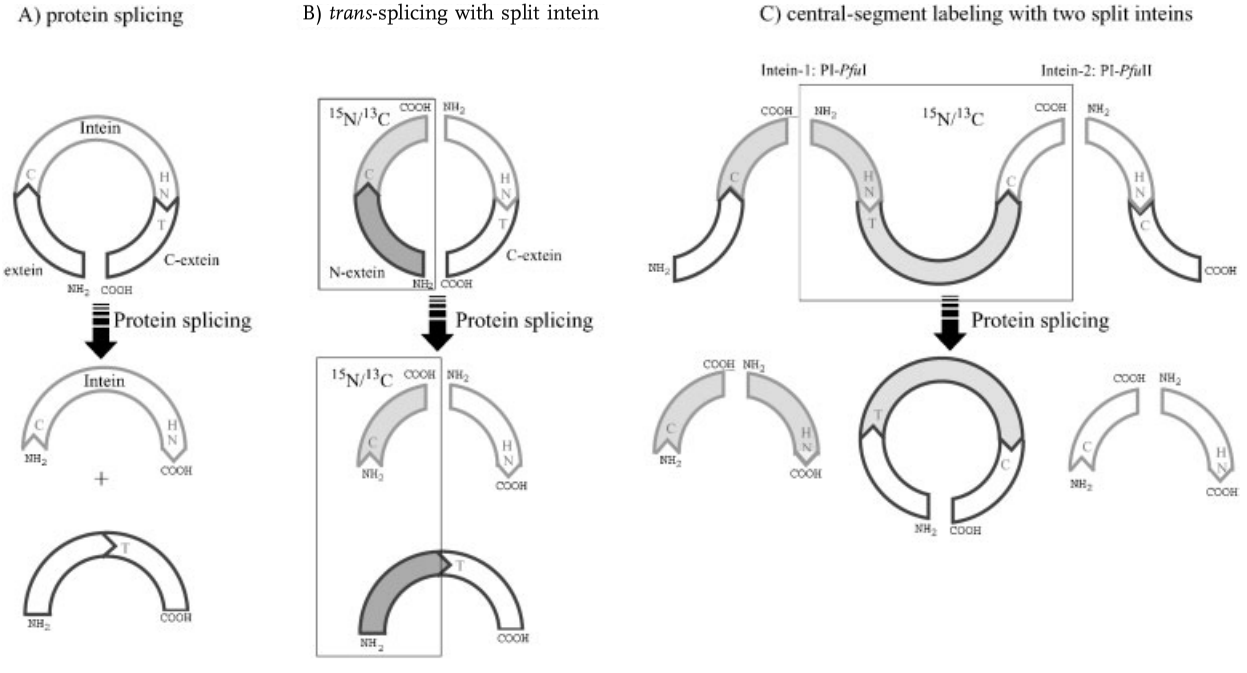


Fig. 1.5 **A** Natural protein splicing, **B** *Trans* splicing with a split intein. The two fragments can be prepared separately and reassembled *in vitro* to form an active intein domain for protein ligation. **C** Central segment labeling using two different split inteins. Residues involved in the protein splicing reaction are shown by a single character code of amino acids. N- and C-termini are indicated by NH_2 and COOH respectively.

of elevated temperatures may often be unfavorable, because many proteins denature at higher temperatures. Therefore, conditions for refolding and ligation such as temperature, pH, additives like glycerol etc. must be carefully optimized for each protein system. On the other hand, the split intein approach does not require any additional thionucleophile, in contrast to the IPL/EPL approach.

Remarkably, this method can even be extended to joining three segments by using two different inteins, PI-*Pfu*I and PI-*Pfu*II (Fig. 1.5 C). Otomo et al. have successfully demonstrated this approach by isotope labeling a central segment of the 370-residue maltose-binding protein [71]. Three protein fragments were constructed for the two ligation reactions. The first fragment was a fusion protein of the N-terminal domain of the split target protein and the N-terminal split intein-1. The second segment consisted of the C-terminal split intein-1 (PI-*Pfu*I), the central part of the split target protein fragment and the N-terminal split intein-2 (PI-*Pfu*II). The third segment contained the C-terminal split intein-2 fused to the C-terminal fragment of the split target protein. Ligation of the first and the second fragment was facilitated by intein-1, while the second and the third fragments were ligated by intein-2. These ligation reactions were highly specific because two different inteins were used. The second fragment was prepared in isotopically enriched medium, and hence the ligated protein was isotopically labeled only in the central part.

The intrinsic limitation of this approach, as in the case of the IPL/EPL approach, is the requirement for specific proximal residues near or at the ligation sites. In inteins, the residues at the junction where ligation occurs are highly conserved because of the chemical mechanism and typically require either cysteine, serine, or threonine [72]. Therefore, at least the N-terminal amino acid of the C-extein must be one of these residues, depending on which intein is used. In addition, the peptide sequence preceding the intein as well as the sequence following the ligation site play an important role for the efficiency of the protein-splicing reaction. For example, in the case of the DnaE intein, the five residues preceding the intein N-terminus and the three residues following the ligation site must seem to be native extein residues in order to achieve efficient splicing [73]. However, these sequential and structural determinants are presently not well understood for all known inteins. Otomo et al. have speculated that the flexibility of the junction region could be one of the important factors for ligation with PI-*Pfu*I. Such requirements would restrict the position of ligation sites, probably to (flexible) linker regions.

One advantage of the split intein over the IPL/EPL approach is the direct use of intein splicing activity, eliminating the requirement for additional thionucleophiles such as thiophenol. An another potential advantage is the ability to perform multiple ligations in a one-pot process, greatly simplifying the reaction procedure for the ligation of several fragments. In contrast, the IPL/EPL approach requires ligation reactions to be performed sequentially for multiple fragment ligations.

Although the use of split inteins for segmental isotope labeling has great potential, the number of inteins which have been adapted to this purpose is currently limited. Additional difficulties arise from the fact that the determinants influencing the success of an intein splicing reaction are not well understood. Moreover, the refolding requirements of split inteins could hinder its use as a general method for the ligation of protein fragments. Hence, further biochemical characterization of inteins is required for the advance of intein-mediated protein ligation methods.

1.4.2 Stabilizing Proteins by Intein-Mediated Backbone Cyclization

Limited protein stability often hampers successful structure elucidation by X-ray crystallography and/or NMR spectroscopy. Relaxation properties are usually improved at elevated temperatures, and multidimensional NMR experiments require sample lifetimes to extend over several days to weeks in order to acquire all the necessary data. In addition, the activity of contaminating proteases that are sometimes present in purified samples can be significant at the experimental temperatures. Therefore, the stability of a target protein can be a concern, in particular for expensive isotope-labeled proteins.

There have been many attempts to improve protein stability and protein properties, utilizing methods such as random mutagenesis, directed evolution, and rational protein design approaches. In general, these methods are far from straightforward and can be time-consuming. In addition, the stabilization of proteins without loss of function is not a trivial problem.

One new approach to stabilizing proteins without changing the primary sequence is to introduce backbone cyclization [74]. No mutations in the primary sequence are introduced by this method, although it might be necessary to insert a flexible linker comprising several residues to join the termini [74–76]. Polymer theory by Flory predicts an improvement in protein stability upon cyclization, because the entropy of the unfolded states should be reduced [77]. Backbone cyclization has long been used for small peptides to reduce the accessible conformational space. Recent advances in intein technology have opened up a new avenue for the cyclization of large proteins, because these proteins can be produced with recombinant techniques in bacterial expression systems [74, 76]. Cyclized proteins can be produced either *in vitro* or *in vivo*, as discussed in the following two sections. Statistical analysis of the structure database reveals that more than 30% of all known proteins might have termini in relatively close proximity, and hence the use of backbone cyclization to stabilize proteins has a good chance of success even in cases where the structure is not yet known [78].

1.4.2.1 *In vitro* Cyclization of Proteins

The IPL/EPL method described in Sect. 1.4.1.1 can be used for cyclization of the backbone polypeptide chains of proteins simply by introducing a nucleophilic thiol group at the N-terminus of the protein (Fig. 1.4 B, 1.6 A). This can be achieved either by creating an N-terminal cysteine by removing residues at the front of the cystein by specific proteolysis or by introducing a cysteine right after the methionine start codon, which is then removed enzymatically *in vivo* [74, 76]. Another method is to use the so-called TWIN system developed by New England Biolab, in which the target protein is fused into the middle of two different modified inteins (Fig. 1.6 B). The N-terminal nucleophilic cysteine is produced by an intein fused to the N-terminus of the target protein. At the same time the C-terminus can be transformed into the corresponding thioester by another intein fused to the C-terminus of the target protein [79]. This system circumvents the requirement for a specific proteolytic site in order to create the N-terminal cysteine, thereby simplifying the cyclization procedure.

