FERMENTATION MONITORING, DESIGN AND OPTIMIZATION

JENS NIELSEN
Technical University of Denmark
Lyngby, Denmark

KEY WORDS
Fermentation
High-performance bioreactors
Metabolic control analysis
Metabolic engineering
Metabolic flux analysis
On-line monitoring
Quantitative physiology

OUTLINE
Introduction
Criteria for Design and Optimization
Strain Construction and Strain Improvement
Modern Experimental Techniques
High-Performance Bioreactors
Quantitative Physiological Studies
Bibliography

INTRODUCTION
The term fermentation is derived from Latin fermentum to ferment, and it has been used to describe the metabolism of sugars by microorganisms since ancient times. Thus, fermentation of fruits is so old that ancient Greeks attributed its discovery to one of their gods, Dionysos. Among the classical fermentation processes are beer brewing, which has documented from as early as 3000 B.C. in Babylonia, soya sauce production in Japan and China, and fermented milk beverages in the Balkans and in Central Asia (1). Before World War II, however, fermentation processes mainly found their application in the food area, and it was with the introduction of the penicillin production that large-scale fermentation was first used in the production of pharmaceuticals. Today fermentation processes are used in the production of many different products and these processes can be divided into nine categories, according to the product (Table 1).

Development of fermentation processes can roughly be divided into four phases (see Fig. 1). First the product is identified. In the case of a pharmaceutical this may be a result of random screening for different therapeutic effects by microbial metabolites (e.g., by high throughput screening of secondary metabolites from Actinomycetes) or it may be the result of a targeted identification of a novel product (e.g., a peptide hormone with known function). Outside the pharmaceutical sector the product may also be chosen after a random screening procedure (e.g., screening for a novel enzyme to be used in detergents) or it may be chosen in a more rational fashion. With the rapid progress in the different genomic sequencing programs (the yeast genome was completely sequenced in 1996 and the genome of several bacteria have been completely sequenced), there is much focus on identification of novel products, and today topics such as bioinformatics and gene function are front-line research topics (2).

The next step in the development phase is to choose a strain. In the past this choice was normally obvious after the product had been identified; for example, Penicillium chrysogenum was chosen for penicillin production because it was the first organism identified to produce penicillin.

Table 1. Categories and Typical Application of Fermentation Processes

<table>
<thead>
<tr>
<th>Category</th>
<th>Typical processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production of microbial cells</td>
<td>Baker's yeast, single cell protein, lactic acid bacteria</td>
</tr>
<tr>
<td>Production of primary metabolites</td>
<td>Ethanol (both in beer and wine production and for technical use), lactic acid, citric acid, acetic acid, amino acids, vitamins</td>
</tr>
<tr>
<td>Production of secondary metabolites</td>
<td>Antibiotics (penicillins, cephalosporins, davalanic acid, tetracycline), skikonin</td>
</tr>
<tr>
<td>Production of microbial enzymes</td>
<td>Lipases, proteases, amylases, glucose isomerase</td>
</tr>
<tr>
<td>Production of pharmaceutical proteins</td>
<td>tPA, human insulin, erythropoietin, growth hormone, hepatitis B vaccine, monoclonal antibodies</td>
</tr>
<tr>
<td>Production of polysaccharides</td>
<td>Xanthan gum, dextran</td>
</tr>
<tr>
<td>Production of DNA</td>
<td>Genes for vaccines or gene therapy</td>
</tr>
<tr>
<td>Biotransformations</td>
<td>Steroids, L-sorbose (from L-sorbitol)</td>
</tr>
<tr>
<td>Production of tissue</td>
<td>Bone marrow, skin</td>
</tr>
</tbody>
</table>

Figure 1. Different phases in the development of a fermentation process.
With the introduction of recombinant DNA (rDNA) technology, it is now possible to choose almost any host for the production. Thus, a strain of Escherichia coli has been constructed that can produce ethanol at a high yield (3), and a recombinant strain of P. chrysogenum can now be used to produce 7-ADCA (a precursor used for synthesis of cephalosporins) directly by fermentation (4). The choice of strain does, however, often depend much on tradition within the company, and most of the fermentation industries have a set of favorite organisms that are used in the production of many different products. Thus the major players on the industrial enzyme market (Novo Nordisk, Genencor, and Gist-brocades) have chosen a few organisms (typically one to two strains of filamentous fungi and one to two species of bacteria) as production vehicles for a wide range of enzymes. With the production of a heterologous protein, that is, expression of a foreign gene in a given organism, it is also necessary to consider many other aspects (e.g., whether the protein is correctly folded and glycosylated) Table 2 gives an overview of the advantages and disadvantages of different cellular systems for the production of recombinant proteins.

After the strain has been constructed, one of the first aims is to produce sufficient materials for further research, and this is typically done in pilot-plant facilities. For a pharmaceutical compound sufficient material must be produced for clinical trials. For other products it may be necessary to carry out tests of the product and examine any possible toxic effects. In parallel to the continuing research in the application of the product, the production process is also designed. The final steps in development are product approval by the proper authorities and construction of the production facility (which in some cases may be through retrofitting of an existing plant). After the process has been developed, there is continuous optimization of the process, and in some cases better strains are introduced. However, for pharmaceutical products new strains are, only introduced when there is a significant improvement in the process because it is costly to obtain approval from the authorities for redesign of the process.

### Table 2. Advantages and Disadvantages of Different Hosts for Production of Recombinant

<table>
<thead>
<tr>
<th>Host</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteria (E. coli)</td>
<td>Wide choice of cloning vectors</td>
<td>Posttranslational modifications lacking</td>
</tr>
<tr>
<td></td>
<td>Gene expression easy to control</td>
<td>High endotoxin content</td>
</tr>
<tr>
<td></td>
<td>Large yields possible</td>
<td>Protein aggregation (inclusion bodies)</td>
</tr>
<tr>
<td></td>
<td>Good protein secretion</td>
<td></td>
</tr>
<tr>
<td>Yeast (S. cerevisiae)</td>
<td>Generally regarded as safe (GRAS)</td>
<td>Less cloning vectors available</td>
</tr>
<tr>
<td></td>
<td>No pathogens for humans</td>
<td>Glycosylation not identical to mammalian glycosylation</td>
</tr>
<tr>
<td></td>
<td>Large-scale production established</td>
<td>Genetics less understood</td>
</tr>
<tr>
<td></td>
<td>Some posttranslational modifications possible</td>
<td></td>
</tr>
<tr>
<td>Filamentous fungi</td>
<td>Experience with large-scale production</td>
<td>High level of heterologous protein expression has not been achieved</td>
</tr>
<tr>
<td></td>
<td>Source of many industrial enzymes</td>
<td>Genetics not well characterized</td>
</tr>
<tr>
<td></td>
<td>Excellent protein secretors</td>
<td></td>
</tr>
<tr>
<td>Mammalian cells</td>
<td>Same biological activity as natural protein</td>
<td>Cells difficult to grow in bioreactors</td>
</tr>
<tr>
<td></td>
<td>Expression vectors available</td>
<td>Expensive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Slow growth</td>
</tr>
<tr>
<td>Cultured insect cells</td>
<td>High level of gene expression possible</td>
<td>Low productivity</td>
</tr>
<tr>
<td></td>
<td>Posttranslational modification possible</td>
<td></td>
</tr>
</tbody>
</table>

### CRITERIA FOR DESIGN AND OPTIMIZATION

The criteria used for design and optimization of fermentation processes depends on the product. Thus, the criteria used for a high-volume-low-value-added product are normally completely different from the criteria used for a low-volume-high-value-added product. For products belonging to the first category (which includes most whole cell products, most primary metabolites, many secondary metabolites, most industrial enzymes, and most polysaccharides) the three most important design parameters are

- Yield of product on the substrate
- Productivity
- Final titer

A high yield of product on the substrate is for these processes very important, since this gives a good utilization of the raw materials, which often accounts for a significant part of the total costs. Thus, in penicillin production the costs of glucose alone may account for up to 15% of the total production costs (5). Productivity is important because it ensures an efficient utilization of the production medium (e.g., purification of the product). Thus, if the product is present in a very low concentration at the end of the fermentation, it may be very expensive to perform purification of the product.

In the production of novel pharmaceuticals, which typically belong to the category of low-volume-high-value-added products, the three previously mentioned design parameters are normally not very important. For these processes quickness to market is generally much more important, and change of the process after implementation is often complicated due to the required U.S. Food and Drug Administration (FDA) approval. In the initial design phase...
it is, however, still important to keep these three design criteria in mind. The requirement for a high final titer is especially important since the cost for purification (or downstream processing) often accounts for more than 90% of the total production costs.

STRAIN CONSTRUCTION AND STRAIN IMPROVEMENT

Today high-performance production strains to be used in fermentation processes are often constructed using rDNA technology. However, the concept of metabolic pathway manipulation is by no means new. Thus, for decades better strains of Saccharomyces to be used for beer fermentation have been obtained through classical breeding and crossing of different strains, and in the production of penicillin productivity has increased by more than 500 times through repeated rounds of mutation and selection of new strains of P. chrysogenum. With the introduction of rDNA technology it has become possible to apply a more rational approach to strain improvement—namely by the introduction of targeted genetic changes resulting in strains with a phenotype that gives a better process. This rational approach has been named cellular and metabolic engineering, and different definitions of the approach have been given (6). The term metabolic engineering was first introduced by Bailey (7), who defined it as “improvement of cellular activities by manipulation of enzymatic transport and regulatory functions of the cell with the use of recombinant DNA technology.” His definition of metabolic engineering includes the following:

1. Inserting new pathways in microorganisms with the aim of producing novel metabolites (e.g., production of polyketides by Streptomyces [8,9]), with the aim of degrading toxic compounds (e.g., in bioremediation), or with the aim of constructing a novel biotransformation system
2. Production of heterologous peptides, such as production of human insulin, erythropoietin, and tPA, or industrial enzymes, such as lipases and cellulases
3. Improvement of pathway fluxes leading to higher yields of metabolites (e.g., increasing the flux toward antibiotics or primary metabolites) or yields of biomass (e.g., increasing the cell mass yield in baker’s yeast production using industrial media containing mixed sugars [10])

What characterizes metabolic engineering is the rational approach to performing genetic changes, and as with all other fields of engineering it consists of two steps: analysis and synthesis (11). As a consequence of the difficulties in performing detailed analysis of cellular metabolism there has mainly been focus on synthesis in the past, such as expression of new genes in various host cells, amplification of endogenous enzymes, and deletion of genes or modulation of enzyme activities. With modern experimental techniques it has, become possible, however, to perform detailed analysis of cellular function through both in vivo and in vitro measurements, and in the following some of these techniques are described.