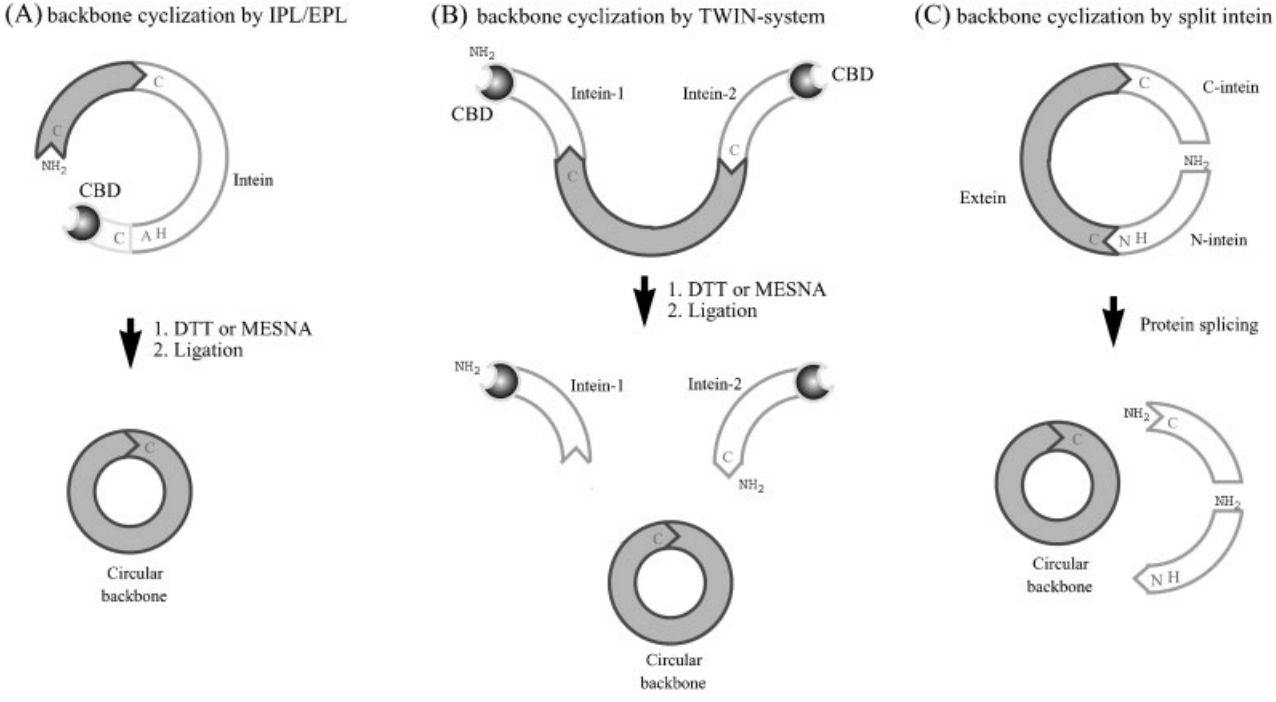


Fig. 1.6 **A** IPL/EPL cyclization with the IMPACT system, **B** backbone cyclization using the TWIN system, **C** backbone cyclization with split inteins. CBD stands for chitin-binding domain. N-termini are indicated by NH_2.

The biggest problems associated with *in vitro* cyclization methods using the IPL/EPL or the TWIN system are competing intermolecular reactions such as polymerization and hydrolysis, which complicate purification as well as reduce yields [74, 79].

1.4.2.2 *In vivo* Cyclization

The principle of *in vivo* cyclization is based on the circular permutation of precursor proteins containing an intein (Fig. 1.6C) [74, 75, 80, 81]. A naturally occurring split intein, DnaE from *Synechocystis* sp. PCC6803, was first successfully used for cyclization. However, similarly to the IPL/EPL approach, a mixture of linear and circular forms is obtained, presumably because of hydrolysis of an intermediate [73, 75]. On the other hand, artificially split inteins such as PI-*Pfu*I, DnaB, and the RecA intein have been successfully applied for *in vivo* cyclization, and only circular forms were observed [80–82], suggesting that the circular permutation approach is more suitable for cyclization. Compared to the IPL/EPL or the TWIN system, *in vivo* cyclization does not require any external thiol group for cyclization, similarly to protein ligation with split inteins. Moreover, there are no undesired products, such as linear forms or polymers, originating from intermolecular reactions.

1.4.2.3 Stability Enhancement by Backbone Cyclization

The effect of backbone cyclization was originally tested on BPTI, but no stabilization effects were observed, presumably because the three disulfide bridges reduce entropic gains [83]. Nevertheless, intein-mediated backbone cyclization has opened the way to a study of cyclization effects on protein stability (including membrane proteins) in more detail. Experimentally improved thermal and/or chemical stability has been shown to result from backbone cyclization of a range of proteins, including β-lactamase, DHFR, *E. coli* IIA^{Glc}, a destabilized mutant of SH3 domain and the N-terminal domain of DnaB [74, 75, 81, 82, 84]. An additional advantage imparted by backbone cyclization is essentially complete resistance to exopeptidases.

1.5 Alternatives to *E. coli* Expression Systems

Structural biology and the structural understanding of the genome, popularly called structural genomics, play an increasingly important part in drug discovery today. Fast and reliable protein expression tools are therefore of prime importance. To this end, the choice of protein expression systems has become increasingly important. While in the not so distant past, only *Escherichia coli*-based expression was used, today a variety of expression systems have been developed ranging from Archaebacteria to mammalian expression vectors. Needless to say, there is no universal expression system, and hence it is often necessary to balance various parameters to achieve optimal expression. For instance, considering the cost of isotopically enriched media, it can be advantageous to sacrifice some native characteristics of a recombinant protein in order to benefit from the higher yields that can be achieved in a more basic expression system. In contrast, specific modifications of the target protein (e.g. glycosylation) predominantly occur in certain cell types, which therefore require the development of special expression vectors.

Here, we describe the various alternatives to the use of *E. coli* expression hosts for heterologous gene expression. Advantages and disadvantages for the different expression systems are discussed, and practical aspects of expression technologies are also described. The feasibility of isotope labeling of recombinantly expressed proteins and their potential use for NMR is also discussed, since the costs and quantities of recombinant proteins produced depend on the system being used.

1.5.1 Expression Vectors

Traditionally, prokaryotic expression, especially employment of *E. coli*-based vectors, has been the system of choice. However, bacteria are unable to provide many vital components required for post-translational modifications including various forms of glycosylation or lipid attachment and protein processing, all of which can also be important for proper protein folding. For this reason, it is not surprising that much time and effort has been dedicated to the development of alternative systems, summarized in Tab. 1.2.

1.5.1.1 *Halobacterium salinarum*

Archaea are interesting organisms in the sense that they represent a phylogenetically distinct group of Prokarya, which is as distantly related to Eubacteria as to Eukarya [85]. *H. salinarum*, the best characterized Archaeon harbors a purple-colored plasma membrane consisting of a complex of one protein, bacterio-opsin (Bop) and its chromophore retinal in a 1:1 ratio [86]. The complex was named bacteriorhodopsin, and it forms typically highly ordered two-dimensional structures in the purple membrane, which allowed its purification and the determination of a high-resolution structure [87]. Recently, a system for heterologous gene expression was constructed for *H. salinarum* [88]. Fusion constructs between the Bop gene and heterologous sequences have been introduced into the *H. salinarum* expression vectors as follows: (i) C-terminally tagged bacteriorhodopsin [88]; C-terminal fusion with (ii) the catalytic subunit of *E. coli* aspartate transcarbamylase [88], (iii) the human muscarinic M1 receptor [88], (iv) the human serotonin 5-HT2c receptor [88], (v) the yeast α-mating factor Ste2 receptor [88]. The Bop-transcarbamylase fusion was well expressed, generating yields of 7 mg receptor per liter of culture. However, introduction of tags at the C-terminus of the Bop gene significantly reduced its expression levels. This was partly because of a decrease in Bop-fusion protein mRNA levels compared to the wild-type Bop. More dramatically, expression studies of fusion constructs between the Bop gene and mammalian GPCRs (G protein-coupled receptors, human muscarinic M1 receptor, platelet-activating receptor and angiotensin-1 receptor) failed to detect fusion protein expression detected by Western blotting [89]. In this case, coding region swaps between Bop and GPCRs improved RNA yields and resulted in detectable levels of Ste2 receptor. These results suggest that *H. salinarum* can be considered as a potentially interesting alternative. The simple and rapid large-scale culture technology is attractive; however, improvements are still required concerning heterologous gene expression. In addition, questions related to codon usage and fusion construct optimization need to be properly addressed.

Tab. 1.2 Features of expression systems

Vector	*Advantage*	*Disadvantage*
Halobacterium salinarum	Rapid expression Colorimetric expression Easy scale-up	Cloning and transformation complicated Requires fusion protein strategy Lack of post-translational modifications
E. coli	Rapid cloning procedure High expression levels Easy scale-up	Toxicity of foreign membrane proteins Lack of post-translational modifications
Saccharomyces cerevisiae	Good secretion machinery Post-translational modifications Easy scale-up	Selection procedure required Tendency to overglycosylation Thick cell wall complicates purification
Schizosacharomyces pombe	Genetics well understood Mammalian promoters applicable	Selection procedure required
Pichia pastoris	High GPCR expression levels	Selection procedure required
Baculovirus	Improved procedure Infection of insect cells High expression yields	Relatively slow virus production Different post-translational processing
Stable mammalian	High authenticity Large-scale set up	Slow procedure to generate cell lines Low recombinant protein yields Stability problems
Transient mammalian	High authenticity Relatively fast methods	Scale-up difficult Transfection methods cell line-specific
Semliki Forest virus	Rapid virus production Broad host range Extreme yields of receptors Large-scale technology established	Safety concerns

1.5.1.2 *Saccharomyces cerevisiae*

Yeast expression vectors have been among those most commonly used since the beginning of gene technology. Vectors based on baker's yeast, *Saccharomyces cerevisiae,* have been especially popular for robust expression of many types of recombinant proteins [90]. For instance, the first commercially available recombinant vaccine, the hepatitis B surface antigen vaccine, was produced from an *S. cerevisiae* vector [91]. Many other recombinant proteins have also been efficiently expressed in yeast including α1-Antitrypsin [92], insulin [93], Epstein-Barr virus envelope protein [94], superoxide dismutase [95] and interferon-α [90].

The genetics and fermentation technology of *S. cerevisiae* are well characterized. Several strong yeast promoters like alcohol dehydrogenase (ADH1), galactose (GAL1/GAL10), 3-phosphoglycerate kinase (PGK) and mating Factor-α (MFα1) have been applied as well as the selection marker genes β-isopropylmalate (LEU2) and oritidine 5′-decarboxylase (URA3) [90]. Among transmembrane proteins, the *S. cerevisiae* α-Factor receptor Ste2p

has been expressed with C-terminal FLAG and His_6 tags [96]. Ste2p belongs to the family of GPCRs with a 7-transmembrane topology. Yields of up to 1 mg of almost homologous receptor were obtained, and the purified receptor was reconstituted into artificial phospholipid vesicles. However, restoration of ligand-binding activity required the addition of solubilized membranes from an Ste2p negative yeast strain. Also, the human dopamine D1A receptor was expressed with C-terminal FLAG and His_6 tags in *S. cerevisiae*, which allowed for purification and reconstitution of receptor [97].