MODERN EXPERIMENTAL TECHNIQUES

Whereas the development in rDNA technology has enabled the design of novel high-performance strains that are tailor-made to specific needs, the revolution in computer technology and in analytical chemistry has allowed a much more fundamental characterization of cellular function than previously possible. Thus, flow cytometry allows quantification of population dynamics, while advances in microscopy and image analysis have allowed studies of cell growth and function at the single cell level (12). In studies of fermentation processes it is important to study the function of whole cultures, and here the development of high-performance bioreactors has been an important contribution. When microbial or cell cultures are grown in these bioreactors, it is ensured that all the cells experience the same conditions, and it is therefore possible to perform quantitative physiological studies, such as study of growth and production kinetics, study of gene expression under well-controlled conditions, quantification of metabolic fluxes, and study of the control of flux through the different cellular pathways.

High-Performance Bioreactors

High-performance bioreactors are characterized as being practically ideal bioreactors with a very low mixing time and a very high gas–liquid mass transfer (13,14). In these bioreactors the cells are subjected to an unchanging environment when they are circulated throughout the liquid medium, and the response to imposed variations in the environment is therefore a consequence of the cellular behavior only. High-performance bioreactors are equipped with a large number of in situ sensors, and to control operating variables a flexible direct digital control (DDC) is used, rather than classical single-purpose controllers. This allows precise control of many operating variables, including temperature, pH, dilution rate, and dissolved oxygen concentration, which therefore can be taken to be culture parameters (Table 3) (15). Many commercially available laboratory bioreactors (volumes less than 10 L) normally fulfill the requirements for being high-performance bioreactors. Several bioreactor companies offer systems with flexible DDCs that enable precise control of the culture variables, and when these bioreactors are designed with

Table 3. In Situ Sensors Used for Measuring Culture Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sensor</th>
<th>Range</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Pt-100</td>
<td>0–150 °C</td>
<td>0.1 °C</td>
</tr>
<tr>
<td>Pressure</td>
<td>Piezoresistor</td>
<td>0–2 bar</td>
<td>20 mbar</td>
</tr>
<tr>
<td>Gas flow</td>
<td>Thermal mass flow meter</td>
<td>0–20 L min⁻¹</td>
<td>0 mL min⁻¹</td>
</tr>
<tr>
<td>pH</td>
<td>pH electrode</td>
<td>2–12</td>
<td>0.02</td>
</tr>
<tr>
<td>pHO₂</td>
<td>Polarographic Clark electrode</td>
<td>0–400 mbar</td>
<td>2 mbar</td>
</tr>
<tr>
<td>pCO₂</td>
<td>Membrane-covered pH electrode</td>
<td>0–100 mbar</td>
<td>2 mbar</td>
</tr>
</tbody>
</table>
the proper stirrer and sparger it is possible to ensure a very low mixing time (less than 1 s) and a high gas–liquid mass transfer ($k\alpha$ values above 0.1 s$^{-1}$). In addition to sensors for the culture parameters, modern high-performance bioreactors are normally equipped with in situ sensors and on-line analyzers that allow frequent monitoring of key culture variables. Figure 2 illustrates a typical high-performance bioreactor system. The bioreactor is equipped with on-line flow injection analyzers for monitoring of important culture variables, exhaust gas analysis, and an automatic sampling system. Hereby the most important culture variables can be measured at a high frequency (30-1,000 h$^{-1}$), whereas other culture parameters can be measured in the automatically withdrawn samples at a lower frequency (1-3 h$^{-1}$). The bioreactor system can be operated as a constant mass, continuous culture with measurement of the feed flow. It is therefore possible to vary the dilution rate (with a precision of 0.005 h$^{-1}$), and through proper design of the feed medium it is possible to operate the bioreactor such that there is a single limiting substrate component, such as glucose or ammonia. Finally, the bioreactor is equipped with gas blending, which allows variations in the composition of the inlet gas to the bioreactor, so the influence of, for example, the oxygen concentration on the growth and production kinetics can be studied at the same specific growth rate (16).

With advanced computer control of culture parameters it is possible to perform very reproducible experiments, which is of paramount importance for a fundamental understanding of the underlying cellular reactions. This is illustrated in Figure 3, where the carbon dioxide evolution rate is shown for two different batch cultivations of the filamentous fungi $P$. chrysogenum. With remarkable reproducibility the CER signal can obviously be used to map the different phases of the batch fermentation, which were carried out with sucrose as the carbon source. In the first phase sucrose is hydrolyzed to glucose and fructose and glucose is metabolized. When glucose is exhausted after 45 h there is a distinct shift in the metabolism, and when fructose is exhausted after 60 h there is another shift in the metabolism (i.e., the CER decreases rapidly). Even when both sugars are exhausted there is still some cellular activity due to metabolism of gluconic acid, which was formed in the initial phase during growth on glucose (17).

**In situ Sensors.** An in situ sensor can be inserted directly into the bioreactor and can be sterilized in place. Table 3 gives the characteristics of in situ sensors used to monitor culture parameters in a typical high-performance bioreactor. Most of these sensors are extremely reliable and are used routinely not only in research laboratories but also in industrial plants.

![Figure 2. Typical high-performance bioreactor system.](image-url)
For monitoring of culture variables there are, however, only a few commercially available in situ sensors available, and besides a few sensors based on enzymatic analysis of medium components (e.g., for glucose analysis [18]), these sensors measure biomass using different measurement principles (see Table 4) (19). There are several commercially available sensors for in situ monitoring of the optical density, which is linearly correlated with the biomass concentration. These sensors have a wide linear range, but there is interference from gas bubbles and solids in the medium. In some sensors special designs are used to reduce this interference such as by passing the sample into an internal measurement chamber (19). Sensors measuring the culture fluorescence (normally the NAD(P)H-related fluorescence) were first introduced as biomass sensors, but despite high sensitivity and correlation of the culture fluorescence with the total cellular activity, the signal from these sensors is difficult to interpret, and there is generally a poor correlation with the biomass concentration (17,20). Measurement of dielectric properties (or capacitance) is one of the later contributions to measurement of biomass, and it has the advantage of a wide measurement range. However, the signal depends on the cellular activity, which although it is an interesting variable, may not necessarily correlate with the biomass concentration. Even though none of the commercially available in situ biomass sensors stands out as the best biomass sensor, several of the sensors are very useful additions to the sensors normally applied in high-performance bioreactors.

**On-Line Analyzers.** With the lack of reliable in situ sensors for measurement of important culture variables, including glucose and other medium components, on-line analysis is frequently used. Here a sample is automatically withdrawn and analyzed. For this purpose flow injection analysis (FIA) is well suited due to its high speed, good precision, and good reliability (21). The major drawback of FIA is that it is a single component analysis, and even though a novel technique termed sequential injection analysis (SIA) offers the possibility to measure several components using the same hardware (22), other analytical systems such as mass spectrometry (MS), high-performance liquid chromatography (HPLC), and gas chromatography (GC) may sometimes be preferred. These analytical systems offer the ability to measure many components in a single run, but, in general, it is difficult to match the high measurement frequency and reliability of FIA.

A prerequisite for on-line monitoring is automatic sampling, and the sampling system normally serves as the sterile barrier between the bioreactor and the analytical system. Automatic sampling can be performed either by withdrawing a sample directly or by using a membrane module. In direct sample withdrawal, a nonfiltered sample is pumped directly from the bioreactor through a syringe or catheter. Direct sample withdrawal is a very simple solution that allows for on-line measurement of the biomass concentration using FIA (23) and measurement of intra-cellular components (24). However, a construction with a syringe is not sufficiently robust to be used in an industrial environment. For sampling via a membrane module there are several commercially available designs (25), which can be divided into two groups:

1. **Membrane modules placed in a recycle loop connected to the bioreactor.** The advantage of these modules is that it is possible to change the membrane while the process is running. However, the continuous pumping of the medium through the membrane loop may have a significant effect on the cellular behavior (especially for shear-sensitive organisms).

2. **In situ membrane modules, which are inserted into the bioreactor and sterilized in place.** The advantage of these modules is that the modules do not influence the bioreactor operation and the withdrawn sample is a good representation of the medium in the bioreactor. However, to ensure mechanical breakage of the membrane it is necessary to apply rather thick membranes, which influences the response time (26).

With a reliable membrane module for automatic sampling it is possible to perform routine on-line monitoring of important medium components, and if FIA is used it is especially possible to obtain high frequency measurements, which allows precise quantification of the fluxes in and out
of the cells (e.g., the substrate uptake rates and the product formation rates). These fluxes form the basis for quantitative physiological research, as discussed later, and high-performance bioreactors equipped with on-line analyzers for the most important culture variables are therefore indispensable tools in such research.

**Measurement of Intracellular Variables.** For quantitative physiological studies, the measurement of intracellular variables is extremely important. These variables should ideally be performed in vivo, while cellular function is maintained, but presently there are no techniques that allow for precise in vivo measurements of intracellular variables; therefore they are taken in vitro. To ensure that the in vitro measurements are good representatives of the in vivo conditions, it is critical to have techniques that allow for rapid sampling and quenching of the cellular activity. For the measurement of intracellular metabolites and cofactors, such as glycolytic intermediates, ATP, and NADH, which have a turnover time of a few seconds, it is extremely important to have an almost instantaneous quenching of the cellular metabolism. Using a specially designed sampling module and rapid cooling of the sample in precooled test tubes containing small glass beads, Theobald et al. (27) were able to measure the intracellular level of many metabolites of the EMP pathway and the cofactors ATP, ADP, and AMP in *S. cerevisiae*. Furthermore, using a freeze-quench method, where the sample is mixed with a reagent and then quenched by pumping it into a reservoir of liquid methanol at −40 °C, de Koning and van Dam studied the uptake kinetics of glucose by *S. cerevisiae* at the scale of 5 ms to 1 s (28). Using these sampling techniques, together with HPLC and enzymatic assays, it is possible to obtain a fundamental understanding of the metabolic levels inside the cells at different growth conditions, which, together with knowledge of the metabolic fluxes (discussed later), will give the ultimate physiological characterization.

**Quantification of Morphology.** For filamentous microorganisms there is in many cases a relation between morphology and product formation. In the past, detailed studies of the growth mechanisms of these organisms have been hampered by the lack of good experimental techniques. However with the rapid development in image analysis it has become possible to study the growth of these organisms in much detail (29). Yet, because of the many different processes that influence morphology, it is important to combine experimental work and mathematical modeling to extract information about the underlying mechanisms. When the morphology is measured for a given number of hyphal elements in a sample taken from a fermentation with filamentous microorganisms, there will be a certain distribution of properties, governed by the underlying mechanisms for growth and hyphal breakage (30). It is difficult to extract information about the different mechanisms from these measurements directly. However, using small growth chambers, where the developing morphology can be monitored on-line, it is possible to obtain fundamental information about the growth mechanisms (12). When this information is combined with computer simulations of hyphal element populations, it is possible to obtain some insight into the mechanisms behind hyphal breakage during submerged fermentations (31).

**Quantitative Physiological Studies**

Despite many years of extensive analysis of glycolysis, all the details in regulation of glycolytic flux have not been elucidated. Even in *S. cerevisiae*, where studies of a large number of mutants in the glycolysis have resulted in a meticulous mapping of regulation patterns, all the control structures have probably not been revealed. Thus, it was only a few years ago that fructose-2,6-bisphosphate was discovered not only as an important regulator of phosphofructokinase but of other enzymes as well. Further physiological studies are therefore extremely important to identify yet-unknown regulatory patterns and regulatory compounds. With the recent advances in genome characterization it is now possible to rapidly construct many different mutants by introduction of gene disruptions or over-expression of key enzymes, and this is a good basis for further studies. However, it is important to combine genetic work with quantitative physiological studies in which gene regulation is studied under well-controlled growth conditions, and metabolic fluxes and their control are quantified.

**Study of Gene Regulation at Well-Controlled Growth Conditions.** To illustrate how gene regulation can be studied in submerged fermentations carried out in high-performance bioreactors, two examples are considered. The first example is regulation of the strong TAKA (α-amylase) promoter of *Aspergillus oryzae*, which is often used as expression promoter for production of industrial enzymes by filamentous fungi. In the past there has been much speculation on the regulation of this promoter, but most studies were based on shake-flask experiments, in which it is difficult to distinguish between repression and induction. However, through the use of continuous cultures it can clearly be demonstrated that there is both glucose repression and maltose induction (32,33). Thus, by adding a glucose pulse to a steady-state chemostat operated at a low dilution rate, when the glucose concentration is very low (about 5 mg L⁻¹; i.e., derepressed conditions), it was found that there is an immediate decrease in the α-amylase synthesis (Fig. 4). This clearly indicates that there is glucose repression. Furthermore, by shifting from a glucose-based medium to a maltose-based medium when the cells were grown under derepressed conditions (i.e., at a low glucose concentration) it was possible to show that there is maltose induction (33). Besides allowing qualitative conclusions to be drawn, measurements of the glucose concentration at the milligram per liter level even allowed quantification of the degree of glucose repression (32). These findings are consistent with mapping of the TAKA promoter, which indicates the presence of several creA binding sites (creA is a regulatory protein associated with glucose repression in filamentous fungi) and a binding site for a protein that is induced by starch (or maltose) (34). In a classical approach to analyze the promoter further, one would construct promoter fragments and identify the function of the individual binding sites. In this analysis shake-
Glucose repression of the TAKA promoter in A. oryzae.

At time zero a glucose pulse (resulting in a concentration of 600 mg L$^{-1}$) is added to a steady-state chemostat where the glucose concentration is low (i.e., derepressed conditions). Right after the glucose pulse is added the production of $\alpha$-amylase decreases instantaneously to the level observed at high glucose concentrations. The $\alpha$-amylase concentration (given as units per milliliter) therefore decreases due to washout from the bioreactor. When the glucose pulse is exhausted (indicated by the arrow) there is derepression and $\alpha$-amylase production increases. Source: From Ref. 32.

Figure 4. Glucose repression of the TAKA promoter in A. oryzae. At time zero a glucose pulse (resulting in a concentration of 600 mg L$^{-1}$) is added to a steady-state chemostat where the glucose concentration is low (i.e., derepressed conditions). Right after the glucose pulse is added the production of $\alpha$-amylase decreases instantaneously to the level observed at high glucose concentrations. The $\alpha$-amylase concentration (given as units per milliliter) therefore decreases due to washout from the bioreactor. When the glucose pulse is exhausted (indicated by the arrow) there is derepression and $\alpha$-amylase production increases. Source: From Ref. 32.

Table 5. Invertase Activity in Different Haploid Laboratory Strains of Saccharomyces cerevisiae during Growth on Glucosea

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genetic construct</th>
<th>Invertase activity [U (g DW)$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2180-1A</td>
<td>Wild-type</td>
<td>61</td>
</tr>
<tr>
<td>TAB</td>
<td>Antisense, plasmid</td>
<td>55</td>
</tr>
<tr>
<td>TA25</td>
<td>Antisense, chromosomal</td>
<td>55</td>
</tr>
<tr>
<td>T301</td>
<td>Disruption</td>
<td>263</td>
</tr>
</tbody>
</table>

Note: The antisense fragments were inserted either in a high-copy plasmid or integrated into the chromosome.

Source: From Ref. 10.

With gene disruption of MIG1 a significantly higher activity of invertase is obtained during glucose consumption (Table 5), resulting in an almost parallel uptake of glucose and sucrose (10), as seen in Figure 5. Disruption of MIG1 was also found to partly alleviate glucose repression on maltose uptake and metabolism in laboratory strains; there was an almost parallel uptake of glucose and maltose, whereas there was no effect of MIG1 disruption on maltose uptake in a polyploid industrial strain (35). This was explained to be a consequence of strong catabolite inactivation of the maltose permease in the industrial strain (35). Based on the physiological characterization of the recombinant strains it was concluded that silencing of MIG1 may alleviate glucose repression, but other effects including catabolite inactivation also have an important influence. Only through careful physiological characterizations was it possible to identify the different effects.

Quantification of Metabolic Fluxes. In connection with optimization of metabolite production, where the aim is to direct as much carbon as possible from the substrate into the metabolic product to ensure a high yield, it is of prime importance to quantify how the carbon fluxes are distributed through the various cellular pathways. A very powerful technique for calculation of the fluxes through various pathways is the so-called metabolic flux analysis, where the intracellular fluxes are calculated using a stoichiometric model for all the major intracellular reactions and applying mass balances around the intracellular metabolites. A set of measured fluxes as input in the calculations is used, and these fluxes typically are the uptake rates of substrates and secretion rates of metabolites (36-38). If one performs experiments with $^{13}$C-enriched carbon sources and measures the fractional enrichment of $^{13}$C in intracellular metabolites, it is possible to apply an additional set of constraints in the model, and a better estimation of the intracellular fluxes may be obtained (39). Finally, if one applies information about the isotopomer distributions for the intracellular metabolites much more information is supplied, and it may even be possible to quantify the net fluxes through reactions that are practically reversible (40).

Besides quantification of the various pathway fluxes, and therefore determination of the carbon flows inside the cell, metabolic flux analysis is useful for the following:

- Identification of possible rigid branch points (or nodes) in the cellular pathways. Through comparison
Figure 5. Alleviation of glucose repression by disruption of MIG1 in S. cerevisiae. Results are shown for two batch fermentations carried out on a glucose–sucrose mixture: (a) with a wild-type haploid strain and (b) with the same strain where MIG1 has been disrupted. Obviously disruption of MIG1 results in metabolism of sucrose at a higher glucose concentration as compared with the wild-type strain (i.e., the sucrose concentration begins to decrease right at the start of the fermentation). □, sucrose concentration (mM); ●, biomass concentration (OD); ■, glucose concentration (mM); ○, fructose concentration (mM). Source: From Ref. 10.

of the distribution of fluxes at different operating conditions and in different mutants it is possible to identify whether a pathway node is rigid or flexible (41). Thus through metabolic flux analysis of various mutants of C. glutamicum, Vallino and Stephanopoulos (42,43) concluded that in lysine production the glucose-6-phosphate node is flexible and the pyruvate node is weakly rigid.

- Identification of the existence of different pathways. Formulation of the reaction stoichiometry, which is the basis for metabolic flux analysis, requires detailed information of the biochemistry. However, for many microorganisms such details in the pathway stoichiometries are unknown, and it may not be known whether a given pathway is active. Furthermore, in many cells there are several isoenzymes whose functions are not known in complete details. By calculating the metabolic fluxes with a different set of cellular pathways it may be possible to identify which pathway is most likely to exist or to obtain indications of the functions of different isoenzymes and/or pathways. Thus, in analysis of anaerobic growth of S. cerevisiae, we found that alcohol dehydrogenase III, a mitochondrial enzyme whose physiological function has not been identified, plays an important role in maintaining the redox level inside the mitochondria (38). However, this type of analysis should always be followed by enzyme assays in which the presence of enzyme activity in cell extracts is tested.

- Calculation of nonmeasured extracellular fluxes. Normally the number of fluxes that can be measured is larger than what is needed for calculation of the intracellular fluxes. In this case it is possible to calculate some of the extracellular fluxes, such as the rate of production of various by-products, by use of the stoichiometric model and the measured fluxes.

- Examination of the influence of alternative pathways on the distribution of fluxes. In connection with optimization of metabolite production it may be possible to identify one or several constraints for increasing the yield of a metabolite on the substrate or the flux leading to the desired metabolite. In these cases it is possible to envision various scenarios, such as examining whether insertion of a new pathway or an isoenzyme (or perhaps deletion of an isoenzyme) can have a positive effect on removing this constraint, leading to an increased flux toward the desired metabolite. In a study of penicillin production we calculated that the yield of penicillin on glucose is likely to be higher if cysteine, one of the precursors for penicillin production, is synthesized by direct sulphydrylation rather than by the transsulfuration pathway claimed to exist in P. chrysogenum (37).

- Calculation of maximum theoretical yields. Based on a stoichiometric model it is normally possible to calculate the maximum theoretical yield of a given metabolite if a given set of constraints is specified. This value is of interest because it specifies an upper limit for the yield in the process. This has been illustrated both for lysine production by C. glutamicum (36) and for penicillin production by P. chrysogenum (37). In these calculations it is necessary to introduce a constraint, for example, that reactions do not run in thermodynamically unfavorable directions.

Control of Metabolic Fluxes. One of the most important aspects of quantitative physiology is control of flux. As dis-
cussed, the concept of metabolic flux analysis is useful for studies of interactions between different pathways and for quantification of flux distributions around branch points, but it does not allow for evaluation of how the fluxes are controlled (i.e., how the rates of synthesis and conversion of metabolites are kept in close balance over a very wide range of external conditions without catastrophic rises or falls in the metabolite concentrations). The discoveries in the 1950s of feedback inhibition, cooperativity, covalent modification of enzymes, and control of enzyme synthesis introduced a number of molecular effects that may play a role in control of fluxes. With these many different mechanisms it is not surprising that disputes often arises over how the flux through a given pathway is controlled. Since the many reports on enzyme regulation typically are verbal and qualitative (e.g., phosphofructokinase is the major flux-controlling enzyme of glycolysis in muscle), it is difficult to discriminate between different findings, and in many cases different findings may seem conflicting. Furthermore, one often finds terms such as rate-limiting steps and bottleneck enzyme when control of flux is discussed. Thus, the early findings that it is the first enzyme in a pathway or the first enzyme after a branch that is under some type of control (e.g., by feedback inhibition) often results in statements such as "the first step in a pathway is the rate-limiting step."