1.5.1.3 *Schizosaccharomyces pombe*

Another yeast strain that has received much attention as an expression host is the fission yeast *Schizosaccharomyces pombe*. In contrast to *S. cerevisiae*, no budding occurs, and the yeast only reproduces by means of fission and by spores [98]. Two types of expression vectors have been developed for *S. pombe*. The chromosomal integration type of vector maintains the foreign gene stably in the chromosome [99], and the episomal vector replicates autonomously in yeast cells [100]. Some mammalian promoters like the human chorionic gonadotropin and CMV promoters are functional in *S. pombe* [101]. The fission yeasts possess many similar features to mammalian cells. *S. pombe* has a signal transduction system similar to the mammalian G protein-coupling system [102], and the mammalian endoplasmatic reticulum retention signal KDEL is also recognized [103]. The glycosylation pattern for *S. pombe* is also different from that of *S. cerevisiae* and other yeast species.

A wide range of mammalian proteins have been expressed in *S. pombe*. In a successful example, the human lipocortin I comprised 50% of the total soluble proteins in yeast cells and showed high activity, indicating that the post-translational modifications were mammalian-like [104]. Membrane proteins including cytochrome P450 were expressed at ten times the levels of those in other yeast systems [105]. Also, GPCRs have been expressed in *S. pombe*, where the human dopamine D2 receptor was correctly inserted into the yeast cell membrane and demonstrated expression levels three times those of *S. cerevisiae* [106].

1.5.1.4 *Pichia pastoris*

The methylotrophic yeasts including *Pichia pastoris, Hansenula polymorpha* and *Kluyveromyces lactis* have become potentially attractive expression hosts for various recombinant proteins [107]. In addition to the relative ease with which molecular biology manipulations can be carried out, *P. pastoris* has demonstrated a capacity for performing many post-translational modifications such as glycosylation, disulfide bond formation and proteolytic processing [108]. *P. pastoris* utilizes the tightly methanol-regulated alcohol oxidase 1 (AOX1) promoter, and the vector is integrated as several copies into the yeast host genome. When human insulin was expressed in *P. pastoris*, the secretion was comparable to that obtained for *S. cerevisiae*. Peptide mapping and mass spectrometry confirmed identical processing of human insulin in yeast and mammalian cells. *P. pastoris* has also been used as a host for expression of GPCRs [109]. The mouse 5-HT_{5A} receptor and the human β_2-adrenergic receptor were fused to the prepropeptide sequence of the *S. cerevisiae* α-factor, which enhanced the expression levels by a factor of three. Multiple chromosomal integrations further improved the expression twofold. In the case of the β_2-adre-

nergic receptor, addition of the antagonist alprenol to the culture medium increased the number of specific binding sites. A similar but weaker effect was seen for the 5-HT_{5A} receptor after addition of yohimbine. The binding activity for the β_2-adrenergic receptor and the 5-HT_{5A} receptor were 25 pmol and 40 pmol, respectively, per milligram of membrane protein. The pharmacological profiles assayed by ligand-displacement analysis were similar to those obtained from receptors expressed in mammalian cells.

1.5.1.5 **Baculovirus**

Heterologous gene expression has been studied to a great extent in insect cells with the aid of baculovirus vectors. The popularity of the baculovirus system is mainly due to the high expression levels obtained for various recombinant proteins resulting from the use of strong viral promoters [110]. Generally, heterologous genes are expressed from the polyhedrin promoter of *Autographa californica* in several insect cell lines such as *Spodoptera frugiperda* (Sf9), *Trichoplusia ni* (Tn), *Mamestra brassicae* and *Estigmene acrea* [111]. Although baculovirus vectors have been used for expression of various mammalian recombinant proteins, a limitation has been the differences in the N-glycosylation pathway between insect and mammalian cells. However, *Estigmene acrea* cells can produce a similar glycosylation pattern as occurs in mammalian cells [112]. Moreover, modifications of baculovirus vectors by replacing the polyhedrin promoter with a CMV promoter made it possible to carry out expression studies in mammalian cell lines [113]. Using this so-called BacMam system, milligram quantities of a cellular adhesion protein (SAF-3) could be produced in CHO cells [114]. Baculovirus vectors have been used extensively to express GPCRs and ligand-gated ion channels [115]. Expression levels up to 60 pmol receptor per milligram have been obtained, which has led to relatively efficient purification procedures. In attempts to further enhance the expression level of the β2-adrenergic receptor, an artificial sequence was introduced, which resulted in approximately double the receptor levels in insect cells [116].

1.5.1.6 **Transient Mammalian Expression**

Several approaches have been taken to develop efficient transient mammalian expression systems. The most straightforward process has been to engineer expression vectors with strong promoters. Relatively high expression levels for cytoplasmic and even some transmembrane proteins have been obtained in adherent cells on a small scale. However, a major problem arises in the scale-up of these growth procedures, which are also relatively expensive [117]. In spite of this difficulty, transient transfection experiments using a modified calcium-phosphate coprecipitation method have been carried out in HEK293 EBNA cells adapted to suspension cultures grown on a 100 L scale [118]. More than 0.5 g of a monoclonal antibody was produced from this system, although similar methods have yet to be developed for receptor expression.

1.5.1.7 **Stable Mammalian Expression**

Generation of various cell lines (BHK, CHO and HEK293) with the target gene inserted downstream of a strong promoter into the genome is a common approach to achieve overexpression in mammalian hosts. However, one drawback is the time-consuming procedure involved in the establishment of stable cell lines, which generally requires 6–8 weeks. Other

problems associated with this approach are related to the relatively low expression levels and the instability of generated cell lines. These issues have been addressed by engineering inducible expression systems, which are usually based on tetracyclin-based regulation (Tet on-off systems) [119]. A highly interesting development has been the generation of a cold-inducible expression system based on the Sindbis virus replicon [120]. Because of a point mutation in one of the replicase genes, the viral replicase complex is totally inactive at 37 °C, whereas a shift in temperature below 34 °C results in high replication activity and high levels of heterologous gene expression. Using this approach, the serotonin transporter gene, characterized by its low expression levels in any system tested, generated reasonable yields (approximately 250,000 copies per cell) [121].

1.5.1.8 **Viral Vectors**

The two common features that have made viral vectors attractive for recombinant protein expression are their high infection rates for a broad range of mammalian cell lines and their strong promoters. Adenovirus vectors have shown high expression levels in, for instance, human embryonic kidney (HEK293) cells, but their use has been to some extent restricted by the fairly complicated virus generation procedure [122]. Another potentially useful class of viruses are the poxviruses. Recombinant gene expression of herpes simplex virus thymidine kinase (TK) has been established for vaccinia virus vectors [123]. Moreover, the engineering of a hybrid bacteriophage-vaccinia virus vector by applying the T7 promoter has simplified and broadened the use of pox virus-based expression systems [124]. However, vaccinia vectors are still quite complicated to use for rapid recombinant protein expression, and they have instead found applications in the field of vaccine development. Alphavirus vectors have proven to be highly efficent for heterologous gene expression. Both Semliki Forest virus- (SFV) [125] and Sindbis virus-based [126] expression vectors have been engineered to rapidly generate high-titer recombinant particles, which are susceptible to a broad range of mammalian cell lines and primary cells in culture [127]. Typically, both GPCRs and ligand-gated ion channels have been expressed at extreme levels, i.e. up to 200 pmol receptor per milligram protein [128]. Large-scale SFV technology has been established, which has generated large quantities of, for instance, mouse serotonin 5-HT3 receptor, purified to homogeneity and subject to structural characterization [129]. Moreover, several GPCRs have been expressed at levels of 5–10 mg receptor yields per liter suspension culture of mammalian host cells [130], which has provided material for large-scale purification.

1.5.2 **Comparison of Expression Systems**

Comparison between different systems for transmembrane protein expression are always difficult to make, and they obviously reflect individual needs and are strongly influenced by personal experience. It is, however, important to define the usefulness of each system by taking into account different aspects such as ease of handling, expression levels, time and labor requirements, safety, costs, and the quality of the produced recombinant protein (Tab. 1.3).

Obviously, prokaryotic systems are easy to use, the costs for their large-scale applications are low, and no safety risks are involved. The drawbacks are their limited capacity for post-translational modifications and generally low yields of complex mammalian

Tab. 1.3 Application of expression systems for mammalian membrane proteins

Vector	*Handling*	*Expr./Scale-up*	*Authenticity*	*Time/Labour*	*Safety*	*Costs*
Halobacterium salinarum	Easy	Low/easy	Low	Short/easy	High	Low
E.coli	Easy	Moderate/easy	Low	Short/easy	High	Low
Saccharomyces cerevisiae	Rel. easy	High/easy	Mod.	Mod./mod.	High	Mod.
Schizosacharomyces pombe	Rel. easy	High/easy	Mod.	Mod./mod	High	Mod.
Pichia pastoris	Rel. easy	High/easy	Acceptable	Mod./mod	High	Mod.
Baculovirus	Mod.	High/mod.	Rel. high	Mod./mod.	High	Rel. high
Stable mammalian	Difficult	Low/difficult	High	Long/intensive	High	High
Transient mammalian	Difficult	Mod./difficult	High	Mod./mod.	High	High
Semliki Forest virus	Easy	Extreme/easy	High	Short/easy	Of concern	High

transmembrane proteins. Yeast expression systems are competitive with bacterial vectors with respect to scaleability, costs and safety. Although the time required from gene construct to expressed recombinant protein is slightly longer, the yields are significantly higher. Yeast can also provide some post-translational and protein-processing capacity, although not identical to mammalian cells. The thick yeast cell wall is of some concern, because it makes the purification of intracellular and transmembrane proteins more complicated. Baculovirus vectors have the advantages of high expression levels in insect cells and the fairly simple though more expensive scale-up procedure. Needless to say, the optimal host for expression of mammalian transmembrane proteins from the viewpoint of a molecular biologist must be mammalian cell lines. Obviously, the drawbacks with the conventional transient or stable expression approaches have been the labor intensiveness and time-consuming procedures. The yields have also been disappointingly low and the costs for large-scale production high. Viral vectors are therefore potentially very attractive because of their high capacity for gene delivery and extreme expression levels of heterologous genes. Naturally, viral vectors always pose a higher safety risk of possible infection of laboratory personnel. Not only have the rapidly generated replication-deficient SFV vectors been demonstrated to be free from any contaminating replication-proficient particles, but also the amounts of residual infectious particles associated with cells or even in the medium are negligible [131]. Today the SFV system is classified as BL1 level in several European countries (Germany, Finland, Sweden, Swizerland, UK) but is currently BL2 in the United States. Moreover, the broad host range provides an additional opportunity for the study of gene expression and protein processing in several mammalian host cell lines in parallel. Evidently, large-scale cultivation of mammalian cells is more expensive than bacterial or yeast cell equivalents, but these costs are significantly reduced by using a serum-free medium for suspension cultures.