The concept of metabolic control analysis (MCA) tells us that these kinds of qualitative statements have no meaning, since flux control is distributed over all the steps in a pathway, but that some steps may exert a higher degree of flux control than others. The concept of MCA was developed from the landmark papers of Kacser and Burns (44) and Heinrich and Rapoport (45). Its basis is a set of param-eters, called elasticity coefficients and control coefficients, that quantify the control within a reaction network. The elasticity coefficients are given by

$$e_{ij} = \frac{X_j}{v_i} \frac{\partial v_i}{\partial X_j}$$  \hspace{1cm} (1)

and they express the sensitivity of the rates of the enzyme-catalyzed reactions ($v_i$) toward the metabolite concentrations ($X_j$). The most often used control coefficients are the flux control coefficients (FCCs), which are given by

$$C_{ij} = \frac{v_i}{J_j} \frac{\partial J_j}{\partial v_i}$$  \hspace{1cm} (2)

The FCCs express the fractional change in the steady-state flux through the pathway ($J_j$) that results from an infinitesimal change in the activity of the enzymes (or reaction rates). The FCCs and the elasticity coefficients are related to each other via the summation theorem, which states that the sum of all the FCCs is 1, and the connectivity theorem, which states that the sum of the product of the elasticity coefficients and the FCCs is zero.

Three different groups of experimental methods are available for determination of the FCCs:

- **Direct methods**, where the control coefficients are determined directly
- **Indirect methods**, where the elasticity coefficients are determined and the control coefficients are calculated from the theorems of MCA
- **Transient metabolite measurements**, where the metabolite concentrations are measured during a tran-

### Table 6. Overview of Methods for Determination of Flux Control Coefficients

<table>
<thead>
<tr>
<th>Method</th>
<th>Procedure</th>
<th>Advantages and disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic manipulations</td>
<td>Alternate the expressed enzyme activity through genetic manipulation, such as by inserting inducible promoters</td>
<td>Robust method that gives direct answers, but is very laborious</td>
</tr>
<tr>
<td>Enzyme titration</td>
<td>Vary the enzyme activity through titration with purified enzymes</td>
<td>Simple and straightforward procedure, but it can only be applied for pathway segments that are completely decoupled from the rest of the cell</td>
</tr>
<tr>
<td>Inhibitor titration</td>
<td>Vary the enzyme activity through titration with specific inhibitors</td>
<td>Simple and easy to apply, but requires the existence of specific inhibitors</td>
</tr>
<tr>
<td>Double modulation</td>
<td>Measure the metabolite levels at different environmental conditions and determine the elasticity coefficients by calculation of differentials</td>
<td>Elegant approach, but it requires two independent changes in the metabolite levels, which are difficult to obtain due to the high degree of coupling between intracellular reactions</td>
</tr>
<tr>
<td>Single modulation</td>
<td>Similar to double modulation, but it is based on knowledge of one of the elasticity coefficients</td>
<td>More robust than double modulation, but it requires knowledge of one elasticity coefficient</td>
</tr>
<tr>
<td>Top down approach</td>
<td>Based on grouping of reactions and then using double modulation, for example</td>
<td>Very useful, but do not directly give all the FCCs of the system</td>
</tr>
<tr>
<td>Kinetic models</td>
<td>Direct calculation of the elasticity coefficients from a kinetic model</td>
<td>Robust, but relies on the availability of a reliable kinetic model for the individual enzymes in the pathway</td>
</tr>
</tbody>
</table>
sient and this information is used to determine the control coefficients (46).

Table 6 gives an overview of the different direct and indirect methods.

Whereas the elasticity coefficients are properties of the individual enzymes, the FCCs are properties of the system. The FCCs are therefore not fixed but change with the environmental conditions. This has been illustrated in analysis of the penicillin biosynthetic pathway (47,48). Based on a kinetic model for the enzymes in this pathway, the FCCs were calculated at different stages of fed-batch fermentation, and a drastic shift in the flux control was found. During the first part of the cultivation the flux control was mainly at the first step in the pathway, the formation of the tripeptide LLD-ACV by isopenicillin N synthetase (IPNS). This shift in flux control is due to intracellular accumulation of LLD-ACV, which is an inhibitor of ACVS. Obviously, it makes no sense to talk about a rate-limiting step or a bottleneck enzyme in this process. Besides the shift in flux control, it is interesting to note that most of the flux control is by the first two steps in the pathway (48). Furthermore, through analysis of the kinetic model it was found that the value of the FCCs depends on the dissolved oxygen concentration, which is a substrate in the IPNS catalyzed reaction (48).

BIBLIOGRAPHY

See also Ammonia Toxicity, Animal Cells; Bioreactors, Air-Lift Reactors; Bioreactors, Continuous Stirred-Tank Reactors; Bioreactors, Fluidized-Bed; Secondary Metabolite Production, Actinomycetes, Other Than Streptomyces.

FERMENTER DESIGN

MARVIN CHARLES
Lehigh University
Bethlehem, Pennsylvania

JACK WILSON
ABEC, Inc.
Allentown, Pennsylvania

KEY WORDS

Bulk mixing
Clean-in-place
Heat transfer
Impellers
Oxygen transfer
Power input
Sampling systems
Seals
Sterilization
Stirred tank

OUTLINE

Introduction
Scope
Design Philosophy
Safety and Regulatory Compliance
Containment: Worker and Community Biosafety
Physical Safety
Product Safety
Design Basis and Other General Considerations
Process Requirements: Basics
Broth Rheology and Its Effects on Transport Processes
Oxygen Transfer/Aeration
Power
Bulk Mixing
Gas Holdup, Foaming, and Aerosol
Heat Transfer
Sterilization
Cleaning
Putting It Together: Preliminary Design Calculations
Mechanical Design

INTRODUCTION

Scope

The scope of this article is limited to design of agitated vessels used for aerobic, single-cell (bacteria and yeast), and filamentous (bacteria and fungi) fermentations. It presents general design principles, combining basic concepts with practical matters including regulatory compliance, safety, maintenance, cleanability, ease of use, and cost, all of which are strongly interrelated. It focuses primarily on fermenters used to produce human and animal health products or their precursors, but the general approaches discussed are applicable to other types of products (e.g., industrial enzymes, food products); the primary differences are emphasized.

Included in the presentation are fermenter vessels; agitation; aeration; heat transfer; sterilization; cleaning; and piping systems. Control and data acquisition are discussed only to the extent that they influence the items already noted.

It is assumed that the reader is familiar with the basics of fermentation, general fermenter construction, and nomenclature. Those not having such background are advised to consult basic references (1-3).

Design Philosophy

Our basic philosophy comprises the following simple principles:

1. Each design case has unique requirements and calls for individualized application of the basic design concepts and practices discussed herein. The one-size-fits-all approach usually results in sleeves that drape into your food, arm holes that cut off circulation, pant legs that are made to trip over, or some combination of these features and others. The magic number approach—all problems have simple solutions based on codified numbers such as fixed geometric ratios (4)—leads to the same place.

2. Successful design requires a systems approach. A fermenter is a system that is part of a process system, and the process is part of a plant system. All these systems interact with each other, with real people, with control systems, and with other systems related to regulatory compliance, safety, documentation (including protocols, SOPs, etc.), change control, maintenance, and so forth. Failure to take these
interactions into account during fermenter design usually results in considerable pain, not only for those guilty of the omission, but also for innocents who were never asked for planning and/or design input, but who must live with the result. Note that appropriate documentation, protocols, SOPs, change control, and so on should be considered important elements of design and operation of any plant, licensed or not.

3. Compromise is always necessary. Nothing in any project is 100% right, and nothing has to be. He or she who looks for 100% of anything will hold up a project needlessly and will generate a lot of animus. (The words actually used in such cases cannot be used here.)

4. The time to be thinking about all of these points is prior to and during design, not when you’re standing on the plant floor trying to validate a continuous mixer that has only one port and no instrumentation.

We discuss next the bases for safety and regulatory considerations, primarily to sensitize readers to these issues early on. They are all too often overlooked during early stages of design, when the focus is on satisfying process requirements (e.g., oxygen transfer). Unfortunately, this frequently leads to important constraints being ignored and hence to designs that require expensive and often cumbersome “fixes.”

SAFETY AND REGULATORY COMPLIANCE

All fermenter designs are influenced by safety and regulatory factors, the extent depends on (1) the nature of the product and its intended use, (2) the process, and (3) the nature and location of the facility. For example, even fermentation products not regulated by the U.S. Food and Drug Administration (FDA) must adhere to safety requirements, which in some cases are very strict. Furthermore, no professional design can escape the effects of a long list of code requirements, for example, the ASME pressure vessel code (5). By the same token, not all products that are FDA regulated are equally affected by safety, regulatory, and code requirements. We must consider each case on the basis of its own unique requirements.

Our objective in the rest of this section is to introduce some of the requirements and to provide some rational bases for their implementation, details of which are discussed later. Much of the discussion is directly applicable to fermenters used for production of human or animal health products or for their precursors, but also applies to other cases to varying extents.

Containment: Worker and Community Biosafety

A fermenter must be an integral part of a system designed to insure safety at three levels: product, worker, and community. Many of the methods used to protect the product also serve to protect the plant personnel and the community (e.g., use of closed systems and HEPA filters); however, there are potential points of conflict (6). For example, some containment practices that call for completely welded hard piping to a contained drain line for any condensate that could be exposed to culture fluid could expose the product to drain line contaminants. One resolution of this problem is to use steam locks in all such lines. All such conflicts encountered to date have been resolved, but not always easily or relatively inexpensively.

Insurance of product safety is tied primarily to compliance with FDA (or similar organizations) regulations. This will be considered in the subsection “Product Safety.” Here we deal with worker and community safety.

Protection must be provided against potential ill effects of fermentation products (e.g., cytotoxins, allergens), organisms (e.g., pathogens, whether recombinant or not), fermentation byproducts (e.g., pharmacologically active precursors of the active product), or some combination. The practice of providing such protection is called containment, which is defined as insuring that deleterious fermentation components can’t be transported to any area inside or outside the plant before they have been rendered harmless. To accomplish this, containment must be exercised at several levels (7). We consider in this article only direct or “primary” containment of fermenters; however, one should not lose sight of the fact that the other levels must be considered during fermenter design, if for no other reason than to ensure that the fermenter design and operation will be consistent with the overall containment strategy.

Containment levels have been defined by the National Institutes of Health (NIH) (8), based on the potential danger of organisms. These levels, in increasing order of potential danger are GILSP, BL1-LS, BL2-LS, BL3-LS, and BL4-LS. (BL4-LS must be handled on a case-by-case basis in consultation with the Centers for Disease Control and others. It is, thank goodness, well beyond our scope.) There are other classifications that have been developed in the United Kingdom (9), the European Union (10), and elsewhere, but these do not differ markedly from the U.S. scheme. It is important to know that (1) all of these guidelines address what must be accomplished but not how they should be accomplished, and (2) there is not complete agreement on what is actually required to implement the guidelines. In addition, the guidelines do not address containment of products. Although many of the considerations for product containment are the same or very similar to those for organism containment, there are some very important differences. Here, we confine our attention to organisms; the interested reader should consult the Refs. 11–14 for more specific information concerning hazardous products. One also should consult U.S. Environmental Protection Agency (EPA) regulations applicable to microorganisms and their products.

Most existing processes use organisms that require either GILSP or BL1-LS containment; nevertheless, many plants are built to satisfy requirements for BL2-LS (actually, somewhere between BL2-LS and BL3-LS). The reasons for this include satisfying FDA requirements (e.g., safety of the product) and the fact that the additional cost is not large and is relatively cheap insurance. There also
is the oft-stated reason that such a route insures greater flexibility for future operation (15). One might add to this the fact that the rules are not well established and that requirements could change substantially between the time of design and the time production begins. It is also worth noting that there is growing interest in using organisms that do, in fact, require BL2-LS and even BL3-LS containment, which will probably focus greater attention on further development of acceptable implementation practices as well as changes in the guidelines. Note that equipment requirements and cost for BL2-LS and BL3-LS are currently about the same (this probably will change); BL3-LS facilities requirements and cost are very much greater than for BL2-LS.

Finally, it is important to note that containment is nothing new. Highly pathogenic organisms have long been used to produce therapeutics and biological warfare components; hence, there is a considerable body of experience dealing with the subject (11,16). There also is considerable guidance to be had from the nuclear industry; nevertheless, the nature, sizes, and large number of new commercial processes, along with new methods, materials, and so on, will force heightened awareness, scrutiny, review, and modernization or change.

Physical Safety

Most physical safety issues are addressed by U.S. Occupational Safety and Health Administration (OSHA) requirements (17) and by various national codes such as the ASME Pressure Vessel Code (5) and the National Electrical Code. There also are many local construction codes that must be satisfied. Among these are various earthquake-resistant construction codes that are also applicable to containment considerations. And then there are the requirements of the final arbiters: the insurance companies.

Product Safety

We focus here on FDA requirements, but the reader is urged to keep in mind that he or she also must consider those of other regulatory agencies and be aware of the unsettling facts that it is not always clear which agency has jurisdiction, and there often are conflicting requirements.

The primary purpose of FDA regulations is to ensure product safety, purity, and efficacy. The legal bases for the regulations and their enforcement can be found in the Code of Federal Regulations (CFR) (18) and a number of other FDA documents such as Points to Consider (19,20), Inspection Guides (21,22), and Guides to Industry (23,24). Industrial implementation of all the regulations, along with ongoing improvements, is referred to as current good manufacturing practices (cGMP).

There is an endless stream of writings in professional journals, short courses, and such concerning the regulations and cGMP practice. Unfortunately, there is a great deal that is not covered clearly in writing, and there is a constant proliferation of new and/or revised guidelines followed closely by lots of (frequently conflicting) interpretations and opinions. In addition there has been a problem of nonuniform application. For example, regulatory scrutiny has been far stricter for fermenters used to produce active molecules directly than for those used to produce precursors; there has been considerable variation from product to product within a given class; and manufacture of human drugs has been regulated much more rigidly than manufacture of animal drugs. While understandable to some extent, such nonuniformity has caused considerable confusion. These gaps are beginning to be closed, and this may have considerable influence on the design of fermenters—particularly those used to produce precursors and animal drugs.

In simplest terms, cGMP requires that the combination of fermentation process equipment and operating protocols must consistently yield product that meets acceptable specifications and is capable of being converted/purified consistently to final product, meeting approved product specifications. From this general sense of intent it can be argued that failure to have control over the fermentation puts the final product at risk. Few would argue with this general concept; however, there is considerable debate as to interpretation and extent.

Requirements that flow from the need to have control are based on relatively simple ideas:

1. If an organism is subject to environmental, medium, or other conditions outside the range in which it is known to yield product meeting acceptable specifications and capable of being converted/purified consistently to final product meeting approved product specifications, then we have no guarantee that it does not produce other products with which the recovery system can not cope and that can escape detection by the analytical methods in place.

2. If a fermenter becomes contaminated with other microbes, said microbes could produce toxins that could be carried undetected to the final product. It is possible that this could occur without altering the behavior of the process organism. It also is possible that products of the contaminant could cause the process organism to make toxins that could go undetected into the final product. Given these possibilities, plus the fact that it is not possible to prove that the contaminant will always be the same, evidence of contamination is evidence for lack of control.

3. If a fermenter is not cleaned properly, deleterious microbial or nonmicrobial products could remain to contaminate the next batch in such a way that impurities could be carried undetected to the final product. Similar statements can be made about other contaminants introduced via other routes as a result of poor cleaning practices.

Again, it is unlikely that many would argue with these points, but again there is considerable debate as to interpretation and extent.

Requirements 1–3 can be translated to specific requirements for, among other things, controllability and reliability of fermentation conditions, sterilization, and aseptic and cleaning operations. Controllability requires that the fermenter and its subsystems be designed and constructed in ways that make possible control of environ-
mental variables within the operational ranges required. This obviously means that mechanical design must be in harmony with control system design.

Sterilization/aseptic operation translates to (1) destroying any microbial contaminants that may be present in any part of the equipment that might contact process fluids and (2) insuring that no microbes (other than the production organism) can enter after the equipment has been sterilized. The latter embodies the concept of the “sterile barrier.” To these ends, the following apply:

- The fermenter and all its ports and direct attachments must be sterilizable.
- All piping that will contact process fluids (including additives) and/or provide paths into the system (e.g., the air exhaust line) must be sterilized initially. Some (e.g., sampling lines, addition lines) must also be sterilizable at any time during a fermentation.
- Inlet gases (e.g., air, oxygen) and all additives (e.g., medium components, acid, base, antifoam) must be sterilized before they contact any sterile process piping.
- All penetrations (e.g., drive shaft, probe ports) must be sterilizable.

The preferred sterilization method is automatic sterilization-in-place (SIP) with steam. This is discussed at length later.

Clean operation requires, among other things, the following:

- The system must be designed such that any surface that can contact a process stream can be cleaned consistently to a level that insures that the product will be free of soils resulting from a fermentation.
- Nothing in the system that comes in contact with process fluid can introduce unacceptable and unidentifiable materials.

Note (for example) that 21 CFR 211 (Part D) (guidelines for design of equipment for licensed facilities) does not have any specific requirements for fermenter sterilization or cleaning; however, the preceding discussion coupled with the intended use and cGMP concepts makes it difficult to argue the point. But again, there remain the questions of rational interpretation and extent of application.

Riding along with cGMP is the practice of validation. This is an extensive, expensive, and controversial area for which and about which reams of reams have written. Following is the formal definition of validation:

Establishing documented evidence that provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes. (25)

What this means is that not only does the equipment have to be designed, installed, and operated so as to run consistently within well-defined ranges of process variables to consistently yield product that meets acceptable specifications and is capable of being converted/purified consistently to final product, meeting approved product specifications, but that there must be adequate documentation to prove this before product can be sold. In addition, one must demonstrate that this goal can be achieved under so-called worst-case conditions; there is often considerable debate as to what this really means (26). It is beyond our scope to consider validation in detail. Suffice it to say that it can translate to checking and testing every component of the equipment and documenting not only its proper function but also its history. The brave of heart are referred to the bulging literature on the subject; Refs. 27–30 provide a start.

All of this means we must take great care to consider all the details, such as materials, corrosion issues, machining methods, vendors, welds, and types of steam and water, keeping in mind that all must be considered as part of a system that must interface “seamlessly” with the systems of cGMP and validation. Failure to consider this before and during design is almost always very costly. Design should be done so as to facilitate validation, and subsequently, routine cGMP operation, maintenance, and so forth. “Simple” things such as judicious placement of access valves and piping isolation can go a very long way to ensuring minimum pain and maximum operability. Having said all this, we reiterate that not all fermenters are subject to all these requirements, and even those that are, are subject to them in varying degrees (at least in practice).

**DESIGN BASIS AND OTHER GENERAL CONSIDERATIONS**

A design basis for fermentation equipment should derive from a facility/process design basis. The latter should include (among other items) the general nature of the facility (e.g., research and development vs. production, single product vs. multiuse); product(s) specifications; regulatory, containment, and other requirements; level of automatic operation; general processing scheme (e.g., batch); nature of individual process steps; productivity, concentrations, and so forth; cleaning requirements; special considerations (e.g., earthquake-proof construction); critical valving and instrumentation; staffing requirements and constraints; architectural and general floor plan constraints; and utilities. Most of these will have some influence on fermenter design—some in more subtle ways than others.

It is difficult to overstate the importance of the initial definition that will derive from the design basis. Obviously, a rational design basis for a fermenter must also be based on fermentation characteristics as well as operating cycle and productivity required (which should derive from the overall process design basis). From these will flow the sizes and number of vessels, and definitions of oxygen transfer, heat transfer, power, and bulk mixing requirements. The design must satisfy these but must also satisfy requirements for regulatory compliance, safety, cleaning, facile operation, and maintenance. All of these factors are highly interactive (e.g., design for oxygen transfer affects design for cleaning); hence, responsible design will almost always require several iterations to ensure the greatest probability of success and to minimize lost time caused by instal-
lations, operation, and various problems. The iterative nature of the design (and, unfortunately, construction) process makes most important the existence of a well-crafted, well-implemented, and well-documented change control process. Finally, as with any engineering project, failure to have a solid basis of design, a complete scope, accurate process flow diagrams and piping and instrumentation diagrams (P&IDs), and accurate process timing will almost certainly result in added cost, lost time, and worse.

**PROCESS REQUIREMENTS: BASICS**

One of the first steps in fermenter design is translation of process demands to oxygen transfer, heat transfer, bulk mixing, and power requirements—the so-called transport processes (TPs). Sound translation is not straightforward because the TPs interact not only with each other, but also with vessel geometry and other design factors. The nature of each TP and the relationships among them vary with the process requirements. They also vary significantly with the basic nature of the organism and the culture broth. The major factors here are the rheological nature of the broth and the sensitivity of the organism to fluid mechanical forces. In some cases, sensitivity of the organism to temperature and other environmental factors, such as pH, impose tighter constraints (e.g., wall temperature control, uniformity of mixing).