1.5.3 Isotope Labeling and NMR

Recent developments in technologies within structural biology should also play an important part for transmembrane proteins. The potential to incorporate stable isotopes would facilitate structure determination by NMR techniques. Although NMR technologies were long considered to be applicable only to smaller proteins, the development of transverse relaxation-optimized spectroscopy (TROSY) has made it possible to use NMR for larger proteins also [5], even integral membrane proteins. For example, the OmpX and OmpA integral membrane proteins of *E. coli* were labeled with $^{13}C/^{15}N/^{2}H$ isotopes and overexpressed as inclusion bodies in bacterial cells. After solubilization in 6 M Gdn-HCl and reconstitution in detergent micelles, solution NMR techniques could be used to identify regular secondary structural elements [132].

Proteins that require non-*E.coli* expression systems are generally too large in size to be used for NMR studies without uniform ^{15}N, ^{13}C and and sometimes ^{2}H labeling. Hence, when using the expression systems described in this chapter, it is important to ensure that cells can be grown on defined, isotopically enriched media. This fact at the moment excludes, for example, the use of fetal calf serum. However, special isotope-enriched defined media are available from commercial suppliers which present fully rich, serum-free (protein-free) media at reasonable costs containing labeled amino acids and carbohydrates and which can be used to effectively express heterologous proteins in insect cells applying the baculovirus vector system [133]. Similarly, rich media are also available for expression in *S. cerevisiae* [134]. Conversely, the methylotropic yeast *P. pastoris* can be grown on minimal media, facilitating its use as a potential host for isotope labeling. In fact, there have been a number of successful examples where *P. pastoris* was used to produce isotopically enriched samples for solution NMR studies, including a cysteine-rich glycosylated domain of thromobomodulin [135], a glycosylated EGF module [136], domains from rat calretinin [137], and tick anticoagulant peptide [138]. Even more importantly, it was shown that expression in yeast enables the production of heterologous proteins in deuterated form [139], which would be very difficult or even impossible to achieve in mammalian expression systems because of the cell toxicity of deuterated water. Similarly, *H. salinarum* can easily be grown in D_2O and on the above-described algae hydrolyzates. Moreover, residue type-specific labeling is possible in this host [26]. Examples of proteins expressed in this host in isotopically labeled form can be found in the literature, e.g. Refs. [25] and [140]. In order to obtain proteins with more native-like glycosylation patterns, CHO mammalian cell expression systems have also been developed for NMR sample generation [141–143]. However, the requirement for rich media in mammalian cell-based expression systems combined with difficulties associated with generating high expression levels have to date prevented a more widespread utilization for NMR. Nonetheless, considering the range of proteins that may be inaccessible to *E. coli*-based expression systems as well as the potential information that can be gained by studying the native-like post-translationally modified forms of proteins, the adaptation of heterologous protein expression systems to the purpose of isotope labeling for solution NMR clearly requires further development.

Reconstitution of GPCRs in appropriate membrane mimetics or membrane preparations is still very difficult, and success requires efficient expression systems to yield enough protein material. As large quantities of an increasing number of recombinant proteins become available, it will be possible to develop techniques for solubilization, purification and reconstitution in a high-throughput format. The first global structural genomics project for membrane proteins, MePNet (Membrane Protein Network) was recently initiated with the aim of comparing the overexpression of 100 GPCRs in three systems based on *E. coli, P. pastoris* and SFV vectors [144]. The goal of this three year program is to verify the expression levels for the 100 targets and establish platforms for solubilization, purification and crystallization technologies which should form a solid base for obtaining novel high-resolution structures of GPCRs. Technological developments arising from this initiative should also benefit the field of solution and solid-state NMR.

1.5.4 Target Proteins

The choice of expression system is to a large extent dictated by the type of target protein. In general, many soluble proteins, which are fairly easy to express and have a low molecular weight, are efficiently expressed from bacterial vectors. As a rule of thumb one can suggest that *E. coli*-based expression should be used whenever possible. However, *E. coli* vectors are not suitable for expression of many authentic mammalian transmembrane proteins. Additionally, the near-completion of the human genome sequence has revealed a multitude of genes as potential targets for structural analyses. Many of these include transmembrane proteins, whose properties quite often make them more difficult to express. Many of these transmembrane proteins, including receptors and channels, are important targets for drug discovery. GPCRs alone stand for approximately 50% of drug targets today. Moreover, a quarter of the top 200 drugs in the United States are based on GPCRs, and the annual sales in 2000 reached more than 18.5 billion US$ [145]. It is therefore understandable that so much interest has been focused on transmembrane proteins today. At present, it is possible to express complicated transmembrane receptors in several vector systems at reasonably high levels, cost-effectively in a near-to-native state. As more recombinant proteins become available, the solubilization, purification and crystallization for NMR technology can also be developed, which will contribute to the achievement of rates of structure determination similar to those common for soluble proteins today. However, the determination of high-resolution structures of membrane proteins, and particularly mammalian ones, has been modest compared to soluble proteins (Tab. 1.4). Recently, methodological advances in NMR spectroscopy have significantly raised the size limit amenable to NMR investigations, and hence membrane protein structures determined by NMR have also appeared in the literature. The outer membrane protein A from *E. coli* was reconstituted in DPC micelles and the structure determined by heteronuclear multidimensional NMR [146]. Similarly, the structure of the outer membrane protein OmpX was elucidated in DHPC micelles [147]. Both proteins were expressed in *E. coli*. However, in the case of OmpA, a series of selectively ^{15}N-labeled samples was prepared, whereas in the case of OmpX, spectroscopy was performed on a single triply ^{2}H,^{13}C,^{15}N-labeled sample. The best successes have been achieved when it

was possible to isolate proteins from their natural sources instead of using recombinantly expressed forms. The results have been particularly poor for GPCRs, and the only example of a mammalian 7-transmembrane protein for which a high-resolution structure has been obtained is bovine rhodopsin [148].

1.6 The Use of Cell-Free Protein Expression for NMR Analysis

The expression of proteins usually requires optimization by trial and error, since conditions leading to production of high yields of active protein are difficult to predict. Therefore, additional time-consuming steps are often involved that may require changes in the choice of expression system, of reaction conditions, and of tests for the expression of protein sub-domains. In structural genomics programs this problem was rapidly identified as a major bottleneck, since high-throughput production of soluble proteins in milligram amounts is essential for success. Hence, systems that allow rapid, productive expression, are easy to manipulate, and can be run in a parallel format are highly desired in order to rapidly screen for the best conditions.

Until recently, cell-free protein expression (also sometimes erroneously named *in vitro* protein expression) did not exhibit the productivity required for preparation of NMR samples, especially considering the high cost of using isotopically labeled starting material. Rather, it was exclusively used as an analytical tool that served to verify correct cloning or to study promotor sites. Because of the very low yields, detection of the expressed product usually required incorporation of a radioactive label (usually via ^{35}S-methionine).

This situation changed fundamentally when Spirin published his substantial improvements in 1988 that resulted in much higher product yields [149]. He developed a set-up in which the coupled transcription/translation reaction could be continuously supplied with all the essential low-molecular-weight components (nucleotides, amino acids, energy components) while maintaining a constant reaction volume by continuously removing the product through a membrane (continuous-flow cell-free system, CFCF). Under these conditions, the system would remain active for more than 1 day, in contrast to the usual upper limit of 2 h which was observed in the conventional batch system. The prolonged expression period indicated that sufficient amounts of all factors necessary for translation were still available in the system, although the membrane clearly allowed the leakage of at least some of them. Soon, these findings were used to run the reaction in a more robust dialysis mode, keeping the reaction volume constant while feeding the system with all necessary low-molecular-weight consumables.