Single-cell organisms (most bacteria and yeasts) tend to tolerate fluid forces very well (there are a few exceptions) and tend to have low-viscosity Newtonian broths. Mycelial organisms (fungi and some mycelial bacteria such as streptomyces) tend to be more prone to damage by fluid mechanical forces than are single-cell organisms and tend to have high-viscosity, non-Newtonian broths. Detailed discussion of all these factors is beyond the scope of this article; however, we now discuss a few points concerning rheological behavior, and point to fluid effects wherever appropriate in the remainder of the article.

**Broth Rheology and Its Effects on Transport Processes**

Newtonian viscosity depends only on composition and temperature; the nature of the fluid motion does not affect it. Non-Newtonian viscosity does depend on the nature of the fluid motion. This dependency is usually expressed in terms of fluid shear rate (a measure of how rapidly fluid velocity changes from one point to another point close by).

There are several types of non-Newtonian behavior (e.g., pseudoplastic, Bingham) that can be described quantitatively by a host of mathematical models. One model used frequently to describe non-Newtonian fermentation broths (31–33) is the so-called power law:

\[ \eta = K \gamma^{(n-1)} \]  

(1)

where \( \eta \) is viscosity, \( \gamma \) is shear rate, and \( n \) and \( K \) are constants. Such relationships can be useful if one has rheological data for the subject broth and if the design correlations used incorporate rheological properties in a meaningful and accurate way. It is seldom that either of these conditions is satisfied, and almost never that both are satisfied.

Also, there is a considerable amount of misleading and/or inaccurate information in the literature concerning broth rheology, its measurement, and its use; the reader is cautioned to tread carefully in this area (4,34). If accurate viscosity information is available, it can be a valuable qualitative and sometimes semiquantitative guide, if interpreted and used properly.

Mass transfer (oxygen), heat transfer, and bulk fluid motion all depend strongly on rheological characteristics. Generally, all are poorer in non-Newtonian than in Newtonian broths; indeed, the rheological characteristics of mycelial broths can and do impose severe constraints. Two of many examples are the following:

1. Heat and oxygen transfer rates tend to be anywhere from 50 to 5% of what they would be for typical (e.g., Escherichia coli) bacterial fermentations.
2. Bulk mixing quality (homogeneity) is much poorer than in typical Newtonian broths; therefore, accurate monitoring and control are much more difficult.

Difficulties associated with viscous, non-Newtonian rheology also extend to other areas such as cleaning.

**Oxygen Transfer/Aeration**

Please note that in this and subsequent sections we include design and operating rules of thumb based on our experience and the experiences of others. They are intended to be helpful guides, not edicts.

**Requirements.** Oxygen transfer rate (OTR) requirements are usually dictated by conditions during the most active part of the growth phase (other phases require oxygen, but supply rate usually is not as high). The requirement is calculated from

\[ OTR = (Y_{x0})^{-1}(\mu X)_{\text{max}} \]  

(2)

where \( \mu \) is the specific growth rate (h\(^{-1}\)), \( X \) is the cell mass concentration (g/L dw), and \( Y_{x0} \) is the cell yield coefficient on oxygen (g cells dw/g \( O_2 \)). Note that the maximum growth rate and maximum cell mass are not always reached simultaneously.

**Satisfying the Oxygen Material Balance: Gas Flow and Linear Velocity.** The very first thing we must do in satisfying an OTR requirement is to insure that we have a closed oxygen material balance. For most practical operating conditions, the rate of change of oxygen inventory in a fermenter is very small compared to oxygen flows in the inlet and outlet and to the OTR. It also can be demonstrated easily that the total molar flow rate of gas does not change enough to cause any loss of sleep. Given these practical realities, the oxygen material balance is

\[ OTR = (1,000 \times 60)(y_{in} - y_{out})/V_L \]  

(3)

where, \( F \) is the gas flow rate (mol/min), \( y \) is the mole fraction of oxygen in the gas, and \( V_L \) is the liquid volume (L). Note that the usual units for OTR are millimoles per liter.
per hour and that the factors of 1,000 and 60 in equation 3 are conversion factors needed to keep unit consistency.

The material balance also can be expressed as

\[ F = \text{OTR} \times V_L/(1,000 \times 60 \cdot \gamma_{in}) \]  
(4)

where \( \varepsilon \) is the oxygen transfer efficiency and \( \gamma_{in} \) is the oxygen mole fraction in the inlet gas. Equation 4 is a more useful form because we know that under practical conditions \( \varepsilon \) will be between 0.15 and 0.35 for typical Newtonian broths, and between 0.05 and 0.15 for typical non-Newtonian broths.

As a general rule, heroic efforts will be required to get OTRs over 300 mmol/(L h) in large fermenters (>5,000 L). Even if the effort is expended and it is successful, it will usually create heat transfer problems (see later) that will require even greater effort to overcome. Our advice is to consider other options very seriously.

The material balance provides information about the gas volumetric flow rate, which often is expressed as the standard flow, \( v_{std} \), in units of standard liters per minute (SLPM):

\[ v_{std} = 22.4F \]  
(5)

This is useful because most flow meters are calibrated in terms of standard flow; however, actual flow is more useful for fermenter design purposes. One problem encountered in calculating the actual flow is that it increases from the bottom to the top of the broth because of pressure change. As a practical matter, however, it is usually adequate to use the average pressure in the tank:

\[ v_{act} = v_{std} \times ((1/P_{avg})(T_f/298)) \]  
(6)

where \( v_{act} \) is the gas flow at average pressure (L/min), \( P_{avg} \) is the average pressure (atm, abs), and \( T_f \) is the fermentation temperature (K). One might want to revisit this for very tall vessels.

The standard gas flow rate also is expressed in standard gas volumes per liquid volume per minute (VVM).

Among the important information that can be obtained from \( v_{act} \) is the gas linear velocity, \( V_g \) (cm/min):

\[ V_g = 4 \times 1000 \cdot v_{act}/(\pi \times D_i^2) \]  
(7)

where \( D_i \) is the inside diameter of the fermenter. \( V_g \) affects mass transfer, deliverable power via the impellers, foaming, gas holdup, and aerosol formation. As a rule of thumb, \( V_g \) should be held below 200 cm/min to avoid problems associated with excessive gas holdup, foaming, and aerosol formation. These effects are discussed later.

**Mass Transfer.** The rate at which oxygen can be transferred is dictated by three factors: the driving force, the resistance to transfer, and the contact area between the gas and liquid phases. This is usually expressed as

\[ \text{OTR} = K_g \cdot \bar{a} \cdot (\Delta P)_{LM} \]  
(8)

where \( K_g \) is the mass transfer coefficient ([mmole M]/[L h atm]), and \( \bar{a} \) is the interfacial area per unit volume (M²/M³). \( \Delta P_{LM} \) is the log mean pressure driving force, defined as

\[ \Delta P_{LM} = (P_{O,in} - P_{O,out})/L \cdot [P_{O,in} - P_{O}*]/(P_{O,out} - P_{O}* ) \]  
(9)

where \( P_{O,in} \) is the partial pressure of oxygen in the gas inlet (atm, abs), \( P_{O,out} \) is the partial pressure of oxygen in the gas outlet (atm, abs), and \( P_{O}* \) is the partial pressure of oxygen that would be in equilibrium with the dissolved oxygen concentration in the existing liquid (atm, abs). Note that use of the log mean driving force assumes that the liquid is mixed perfectly (homogeneous liquid phase) and that the gas moves in plug flow through the fermenter. These assumptions are acceptable for most low-viscosity Newtonian broths, but they can be quite poor for high-viscosity, non-Newtonian broths. Unfortunately, no useful alternatives have yet been proposed; therefore, extra caution should be exercised in interpreting calculated results for non-Newtonian broths.

Other mass transfer coefficients are used for driving forces other than \( \Delta P_{LM} \). The most common among these is \( K_g (\text{min}^{-1}) \), which is used in conjunction with a dissolved oxygen concentration driving force. The reader is cautioned that \( K_g \) and \( K_{ga} \) are sometimes confused with each other—even in the literature.

A given OTR requirement can be satisfied by various combinations of driving force and mass transfer coefficient. It is important to note, however, the following:

- Factors that influence driving force also can influence \( K_{ga} \).
- Each combination will have different effects on the ultimate design of the vessel, piping system, and support systems, as well as on operational factors related to regulatory compliance cleaning, maintenance, and so on. Such effects are discussed as we proceed.

**Driving Force.** Driving force is affected by total pressure, gas inlet oxygen mole fraction, total gas flow rate, and the dissolved oxygen concentration. Total pressure and oxygen mole fraction affect the oxygen partial pressure directly. The effect of total gas flow is a bit more subtle. For a given oxygen transfer rate, increasing the total gas flow increases the mole fraction of the outlet gas flow, thereby increasing the average mole fraction of oxygen in the gas; hence, the average driving force is increased. This can be seen quantitatively via an oxygen material balance which gives

\[ P_{O,out} = P_{O,in} - \text{OTR} \times P/(1,000F) \]  
(10)

where \( P \) is the total absolute pressure (atm) and \( F \) is the total molar flow (mmol/h) of gas.

The characteristics of \( K_g \) and \( a \) are complex and not completely understood. Our first clue to this is that \( K_{ga} \) is expressed as a product in terms of empirical correlations (we don't understand enough about them to get reliable quantitative guidance from first principles). Typical correlations have the form...
where $P_g$ is the power delivered to aerated broth (HP), $V_i$ is the unaerated liquid volume (L), and $\delta$, $\alpha$, and $K$ are usually advertised as constants. There also a several variations on this form. A discussion of all the fine points of such correlations is beyond our scope; however, the reader should take note of the fact that published correlations usually cannot be relied upon to provide accurate predictions. They should be used as qualitative or semiquantitative guides only. Reasons for this caveat include the following:

- They are usually applicable only under conditions at or near the ones used to develop them. Unfortunately, most are developed at scales and under conditions that are not realistic for production.
- They can have pronounced dependence on fermenter geometry.
- They do not scale up well, or at all. The "constants" in equation 11 and its relatives usually aren't constant.

These caveats should be made more emphatically for viscous non-Newtonian broths than for Newtonian broths because non-Newtonian rheological characteristics can have profound effects on $K_ga$ (as well as the other transport properties). Most $K_ga$ correlations do not account for broth rheology. Those that do, usually do so inadequately or in ways valid only for a particular broth. It also is important to note that the broth rheology can vary dramatically during a fermentation and that the rheological characteristics can be affected by the fermentation conditions, ranging from the nature of the inoculum to the history of the agitation speed.

As noted earlier, we can get some qualitative guidance from the form of equation 11. We’ll do that, but please keep in mind that the relationships are highly nonlinear, the factors that influence $K_ga$ are dependent on each other in practical operation, and it is not always possible to control variables at the levels you would like. For example, equation 11 predicts that increasing power input per unit liquid volume will increase $K_ga$. This usually turns out to be true in practice if the power can actually be delivered to the fluid and delivered in a manner that will contribute to $K_ga$. Such might not be true in any given case.

The fact that a drive’s power rating is 100 hp doesn’t mean that 100 hp can be delivered to the broth. Deliverability depends on, among other things, geometry (vessel and agitator), rheology, agitator speed, and gas linear velocity (see later). The effects of all are complex and interactive; the interrelations are more complex for non-Newtonian broths than for Newtonian broths. The manner in which power is delivered depends largely on impeller type and broth rheology. For example, oxygen transfer rates to highly viscous, non-Newtonian mycelial broths have been shown to be larger for A-315 (hydrofoil) impellers than for Rushton (disk turbine) impellers delivering the same power (35). There appears to be no meaningful difference for Newtonian broths.

The bottom line is that one should try to obtain experimentally determined correlations for the subject organism and broth at meaningful scales. Failing that (which usually is the case because of time and resource constraints), one should try to use correlations developed for systems as similar as possible to the one at hand. Considerable art is involved here.

The reader can find more about $K_ga$ in the literature (35–38).

**Power**

Power delivered to the broth is used for micromixing and gas dispersion, which are related to mass transfer, and macromixing, which provides overall homogeneity (discussed later). Agitated vessel power is delivered via two mechanisms: direct mechanical power from the impellers and gas expansion. The bulk (>90%) of the power comes from the impellers as long as the fluid motion is under their control, a condition that prevails so long as the impellers are not flooded. There are reasonably reliable correlations available to determine flooding conditions for Newtonian broths (39); however, flooding usually is not a problem under typical conditions used in most Newtonian broth fermentation. Flooding is more likely in highly viscous, pseudoplastic broths typical of mycelial and polysaccharide fermentations because the viscosity near the impeller is much lower than in the rest of the broth. As a result (assuming air is introduced under the impeller), air tends to channel toward the impeller, thereby enshrouding it and decreasing the deliverable power (4). This can dramatically decrease overall mass transfer (and heat transfer) rates and quality of bulk mixing quality (see later). Reliable flooding correlations for non-Newtonian broths have not been published, but some companies have accumulated a considerable amount of data for the fermentations they practice.

Calculation of power input relies on empirical correlations. As with $K_ga$, many correlations have been published (see for example Refs. 40-42), but they are not particularly reliable for many of the same reasons given for the unreliability of $K_ga$ correlations; therefore, we give the same advice here as we did for $K_ga$.

Among the more popular approaches used is the one that relies on the aeration number, $N_A$ (dimensionless), defined as

$$N_A = \frac{v}{N^2 D_i^3}$$

(12)

where $v$ is the gas volumetric flow rate (m$^3$/min), $N$ is the agitation speed (min$^{-1}$), and $D_i$ is the impeller diameter (m). Published correlations give the ratio of gassed power to ungassed power as empirical functions of $N_A$:

$$\frac{P_g}{P_u} = \text{Func}(N_A)$$

(13)

where $P_u$ is the power delivered to the same broth agitated under the same conditions but unaerated.

Each such correlation pertains to a specific impeller type, a single fermenter geometry, and a fairly narrow range of operating conditions. In some such correlations...
(40), a unique function is presented for each impeller type (Fig. 1). In others the relationship is shown not to be unique (43). In most cases, the correlations are for a single aerated impeller. The situation is more confusing for non-Newtonian broths.

Obviously, to use equation 13 one must be able to calculate the ungassed power. This can be done via other empirical correlations that give ungassed power in terms of the dimensionless power number \( N_p \), defined as

\[
N_p = P_{ug} \times \frac{g_c}{N^3} \times D_i^2 \times \rho
\]

where \( g_c \) is Newton’s law conversion factor, and \( \rho \) is the ungassed broth density (lb/ft^3). There are published correlations that give \( N_p \) as a function of Reynolds number for a wide range of single impellers (44). Each is specific to a particular vessel geometry, liquid height, and so forth. Interestingly enough, there are fairly reliable correlations for Newtonian and non-Newtonian fluids so long as the constraints of the correlations are observed.

It is important to note that some impellers that have high power numbers under unaerated conditions can “unload” considerably under aerated conditions. For example, it is fairly typical for a Rushton turbine to deliver, under aerated conditions, only 40% of the power it can deliver when unaerated. On the other hand, the SCABA 6SRGT (45), which is a curved-blade disc turbine, does not tend to unload at all. Some hydrofoils also exhibit very little unloading. This is an important consideration when one is trying to rank impellers in terms of how much power they can deliver for oxygen transfer purposes. (Often the ranking tends to be done qualitatively or “semiquantitatively.”)

Multimipeller Systems. Most fermenters are equipped with more than one impeller. The reasons for this are to improve bulk mixing (see later) and power distribution and to avoid the need for very large impellers and/or very high agitation speeds. Impeller size is limited practically not only by vessel internals but also by the following:

- Torque transmitted to the drive shaft. The larger the torque, the stronger and thicker the shaft must be.

This also translates to higher torque and more expensive gear boxes.

- The size of the vessel manway. This is particularly important if the impellers must be single piece for better cleaning/sterilization characteristics.

Agitator speed is limited by the natural frequency of the agitation system (4). This is because severe and potentially dangerous vibrations will occur if the rotational speed of the agitator approaches the natural frequency. We recommend that the maximum shaft speed not exceed 70% of the natural frequency.

Having said all this, we are left with the problem of doing the power calculations. This usually involves an iterative calculation in which the parameters are (assuming we have selected impeller types) number of impellers, impeller diameter(s), and shaft speed. (There are other considerations we discuss later.) To do this we use the power correlations already noted. But these usually apply only to single impellers. In most cases, they apply to the lowest impeller, which is the one under which inlet air is introduced. Several approaches have been suggested. The simplest is to treat all the impellers as though each is the only one present. Among other things, this assumes that the impellers do not interact with each other. There is empirical evidence (46pp.264–266) that Rushton turbines will not interact substantively in Newtonian broths if they are spaced at least an impeller diameter apart. Obviously, the assumption of noninteraction is not meaningful for impellers that produce significant axial flow (see later).

Another approach used frequently, but not recommended by the authors, is to use the correlations for the lowest impeller but to use them for the upper impellers:

\[
P_{g}/P_{ug} = (1 + \text{HU})^{-1}
\]

where HU is the gas holdup. You guessed it: there are empirical correlations for HU. One example for Newtonian broths is (47)

\[
\text{HU} = 1.8(P_gV_L)^{0.14}V_S^{0.75}
\]

The usual caveats apply.

Typical values for gas holdup range from about 0.1 to 0.3; therefore, equation 15 predicts about 10–20% unloading for the upper impellers. Aeration number correlations predict around 50–60% unloading. Given the uncertainty in all this, and assuming that no other (reliable) information is available, we suggest the more conservative approach of applying the aeration number correlation to each impeller independently.

The added cost for the larger drive will not be a major factor for vessels up to about 5,000 L; this is fairly cheap insurance. One should do pilot agitation studies at a meaningful scale for much larger vessels.

Organism Sensitivity to Fluid Mechanical Forces. Finally, an additional constraint must be imposed if the organism is sensitive to fluid mechanical forces. This is seldom true of unicellular organisms. There are, however, some mycelial organisms that are sensitive. The extent of sensitivity
and the nature of the forces that cause damage should be
determined experimentally. One also should keep in mind
that the character of fluid forces changes significantly with
scale. For example, some turbulent forces that are negli-
gible at small scale can be large and potentially destructive
at large scale. Reliable guidance in this area is almost non-
existent. There is, however, a rough rule of thumb that
states that damage will occur if impeller tip speed exceeds
1,500 ft/min. Bear in mind that this rule was developed
almost 50 years ago and was based on information for or-
ganisms used at that time in vessels that had working vol-
umes in the 30,000 to 50,000-gal range. It is clear that
many factors (including broth rheology) will affect the po-
tential for damage. Unfortunately, most of the information
is seldom available when needed. In such cases, the “any
port in a storm” approach to design usually becomes op-
erable. Fortunately, we seldom encounter a circumstance
in which the rule of thumb must be violated.

Bulk Mixing

Good bulk mixing is needed to insure homogeneity, which
is required for reliable data acquisition and control. There
are several important factors that affect mixing quality.

Broth rheology. This is the primary factor affecting our
ability to provide good mixing quality, but we seldom have
any control over it. We simply must recognize that we will
have to go to greater and greater lengths as the broth be-
comes more viscous and non-Newtonian. Also, one must
bear in mind that the nonlinear behavior of most non-
Newtonian fluids often results in nonobvious responses to
various actions. For example, increasing agitator speed in
a highly non-Newtonian, pseudoplastic broth often leads
to a decrease in power input and a decrease in bulk mixing
quality.

Type(s) of impeller(s). Impeller types fall into three basic
categories: radial flow (e.g., Rushton turbine), axial flow
(e.g., marine propellor), and mixed axial and radial flow
(e.g., Lightnin’ A-315). Radial flow impellers tend to deliver
the high power required to enhance micromixing and mass
transfer but do not promote top-to-bottom mixing; there-
fore, they are not the best choice for promoting homoge-
neity. Multiple radial flow impellers often are used in an
attempt to compensate for poor bulk mixing. This can work
fairly well for low-viscosity Newtonian broths when the ra-
tio of liquid height to tank diameter does not exceed 1.7–
2, but it is a very poor solution when the broth is very
viscous and highly non-Newtonian. Anyone who has ob-
served classical streptomycete or xanthan gum fermenta-
tions can attest to this fact. Pure axial flow impellers do
not deliver much power but do tend to promote good top-
to-bottom mixing and hence contribute significantly to
bulk homogeneity. They do not, however, contribute much
to mass transfer. Some mixed flow types appear to provide
a good balance between bulk mixing and mass transfer
requirements, particularly for non-Newtonian broths. We
recommend that they be considered seriously for such
cases. We also have had considerable success with combi-
nation systems, for example, those with turbines as the
lower impellers and a hydrofoil on top. This appears to
work well for Newtonian and non-Newtonian broths.

Impeller size(s). In most cases, large-diameter impellers
distribute power better and promote bulk mixing better
than do small-diameter impellers. As a rough rule of
thumb, based on considerable empirical evidence, we rec-
 recommend D/D1 ratios of between 0.4 and 0.5 unless some
other constraint (e.g., torque) is violated. Exceptions to this
are some cases in which an axial flow impeller is the top
impeller. In such cases, the ratio should not exceed 0.35 to
avoid the risk of vortexing. This is particularly true for low-
viscosity Newtonian broths.

Vessel geometry. We suggest that the ratio of the aerated
liquid height to the tank diameter be kept below 2.0 to
minimize bulk mixing problems. This must be reconciled
with impeller diameter and impeller spacing requirements
(see later) and with architectural, shipping, and other con-
straints.