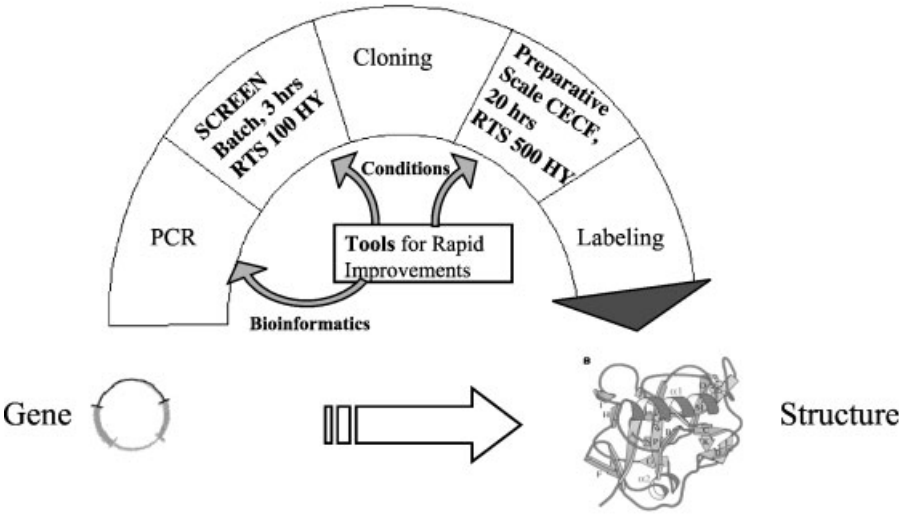
In the following years the productivity of the method was tremendously improved, so that milligram amounts of protein per milliliter of reaction mixture could be obtained [150–154]. This finally opened the door to cell-free protein expression to be used for the production of isotopically enriched proteins suitable for NMR analysis. However, although widely considered to be a promising method for labeling proteins at specific positions and therefore facilitating the process of chemical shift assignment as well as reducing spectral overlap [155, 156], in fact it was used in very few laboratories worldwide. This was mainly due to the fact that the preparation of the lyzate was tedious and the expression levels of each batch varied from lot to lot.

Recently, a system for cell-free protein production has became commercially available (the Rapid Translation System, RTS) [157]. In the following sections this system will be described, and advantages as well as limitations will be discussed.

1.6.1 The Cell-Free Protein Expression System RTS

The RTS system includes two different technology platforms for cell-free protein expression as well as a number of tools for finding optimal conditions (Scheme 1.1). All expression systems use the T7-polymerase for transcription and an *E. coli* lyzate with reduced nuclease and protease activity for translation. The conditions are optimized for a coupled transcription/translation reaction so that the DNA can be directly used as the template.

The first platform (RTS 100 HY) is designed as a screening tool. It uses the batch format, so that the reaction time does not exceed 3 h. In particular, in the RTS 100 HY system the exonuclease activity is reduced, so that the direct use of PCR-generated DNA templates is possible. To facilitate the generation of the PCR templates there is a special product available (linear template kit), which introduces all regulatory elements [T7-promotor, gene10 enhancer sequence and the Shine-Dalgarno (RBS) sequence]. Consequently, RTS 100 HY can be used for the rapid evaluation of the best template, without spending time with cloning, and for optimization of the reaction conditions (e.g. temperature, choice of additives like detergents, chaperones etc.). In addition, a bioinformatic tool (the program ProteoExpert) facilitates the process of designing the optimal template by analyzing and improving the secondary structure of the corresponding mRNA (without changing the amino acid sequence of the protein).



Scheme 1.1 From gene to protein structure via cell-free protein expression.

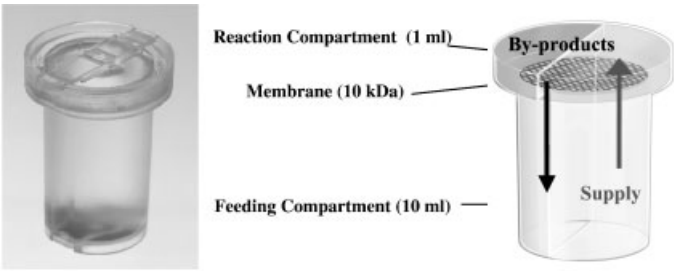


Fig. 1.7 Device of the RTS 500 format using the Continuous Exchange Cell-Free principle.

The second platform (RTS 500 HY) is designed for the production of proteins on a preparative scale. It is based on the CECF principle and utilizes a device with two chambers (Fig. 1.7). This design, with a proper choice of reaction conditions, gives a reaction time of 24 h, yielding up to 6 mg of protein per ml. A scaled-up version is also available (RTS 9000 HY), providing up to 50 mg protein per run.

In all formats, the amino acids are supplied separately, so they can be conveniently exchanged for labeled ones.

1.6.2 From PCR Product to ^{15}N-Labeled Protein

As an example of how cell-free protein expression can be used to rapidly generate a protein sample suitable for NMR analysis, an SH3-domain (8 kDa) was expressed using the RTS system [158]. Since it was initially found that the yield using a template carrying the wild-type sequence was too low, the sequence of the template was analyzed using the ProteoExpert program. The suggested sequences were subsequently evaluated by running expressions in the batch mode (RTS 100) using PCR-generated templates (linear template kit). As a result, all ten test sequences showed significantly higher yields than those with the original wild-type RNA template (data not shown). One of these was selected and ligated into the TOPO Cloning vector (Invitrogen Corp.) for expression in RTS 500 HY. Comparison with expression levels obtained with the wild-type template showed that also by using a circular template the yield could be improved more than fivefold (Fig. 1.8), resulting in approximately 3 mg product/mL reaction mix. In the next step, uniformly ^{15}N-labeled protein was produced by using a mixture of ^{15}N-labeled amino acids. Expression levels of the ^{15}N-labeled reaction were identical to the first reaction performed with unlabeled amino acids. The product was purified to homogeneity, and a [^{15}N,^{1}H]-HSQC spectrum was obtained (Fig. 1.9) to confirm the overall integrity of the protein fold.

For the residue-type assignment of the cross peaks, the SH3 domain was expressed with specifically labeled glycine or arginine residues by using amino acid mixtures where only Gly and Arg were ^{15}N-labeled. Again, yields for the labeled proteins were identical to those of the unlabeled product. Analysis of the corresponding [^{15}N,^{1}H]-HSQC spectra

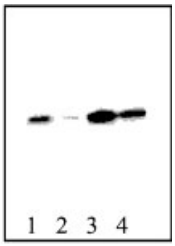


Fig. 1.8 Western blot analysis (via His_6-tag) of the expression of the SH3 domain using the wild-type DNA-sequence (lanes 1 and 2) or the optimized DNA-sequence (lanes 3 and 4). On lanes 1 and 3 0.25 µL DNA were loaded on the gel, while on lanes 2 and 4 0.125µL were applied.

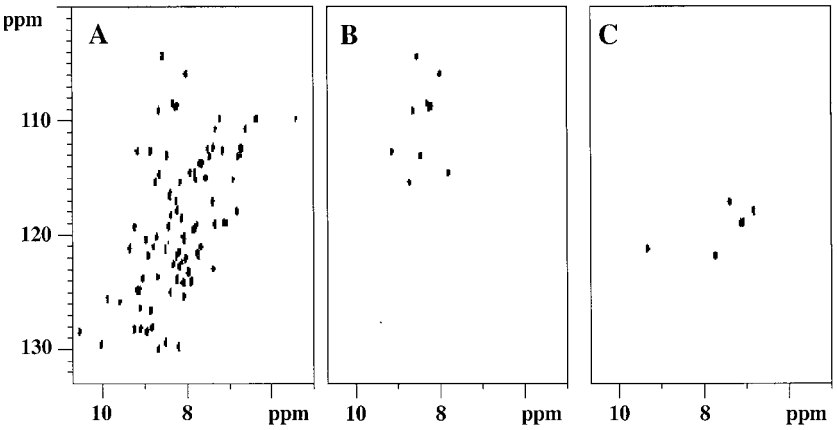


Fig. 1.9 [^{15}N,^{1}H]-HSQC spectrum of ^{15}N-uniformly labeled SH3 domain (**A**) and of a sample selectively labeled with only ^{15}N-Gly (**B**) or ^{15}N-Arg (**C**).

(Fig. 1.9) verified that the numbers of cross peaks for the ^{15}N-Gly as well as for the ^{15}N-Arg labeled protein were identical to the predicted ones (10 for the 10 Gly and 6 for the 3 Arg, respectively). Importantly, no scrambling of the ^{15}N-labels could be detected. All cross peaks were contained in the [^{15}N,^{1}H]-HSQC spectrum of the uniformly labeled protein (Fig. 1.9) and no additional signals occurred. Therefore, by simply overlaying the HSQC spectra, the cross peaks belonging to the glycine and the arginine residues could be readily assigned.

1.6.3 Discussion and Outlook

This example demonstrates how cell-free protein expression can be used to rapidly optimize the reaction conditions as well as to shorten the time for spectral assignments by producing individually labeled proteins.

To discuss in general the applicabilities of cell-free protein expression technology [159], two features are most valuable. First, only the protein of interest is produced, because the highly active T7-polymerase is used for translation. Consequently, labeled amino acids are almost exclusively incorporated into the newly produced protein. Since the labeled amino acids can be supplied as a mixture or added individually, the time for assigning the cross peaks to the particular amino acid(s) can be significantly reduced. This approach will enable partial assignments to be made in molecules that are far too large to allow spectral assignments from uniformly labeled protein by classical methods.

The second important feature is that cell-free protein expression can be considered as an "open system", meaning that no lipid membrane barriers are present. Consequently, chemicals, proteins (e.g. chaperones) as well as PCR-generated templates can be added directly to the reaction solution. Even major changes of the reaction conditions are possible (e.g. using a redox system to produce active proteins containing correctly formed disulfide bonds [160]). These features, as well as the ease of sample handling, dramatically reduce the time for optimizing the expression conditions (in fact, pipetting robots can be designed to run the reactions). Moreover, proteins can be synthesized which display cell toxicity and which therefore can hardly be expressed in classical systems.

However, certain limitations do exist that need to be considered. Although enzymes necessary for post-translational modifications can be added, in principle there is currently no productive system available for the preparation of glycosylated proteins, although some interesting results have already been obtained [161]. Also, the expression of functional membrane proteins in quantities necessary for structural analysis will be a challenging task for the future.

Nonetheless, the speed and flexibility of this emerging technology could provide the key to meeting the demands of high-throughput structure determinations.

1.7 References

1 Bax, A.; Grzesiek, S. *Acc. Chem. Res.* **1993**, *26*, 131–138.
2 Zimmerman, D.E.; Kulikowski, C.A.; Huang, Y.P.; Feng, W.Q.; Tashiro, M.; Shimotakahara, S.; Chien, C.Y.; Powers, R.; Montelione, G.T. *J. Mol. Biol.* **1997**, *269*, 592–610.
3 Yamazaki, T.; Lee, W.; Arrowsmith, C.H.; Muhandiram, D.R.; Kay, L.E. *J. Am. Chem. Soc.* **1994**, *116*, 11655–11666.
4 LeMaster, D.M.; Richards, F.M. *Biochemistry* **1988**, *27*, 142–150.
5 Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366–12371.
6 Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K. *J. Am. Chem. Soc.* **2000**, *122*, 7543–7548.
7 Neri, D.; Szyperski, T.; Otting, G.; Senn, H.; Wüthrich, K. *Biochemistry* **1989**, *28*, 7510–7516.
8 Muchmore, D.C.; McIntosh, L.P.; Russel, C.B.; Anderson, D.E.; Dahlquist, F.W. In *Methods Enzymol.*; Oppenheimer, N.J., James, T.L., Eds.; Academic Press: San Diego, **1989**; Vol. 177, pp 44–73.
9 Patzelt, H.; Simon, B.; Ferlaak, A.; Kessler, B.; Kühne, R.; Schmieder, P.; Oesterhelt, D.; Oschkinat, H. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9765–9770.
10 Evans, T.C., Jr.; Benner, J.; Xu, M.Q. *Protein Sci.* **1998**, *7*, 2256–2264.
11 Pardi, A. *Curr. Opin. Struct. Biol.* **1992**, *2*, 832–835.
12 Markley, J.L.; Kainosho, M. In *NMR of Macromolecules: A Practical Approach*; Roberts, G.C.K., Ed.; Oxford University Press: Oxford, **1993**; pp 101–152.
13 LeMaster, D.M. *Progr. NMR Spectrosc.* **1994**, *26*, 371–419.
14 Goto, N.K.; Kay, L.E. *Curr. Opin. Struct. Biol.* **2000**, *10*, 585–592.
15 Bolton, P.H. *Progr. NMR Spectrosc.* **1990**, *22*, 423–452.
16 Kainosho, M. *Nature Struct. Biol.* **1997**, *4 Suppl.*, 858–861.
17 Lian, L.-Y.; Middleton, D.A. *Prog. NMR Spectrosc.* **2001**, *39*, 171–190.
18 Makrides, S.C. *Microbiol. Rev.* **1996**, *60*, 512–538.
19 Crespi, H.L.; Katz, J.J. *Methods Enzymol.* 1972, *26*, 627–637.
20 Brodin, P.; Drakenberg, T.; Thulin, E.; Forsen, S.; Grundstroem, T. *Protein Eng.* **1989**, *2*, 353–357.
21 Sorensen, P.; Poulsen, F.M. *J. Biomol. NMR* **1992**, *2*, 99–101.
22 Patzelt, H.; Kessler, B.; Oschkinat, H.; Oesterhelt, D. *Phytochemistry* **1999**, *50*, 215–217.
23 Patzelt, H.; Ulrich, A.S.; Egbringhoff, H.; Dux, P.; Ashurst, J.; Simon, B.; Oschkinat, H.; Oesterhelt, D. *J. Biomol. NMR* **1997**, *10*, 95–106.
24 Takeda, H. *Phytochemistry* **1996**, *42*, 673–675.
25 Schweimer, K.; Marg, B.-L.; Oesterhelt, D.; Rösch, P.; Sticht, H. *J. Biomol. NMR* **2000**, *16*, 347–348.
26 Reat, V.; Patzelt, H.; Ferrand, M.; Pfister, C.; Oesterhelt, D.; Zaccai, G. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 4970–4975.
27 Celeste, S.G.; Jay, G.D.; Brenda, G.S.; Engin, S.H. *Arch. Biochem. Biophys.* **1995**, *319*, 203–210.
28 Glasoe, P.K.; Long, F.A. *J. Phys. Chem.* **1960**, *64*, 188–190.
29 Nietlispach, D.; Clowes, R.T.; Broadhurst, R.W.; Ito, Y.; Keeler, J.; Kelly, M.; Ashurst, J.; Oschkinat, H.; Domaille, P.J.; Laue, E.D. *J. Am. Chem. Soc.* **1996**, *118*, 407–415.
30 For more information see: http://www.genome.ad.jp/kegg/kegg2.html
31 Venters, R.A.; Metzler, W.J.; Spicer, L.D.; Mueller, L.; Farmer, B.T. *J. Am. Chem. Soc.* **1995**, *117*, 9592–9593.
32 Gardner, K.H.; Rosen, M.K.; Kay, L.E. *Biochemistry* **1997**, *36*, 1389–1401.
33 Crespi, H.L.; Rosenberg, R.M.; Katz, J.J. *Science* **1968**, *161*, 795–796.
34 Crespi, H.L.; Katz, J.J. *Nature* **1969**, *224*, 560–562.
35 Rosen, M.K.; Gardner, K.H.; Willis, R.C.; Parris, W.E.; Pawson, T.; Kay, L.E. *J. Mol. Biol.* **1996**, *263*, 627–636.
36 Janin, J.; Miller, S.; Chothia, C. *J. Mol. Biol.* **1988**, *204*, 155–164.
37 Kay, L.E.; Bull, T.E.; Nicholson, L.K.; Griesinger, C.; Schwalbe, H.; Bax, A.; Torchia, D.A. *J. Magn. Reson.* **1992**, *100*, 538–558.

38 Gottschalk, G. *Bacterial Metabolism*; 2 ed.; Springer Verlag: New York, **1986**.

39 Lee, A.L.; Urbauer, J.L.; Wand, A.J. *J. Biomol. NMR* **1997**, *9*, 437–440.

40 Ishima, R.; Louis, J.M.; Torchia, D.A. *J. Biomol. NMR* **2001**, *21*, 167–171.

41 Mulder, F.A.; Hon, B.; Mittermaier, A.; Dahlquist, F.W.; Kay, L.E. *J. Am. Chem. Soc.* **2002**, *124*, 1443–1451.

42 Gardner, K.H.; Kay, L.E. *J. Am. Chem. Soc.* **1997**, *119*, 7599–7600.

43 Goto, N.K.; Gardner, K.H.; Mueller, G.A.; Willis, R.C.; Kay, L. E. *J. Biomol. NMR* **1999**, *13*, 369–374.

44 Kelly, G.; Prasannan, S.; Daniell, S.; Fleming, K.; Frankel, G.; Dougan, G.; Connerton, I.; Matthews, S. *Nat. Struct. Biol.* **1999**, *6*, 313–318.

45 Mueller, G.A.; Choy, W.Y.; Yang, D.; Forman-Kay, J.D.; Venters, R.A.; Kay, L.E. *J. Mol. Biol.* **2000**, *300*, 197–212.

46 Berardi, M.J.; Sun, C.H.; Zehr, M.; Abildgaard, F.; Peng, J.; Speck, N.A.; Bushweller, J.H. *Struct. Fold. Des.* **1999**, *7*, 1247–1256.

47 Liu, X.; Wang, H.; Eberstadt, M.; Schnuchel, A.; Olejniczak, E.T.; Meadows, R.P.; Schkeryantz, J.M.; Janowick, D.A.; Harlan, J.E.; Harris, E.A.; Staunton, D.E.; Fesik, S.W. *Cell* **1998**, *95*, 269–277.

48 Aghazadeh, B.; Zhu, K.; Kubiseski, T.J.; Liu, G.A.; Pawson, T.; Zheng, Y.; Rosen, M.K. *Nat. Struct. Biol.* **1998**, *5*, 1098–1107.

49 DeRose, E.F.; Li, D.W.; Darden, T.; Harvey, S.; Perrino, F.W.; Schaaper, R.M.; London, R.E. *Biochemistry* **2002**, *41*, 94–110.

50 Vuister, G.W.; Kim, S.J.; Wu, C.; Bax, A. *J. Am. Chem. Soc.* **1994**, *116*, 9206–9210.

51 Medek, A.; Olejniczak, E.T.; Meadows, R.P.; Fesik, S. W. *J. Biomol. NMR* **2000**, *18*, 229–238.

52 Wang, H.; Janowick, D.A.; Schkeryantz, J.M.; Liu, X.H.; Fesik, S. W. *J. Am. Chem. Soc.* **1999**, *121*, 1611–1612.

53 Coughlin, P.E.; Anderson, F.E.; Oliver, E.J.; Brown, J.M.; Homans, S.W.; Pollak, S.; Lustbader, J.W. *J. Am. Chem. Soc.* **1999**, *121*, 11871–11874.

54 Giesen, A.W.; Bae, L.C.; Barrett, C.L.; Chyba, J.A.; Chaykovsky, M.M.; Cheng, M.C.; Murray, J.H.; Oliver, E.J.; Sullivan, S.M.; Brown, J.M.; Dahlquist, F.W.; Homans, S.W. *J. Biomol. NMR* **2001**, *19*, 255–260.

55 Giesen, A.W.; Bae, L.C.; Barrett, C.L.; Chyba, J.A.; Chaykovsky, M.M.; Cheng, M.C.; Murray, J.H.; Oliver, E.J.; Sullivan, S.M.; Brown, J.M.; Homans, S.W. *J. Biomol. NMR* **2002**, *22*, 21–26.

56 Hajduk, P.J.; Augeri, D.J.; Mack, J.; Mendoza, R.; Yang, J.G.; Betz, S.F.; Fesik, S.W. *J. Am. Chem. Soc.* **2000**, *122*, 7898–7904.

57 Hirata, R.; Ohsumk, Y.; Nakano, A.; Kawasaki, H.; Suzuki, K.; Anraku, Y. *J. Biol. Chem.* **1990**, *265*, 6726–6733.

58 Kane, P.M.; Yamashiro, C.T.; Wolczyk, D.F.; Neff, N.; Goebl, M.; Stevens, T.H. *Science* **1990**, *250*, 651–657.

59 Perler, F.B. *Nucleic Acids Res* **2002**, *30*, 383–384.

60 Chong, S.; Montello, G.E.; Zhang, A.; Cantor, E.J.; Liao, W.; Xu, M.Q.; Benner, J. *Nucleic Acids Res* **1998**, *26*, 5109–5115.

61 Dawson, P. E.; Kent, S. B. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.

62 Xu, R.; Ayers, B.; Cowburn, D.; Muir, T.W. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 388–393.

63 Cotton, G.J.; Ayers, B.; Xu, R.; Muir, T.W. *J. Am. Chem. Soc.* **1999**, *121*, 1100–1101.

64 Low, D.W.; Hill, M.G.; Carrasco, M.R.; Kent, S.B.; Botti, P. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6554–6559.

65 Welker, E.; Scheraga, H.A. *Biochem. Biophys. Res. Commun.* **1999**, *254*, 147–151.

66 Chong, S.; Mersha, F.B.; Comb, D.G.; Scott, M.E.; Landry, D.; Vence, L.M.; Perler, F.B.; Benner, J.; Kucera, R.B.; Hirvonen, C.A.; Pelletier, J.J.; Paulus, H.; Xu, M.Q. *Gene* **1997**, *192*, 271–281.

67 Mills, K.V.; Lew, B.M.; Jiang, S.; Paulus, H. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3543–3548.

68 Wu, H.; Xu, M. Q.; Liu, X. Q. *Biochim. Biophys. Acta* **1998**, *1387*, 422–432.

69 Southworth, M.W.; Adam, E.; Panne, D.; Byer, R.; Kautz, R.; Perler, F.B. *EMBO J.* **1998**, *17*, 918–926.

70 Yamazaki, T.; Otomo, T.; Oda, N.; Kyogoku, Y.; Uegaki, K.; Ito, N.; Ishino, Y.; Nakamura, H. *J. Am. Chem. Soc.* **1998**, *120*, 5591–5592.

71 Otomo, T.; Teruya, K.; Uegaki, K.; Yamazaki, T.; Kyogoku, Y. *J. Biomol. NMR* **1999**, *14*, 105–114.

72 Paulus, H. *Annu. Rev. Biochem.* **2000**, *69*, 447–496.

73 Evans, T.C., Jr.; Martin, D.; Kolly, R.; Panne, D.; Sun, L.; Ghosh, I.; Chen, L.; Benner, J.; Liu, X. Q.; Xu, M.Q. *J. Biol. Chem.* **2000**, *275*, 9091–9094.

74 Iwai, H.; Plückthun, A. *FEBS Lett.* **1999**, *459*, 166–172.

75 Scott, C.P.; Abel-Santos, E.; Wall, M.; Wahnon, D.C.; Benkovic, S.J. *Proc. Natl. Acad. Sc. USA* **1999**, *96*, 13638–13643.

76 Camarero, J.A.; Muir, T.W. *J. Am. Chem. Soc.* **1999**, *121*, 5597–5598.

77 Flory, P.J. *J. Am. Chem. Soc.* **1956**, *78*, 5222–5235.

78 Trabi, M.; Craik, D.J. *Trends Biochem. Sci.* **2002**, *27*, 132–138.

79 Evans, T.C., Jr.; Benner, J.; Xu, M.Q. *J. Biol. Chem.* **1999**, *274*, 3923–3926.

80 Iwai, H.; Lingel, A.; Plückthun, A. *J. Biol. Chem.* **2001**, *276*, 16548–16554.

81 Williams, N.K.; Prosselkov, P.; Liepinsh, E.; Line, I.; Shapiro, A.; Littler, D.R.; Curmi, P.M.; Otting, G.; Dixon, N.E. *J. Biol. Chem.* **2002**, *277*, 7790–7798.

82 Siebold, C.; Flukiger, K.; Beutler, R.; Erni, B. *FEBS Lett.* **2001**, *504*, 104–111.

83 Goldenberg, D.P.; Creighton, T.E. *J. Mol. Biol.* **1983**, *165*, 407–413.

84 Camarero, J.A.; Fushman, D.; Sato, S.; Giriat, I.; Cowburn, D.; Raleigh, D.P.; Muir, T. W. *J. Mol. Biol.* **2001**, *308*, 1045–1062.

85 Woese, C.R.; Fox, G.E. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5088–5090.

86 Murata, K.; Mitsuoka, K.; Hirai, T.; Walz, T.; Agre, P.; Heymann, J.B.; Engel, A.; Fujiyoshi, Y. *Nature* **2000**, *407*, 599–605.

87 Luecke, H.; Schobert, B.; Richter, H.T.; Cartailler, J.P.; Lanyi, J.K. *J. Mol. Biol.* **1999**, *291*, 899–911.

88 Turner, G.J.; Reusch, R.; Winter-Vann, A.M.; Martinez, L.; Betlach, M.C. *Protein Expr. Purif.* **1999**, *17*, 312–323.

89 Winter-Vann, A.M.; Martinez, L.; Bartus, C.; Levay, A.; Turner, G.J. In *Perspectives on Solid State NMR in Biology*; Kühne, S. R., de Groot, H. J. M., Eds.; Kluwer: Netherlands, **2001**; pp 141–159.

90 Hinnen, A.; Buxton, F.; Chaudhuri, B.; Heim, J.; Hottiger, T.; Meyhack, B.; Pohlig, G. In *Gene expression in recombinant Microorganisms*; Smith, A., Ed.; Marcel Dekker: New York, **1994**; pp 121–193.

91 Valenzuela, P.; Medina, A.; Rutter, W.J.; Ammerer, G.; Hall, B.D. *Nature* **1982**, *298*, 347–350.

92 Cabezon, T.; De Wilde, M.; Herion, P.; Loriau, R.; Bollen, A. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 6594–6598.

93 Emini, E.A.; Ellis, R.W.; Miller, W.J.; McAleer, W.J.; Scolnick, E.M.; Gerety, R.J. *J. Infect.* **1986**, *13 Suppl. A*, 3–9.

94 Schultz, L.D.; Tanner, J.; Hofmann, K.J.; Emini, E.A.; Condra, J.H.; Jones, R.E.; Kieff, E.; Ellis, R.W. *Gene* **1987**, *54*, 113–123.

95 Hallewell, R.A.; Mills, R.; Tekamp-Olson, P.; Blacher, R.; Rosenberg, S.; Otting, F.; Masiarz, F.R.; Scandella, C.J. *Biotechnol.* **1987**, *5*, 363–366.

96 David, N.E.; Gee, M.; Andersen, B.; Naider, F.; Thorner, J.; Stevens, R.C. *J. Biol. Chem.* **1997**, *272*, 15553–15561.

97 Andersen, B.; Stevens, R.C. *Protein Expr. Purif.* **1998**, *13*, 111–119.

98 Giga-Hama, Y.; Kumagai, H. *Biotechnol. Appl. Biochem.* **1999**, *30*, 235–244.

99 Grallert, B.; Nurse, P.; Patterson, T.E. *Mol. Gen. Genet.* **1993**, *238*, 26–32.

100 Maundrell, K.; Hutchison, A.; Shall, S. *EMBO J.* **1988**, *7*, 2203–2209.

101 Toyama, R.; Okayama, H. *FEBS Lett.* **1990**, *268*, 217–221.

102 Xu, H.P.; White, M.; Marcus, S.; Wigler, M. *Mol. Cell Biol.* **1994**, *14*, 50–58.

103 Hildebrandt, V. In *Foreign Gene Expression in Fission Yeast Schizosaccharomyces pombe*; Giga-Hama, Y., Kumagai, H., Eds.; Springer Verlag: Berlin, **1997**; pp 97–110.

104 Giga-Hama, Y.; Tohda, H.; Okada, H.; Owada, M.K.; Okayama, H.; Kumagai, H. *Biotechnology* **1994**, *12*, 400–404.

105 Yasumori, T. In *Foreign Gene expression in Fission Yeast Schizosaccharomyces pombe*; Giga-Hama, Y., Kumagai, H., Eds.; Springer Verlag: Berlin, **1997**; pp 111–121.

106 Sander, P.; Grunewald, S.; Reilander, H.; Michel, H. *FEBS Lett.* **1994**, *344*, 41–46.

107 Kjeldsen, T.; Pettersson, A.F.; Hach, M. *Biotechnol. Appl. Biochem.* **1999**, *29*, 79–86.

108 Cereghino, J.L.; Cregg, J.M. *FEMS Microbiol. Rev.* **2000**, *24*, 45–66.

109 Weiss, H.M.; Haase, W.; Michel, H.; Reilander, H. *Biochem. J.* **1998**, *330*, 1137–1147.

110 Luque, T.; O'Reilly, D. R. *Mol. Biotechnol.* **1999**, *13*, 153–163.

111 Possee, R. D. *Curr. Opin. Biotechnol.* **1997**, *8*, 569–572.

112 Ogonah, O. W.; Freedman, R. B.; Jenkins, N.; Patel, K.; Rooney, B. C. *Nat. Biotechnol.* **1996**, *14*, 197–202.

113 Condreay, J. P.; Witherspoon, S. M.; Clay, W. C.; Kost, T. A. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 127–132.

114 Payne, L. A.; Fornwald, J. A.; Kane, J. F.; McNulty, D. E.; Trill, J. J.; Ramos, L. In *Animal Cell Technology: From Target to Market*; Lindner-Olsson, E., Chatissavidou, N., Lillau, E., Eds.; Kluwer: Netherlands, **2001**; pp 94–100.

115 Mazina, K. E.; Strader, C. D.; Fong, T. M. *J. Recept. Res.* **1994**, *14*, 63–73.

116 Guan, X. M.; Kobilka, T. S.; Kobilka, B. K. *J. Biol. Chem.* **1992**, *267*, 21995–21998.

117 Gu, H.; Wall, S. C.; Rudnick, G. *J. Biol. Chem.* **1994**, *269*, 7124–7130.

118 Girard, P.; Derouazi, M.; Baumgartner, G.; Bourgeois, M.; Jordan, M.; Wurm, F. M. In *Animal Cell Technology: From Target to Market*; Lindner-Olsson, E., Chatzissavidou, N., Lillau, E., Eds.; Kluwer Academic Publishers: Netherlands, **2001**; pp 37–42.

119 Fux, C.; Moser, S.; Schlatter, S.; Rimann, M.; Bailey, J. E.; Fussenegger, M. *Nucleic Acids Res.* **2001**, *29*, E19.

120 Boorsma, M.; Nieba, L.; Koller, D.; Bachmann, M. F.; Bailey, J. E.; Renner, W. A. *Nat. Biotechnol.* **2000**, *18*, 429–432.

121 Tate, C. G. *FEBS Lett.* **2001**, *504*, 94–98.

122 Graham, F. L. *Immunol. Today* **2000**, *21*, 426–428.

123 Panicali, D.; Paoletti, E. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 4927–4931.

124 Ward, G. A.; Stover, C. K.; Moss, B.; Fuerst, T. R. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6773–6777.

125 Liljestrom, P.; Garoff, H. *Biotechnol.* **1991**, *9*, 1356–1361.

126 Xiong, C.; Levis, R.; Shen, P.; Schlesinger, S.; Rice, C. M.; Huang, H. V. *Science* **1989**, *243*, 1188–1191.

127 Lundstrom, K. *J. Recept. Signal. Transduct. Res.* **1999**, *19*, 673–686.

128 Lundstrom, K. In *Perspectives on Solid-State NMR in Biology*; Kühne, S. R., de Groot, J. J. M., Eds.; Kluwer: The Netherlands, **2001**; pp 131–139.

129 Hovius, R.; Tairi, A.-P.; Blasey, H.; Bernard, A. R.; Lundstrom, K.; Vogel, H. *J. Neurochem.* **1998**, *70*, 824–834.

130 Lundstrom, K.; Schweitzer, C.; Rotmann, D.; Hermann, D.; Schneider, E. M.; Ehrengruber, M. U. *FEBS Lett.* **2001**, *504*, 99–103.

131 Lundstrom, K.; Rotmann, D.; Herrmann, D.; Schlaeger, E.-J. *Cytotechnology* **2001**, *35*, 213–221.

132 Fernandez, C.; Hilty, C.; Bonjour, S.; Adeishvili, K.; Pervushin, K.; Wüthrich, K. *FEBS Lett.* **2001**, *504*, 173–178.

133 Bellizzi, J. J.; Widom, J.; Kemp, C. W.; Clardy, J. *Structure* **1999**, *7*, R263–267.

134 Bushnell, D. A.; Cramer, P.; Kornberg, R. D. *Structure* **2001**, *9*, R11–R14.

135 Wood, M. J.; Komives, E. A. *J. Biomol. NMR* **1999**, *13*, 149–159.

136 Morgan, W. D.; Birdsall, B.; Frenkiel, T. A.; Gradwell, M. G.; Burghaus, P. A.; Syed, S. E. H.; Uthaipibull, C.; Holder, A. A.; Feeney, J. *J. Molec. Biol.* **1999**, *289*, 113–122.

137 Palczewska, M.; Groves, P.; Kuznicki, J. *Protein Expr. Purif.* **1999**, *17*, 465–476.

138 Laroche, Y.; Storme, V.; Demeutter, J.; Messens, J.; Lauwereys, M. *Bio-Technology* **1994**, *12*, 1119–1124.

139 Morgan, W. D.; Kragt, A.; Feeney, J. *J. Biomol. NMR* **2000**, *17*, 337–347.

140 Helmle, M.; Patzelt, H.; Ockenfels, A.; Gartner, W.; Oesterhelt, D.; Bechinger, B. *Biochemistry* **2000**, *39*, 10066–10071.

141 Wyss, D. F.; Withka, J. M.; Knoppers, M. H.; Sterne, K. A.; Recny, M. A.; Wagner, G. *Biochemistry* **1993**, *32*, 10995–11006.

142 Hansen, A. P.; Petros, A. M.; Mazar, A. P.; Pederson, T. M.; Rueter, A.; Fesik, S. W. *Biochemistry* **1992**, *31*, 12713–12718.

143 Lustbader, J. W.; Birken, S.; Pollak, S.; Pound, A.; Chait, B. T.; Mirza, U. A.; Ramnarain, S.; Canfield, R. E.; Brown, J. M. *J. Biomol. NMR* **1996**, *7*, 295–304.

144 See http://www.mepnet.org for further information

145 Bailey, W. J.; Vanti, W. B.; George, S. R.; Levins, R.; Swarminathan, S.; Bonini, J. A.; Smith, K. E.; Weinshank, R. L.; O'Dowd, B. F. *Expert Opin. Ther. Patents* **2001**, *11*, 1861–1887.

146 Arora, A.; Abildgaard, F.; Bushweller, J. H.; Tamm, L. K. *Nat. Struct. Biol.* **2001**, *8*, 334–338.

147 Fernandez, C.; Adeishvili, K.; Wüthrich, K. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2358–2363.

148 Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C.A.; Motoshima, H.; Fox, B.A.; Le Trong, I.; Teller, D.C.; Okada, T.; Stenkamp, R.E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739–745.

149 Spirin, A.S.; Baranov, V.I.; Ryabova, L.A.; Ovodov, S.Y.; Alakhov, Y.B. *Science* **1988**, *242*, 1162–1164.

150 Kim, D.M.; Kigawa, T.; Choi, C.Y.; Yokoyama, S. *Eur. J. Biochem.* **1996**, *239*, 881–886.

151 Kigawa, T.; Yabuki, T.; Yoshida, Y.; Tsutsui, M.; Ito, Y.; Shibata, T.; Yokoyama, S. *FEBS Lett.* **1999**, *442*, 15–19.

152 Madin, K.; Sawasaki, T.; Ogasawara, T.; Endo, Y. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 559–564.

153 Kim, D.M.; Swartz, J.R. *Biotechnol. Prog.* **2000**, *16*, 385–390.

154 Kim, D.M.; Swartz, J.R. *Biotechnol. Bioeng.* **2001**, *74*, 309–316.

155 Kigawa, T.; Muto, Y.; Yokoyama, S. *J. Biomol. NMR* **1995**, *6*, 129–134.

156 Yabuki, T.; Kigawa, T.; Dohmae, N.; Takio, K.; Terada, T.; Ito, Y.; Laue, E. D.; Cooper, J.A.; Kainosho, M.; Yokoyama, S. *J. Biomol. NMR* **1998**, *11*, 295–306.

157 See www.proteinexpression.com for more information.

158 Fernholz, E.; Besir, H.; Kühlewein, A.; Mayr, D.; Schmitt, R.; Schwaiger, M. unpublished results **2002**.

159 Spirin, A.S. *Cell-Free Translation Systems*; Springer Verlag: Berlin, **2002**.

160 Fernholz, E.; Zaiss, K.; Besir, H.; Mutter, W. In *Cell–Free Translation Systems*; Spirin, A. S., Ed.; Springer Verlag: Berlin, **2002**; pp 175–179.

161 Tarui, H.; Imanishi, S.; Hara, T. *J. Biosci. Bioeng.* **2000**, *90*, 508–514.

162 LeMaster, D.M.; Kushlan, D.M. *J. Am. Chem. Soc.* **1996**, *118*, 9255–9264.

163 Ishima, R.; Louis, J.M.; Torchia, D.A. *J. Am. Chem. Soc.* **1999**, *121*, 11589–11590.

164 Toyoshima, C.; Nakasako, M.; Nomura, H.; Ogawa, H. *Nature* **2000**, *405*, 647–655.

165 Hunte, C.; Koepke, J.; Lange, C.; Rossmanith, T.; Michel, H. *Struct. Fold. Des.* **2000**, *8*, 669–684.

166 Yoshikawa, S.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yamashita, E.; Inoue, N.; Yao, M.; Fei, M. J.; Libeu, C. P.; Mizushima, T.; Yamaguchi, H.; Tomizaki, T.; Tsukihara, T. *Science* **1998**, *280*, 1723–1729.

167 Stock, D.; Leslie, A. G.; Walker, J. E. *Science* **1999**, *286*, 1700–1705.

168 Lancaster, C.R.; Kroger, A.; Auer, M.; Michel, H. *Nature* **1999**, *402*, 377–385.

169 Unger, V.M.; Kumar, N.M.; Gilula, N. B.; Yeager, M. *Science* **1999**, *283*, 1176–1180.

170 Fu, D.; Libson, A.; Miercke, L.J.; Weitzman, C.; Nollert, P.; Krucinski, J.; Stroud, R. M. *Science* **2000**, *290*, 481–486.

171 Kolbe, M.; Besir, H.; Essen, L.O.; Oesterhelt, D. *Science* **2000**, *288*, 1390–1396.

172 Auer, M.; Scarborough, G.A.; Kühlbrandt, W. *Nature* **1998**, *392*, 840–843.

173 Chang, G.; Spencer, R.H.; Lee, A.T.; Barclay, M.T.; Rees, D. C. *Science* **1998**, *282*, 2220–2226.

174 Williams, K.A. *Nature* **2000**, *403*, 112–115.

175 Miyazawa, A.; Fujiyoshi, Y.; Stowell, M.; Unwin, N. *J. Mol. Biol.* **1999**, *288*, 765–786.

176 Lancaster, C.R.; Michel, H. *J. Mol. Biol.* **1999**, *286*, 883–898.

177 Zouni, A.; Witt, H.T.; Kern, J.; Fromme, P.; Krauss, N.; Saenger, W.; Orth, P. *Nature* **2001**, *409*, 739–743.

178 Meyer, J.E.; Hofnung, M.; Schulz, G.E. *J. Mol. Biol.* **1997**, *66*, 761–765.

179 Doyle, D.A.; Morais Cabral, J.; Pfuetzner, R.A.; Kuo, A.; Gulbis, J.M.; Cohen, S.L.; Chait, B. T.; MacKinnon, R. *Science* **1998**, *280*, 69–77.