Impeller spacing. As noted earlier, there is a body of
empirical evidence (48) that supports the spacing of Rush-
ton turbines 1–1.5 impeller diameters apart in Newtonian
broths. This spacing appears to provide a balance between
preventing impellers from interfering with their neighbors
and at the same time avoiding dead zones between them.

The same sources of empirical data support placing the
lowest impeller one impeller diameter from the bottom of
the vessel (assuming a standard dished head) and placing
the uppermost impeller one impeller diameter below the
liquid surface. We have found these rules to work reason-
ably well for Newtonian broths and a few non-Newtonian
broths, and for radial flow impellers other than Rushtons.
Please note that these conclusions are based on limited
observations. Finally, keep in mind that one cannot avoid
interaction among impellers when at least one of them is
an axial flow type or has a significant axial flow component,
and that spacing effects in highly non-Newtonian broths
are not well understood.

Baffling. Baffles are placed in vessels to minimize fluid
swirling and vortex formation (see Ref. 48 for flow pattern
diagrams). The usual approach is to use four baffles on 90°
centers, each baffle having a width equal to 0.1 of the D1.
Baffling tends to increase transmittable power and to im-
prove mixing (except for the dead spots, which tend to form
behind the baffles). Elimination of significant vortexing is
important for safety as well as for improved bulk mixing.
A vortex can reach down to the impeller in an unbaffled
vessel. When this happens, a sizable portion of the impeller
becomes air enshrouded, thereby decreasing resistance on
the impeller. This tends to drive up the impeller speed, at
least briefly (or the speed may be increased intentionally
in an attempt to increase the power input). Unfortunately,
vortices are unstable and can collapse shortly after for-
amination, resulting in very large and potentially cata-
strophic stresses (impulse) on the impeller, particularly if
the impeller speed is increased significantly while it is en-
shrouded.

Gas flow. Gas flow has a complex effect on bulk mixing.
The flow alone does tend to promote bulk mixing (as in
bubble tanks), but it also tends to decrease the effect of the
impellers, particularly at high values of gas linear velocity.
Fortunately, this is not a major problem in most cases, but
it can be for very large fermenters and for highly viscous
non-Newtonian broths.
Gas Holdup, Foaming, and Aerosol

There are several problems that can interfere seriously with a fermentation. Gas holdup (discussed earlier) decreases the effective volume of a fermenter. Foaming and aerosol formation can constrain operation, cause major cleaning and asepsis problems, and in the extreme cause termination of a fermentation. All three are dependent on broth characteristics, power input, and gas flow rate. There is some information in the literature, but these points have not been given the attention that even begins to reflect their importance. In general, all one can say is that all three become bigger problems for a given fermentation as gas flow and power input increase. Beyond that, each case is a new adventure. More will be said about the practical aspects later.

Heat Transfer

Heat Loads. Heat transfer is required during fermentation to maintain constant temperature conditions, and at other times (i.e., sterilization, induction) to increase or decrease broth temperature. In most cases, cooling is required during most of an active, aerobic fermentation. A good approximation to the total heat load, \( Q_{\text{tot}} \), during such fermentations in which the primary carbon source is glucose or a similar carbohydrate is (49)

\[
Q_{\text{tot}} = Q_{\text{metab}} + Q_{\text{mech}}
\]

\[
Q_{\text{tot}} = 0.48(\text{OTR})V_LY_{\text{xib}} + 2545P_g
\]  

(17)

where \( Q_{\text{metab}} \) is the heat generated by metabolism (Btu/h), \( Q_{\text{mech}} \) is the heat generated by power input (Btu/h), and \( Y_{\text{xib}} \) is the yield of cells on oxygen (dry wt cells per g oxygen consumed). Heat transfer can be more of a limiting factor in aerobic fermentations than is oxygen transfer—particularly in large fermenters. The major heat transfer demand is usually during the maximum growth period (i.e., when \( \mu X \) is greatest). It is always wise, however, to check other loads. For example, rapid cool down after an induction phase could put enormous demands on the cooling system. Cooling a 10,000-L (wv) fermenter from 38 °C (104 °F) to 10 °C (50 °F) in 30 min would require an average heat transfer rate of approximately \( 2.4 \times 10^6 \) Btu/h. The same heat transfer rate would support an oxygen transfer rate of approximately \( 495 \text{ mmol/L(h)} \), which is very high by almost anyone’s standards. Similarly, one should check other potentially high loads such as cool down after sterilization (see later). In any case, careful thought and solid, empirical facts (e.g., product thermal degradation rates) should form the basis of any specification that will require extraordinary transfer rates.

Heat Transfer Rate. The rate at which heat can be transferred is dependent on (1) the driving force for heat transfer, (2) the area across which transfer must occur and, (3) the resistance to heat transfer:

\[
Q = (1/\text{resistance}) \times (\text{area}) \times (\text{driving force})
\]  

(18)

The driving force for agitated vessels is usually taken to be the log mean temperature difference, \( \Delta T_{\text{LM}} \), defined as

\[
\Delta T_{\text{LM}} = (T_{\text{co}} - T_{\text{ci}})\ln[(T_{\text{f}} - T_{\text{ci}})/(T_{\text{f}} - T_{\text{co}})]
\]  

(19)

where \( T_{\text{co}} \) is the coolant outlet temperature (°F), \( T_{\text{ci}} \) is the coolant inlet temperature (°F), and \( T_{\text{f}} \) is the fermentation temperature (°F). Note that use of equation 19 assumes that the broth is homogeneous and can be characterized by a single temperature.

The transfer area, \( A_{\text{ch}} \), is the actual contact area (not necessarily the same as the available area) between the broth and the heat exchange surface, which is usually a jacket and/or an internal coil. The actual area can be found if the height of aerated liquid in the tank is known, along with the geometry of the jacket and any internal exchange area (e.g., a coil). The aerated liquid height measured along the centerline of the vessel, \( h_{\text{LA}} \) (ft), is

\[
h_{\text{LA}} = [(1 + HU)V_{\text{LU}} - V_D]/(0.785D^2) + \text{IDD}
\]  

(20)

where \( V_{\text{LU}} \) is the unaerated liquid volume (ft³), \( V_D \) is the lower dish volume (ft³), \( \text{IDD} \) is the inside depth of lower dish (ft), and \( D_i \) is the tank inside diameter (ft).

The resistance to heat transfer is a little more complicated. It is composed of five resistances in series and is expressed as

\[
1/U = 1/h_i + 1/h_o + 1/h_{\text{fi}} + 1/h_{\text{fo}} + t/k
\]  

(21)

where \( U \) is the overall heat transfer coefficient (Btu/h·°F ft²), \( h_i \) is the broth film heat transfer coefficient (Btu/h·°F ft²), \( h_o \) is the jacket fluid heat transfer coefficient (Btu/h·°F ft²), \( h_{\text{fi}} \) is the broth fouling factor (Btu/h·°F ft²), \( h_{\text{fo}} \) is the jacket fluid fouling factor (Btu/h·°F ft²), \( t \) is the fermenter wall thickness (ft), and \( k \) is the fermenter wall thermal conductivity (Btu/h·°F ft). The resistance of the tank wall can be predicted accurately because both \( k \) and wall thickness are known accurately. It should be noted that the \( k \) for stainless steel is very low (about 10 Btu/°F ft ft). This means that the tank wall should be kept as thin as possible to minimize its contribution to heat transfer resistance (see later).

Correlations for \( h_o \) can be found in the literature (46pp. 282–283) but are seldom as reliable as those available from vendors for the specific jackets they fabricate. The problem in predicting \( U \) is in predicting reliable values of \( h_i \), \( h_{\text{fi}} \), and \( h_{\text{fo}} \). Published correlations (50) for \( h_i \) values for aerated, Newtonian fermentation broths have been published but are not very reliable. Correlations for non-Newtonian broths have also been published (51), but as was the case for mass transfer coefficients, \( h_i \) is strongly dependent on broth rheological characteristics. Unfortunately, existing correlations do not do a very good job of accounting adequately for broth rheology, a problem exacerbated by the fact that good rheological information is seldom available. Finally, the fouling factors are essentially impossible to predict. The best one can do is keep fouling to a minimum by means of good cleaning practices and proper filtration of coolants. Effects of individual resistances on \( U \) are illustrated in Figure 2.

The following rough guides (based on many years of experience) are offered because the authors are unable to offer any collection of reliable prediction tools:
It also should be noted that $U$’s for heating usually are a bit higher than $U$’s for cooling. The heat transfer rate can now be expressed as

$$Q = UAT_{LM} \quad (22)$$

Similar equations can be developed for heat transfer via an internal coil. The only difference is that the heat transfer coefficients for coils are usually a bit higher than those for jackets. That having been said, we will try everything possible to discourage you from using internal coils. The reasons relate primarily to mechanical and cleaning considerations and are discussed later.

In applying the preceding, we suggest that the reader consider the following suggestions:

- Avoid the use of internal cooling surfaces (e.g., coils). They make cleaning difficult to impossible, are potential sources of contamination (leaking of coolant), put a lot of mechanical strain on the vessel, and can cause deterioration of bulk mixing (particularly for viscous non-Newtonian broths).
- Avoid subfreezing coolant. Sooner or later, subfreezing coolant will cause severe valve freeze-up and worse.
- Try to keep coolant flow rates low enough to avoid pipe diameters greater than 3 in.; cost increases considerably and availability of fittings and so forth decreases as size increases.

Finally, there are some cases in which high wall temperatures have been cited as causing problems with temperature-sensitive organisms. The wall temperature, $T_w$, depends on the temperatures of the broth and of the jacket fluid, as well as on the relative values of the heat transfer resistances. A simple energy balance across the heat transfer path yields

$$T_w = \frac{(T_f + \beta T_u)}{(1 + \beta)} \quad (23)$$

where

$$\beta = \frac{k}{h_u(t + k/h_u)} \quad (24)$$

Note that this does not account for locally high wall temperatures that would be caused by steam applied to steam locks during fermentation.

**Sterilization**

**Introduction.** Sterilization and aseptic operation taken together (as they must be) have a much greater influence on fermenter design and operation than does any other requirement.

In theory (1) sterilization destroys or removes all foreign organisms in all process equipment (including piping, seals, etc.) that might come into contact with the process fluid, and (2) aseptic operation insures that no contaminating organisms enter the fermenter after sterilization. These ideals are not attainable in practice because (among other reasons) (1) there is a finite probability that a very low concentration of contaminating organisms will not be captured in a sample used to test for contamination, and (2) there is a finite probability of false positives due to sample contamination and so forth. It also is important to note that there is a continuing debate concerning the definitions of pure culture and sterility (6,51).

The driving forces for sterilization and aseptic operation range from minimizing product losses to insuring strict compliance with regulatory requirements. Among the many specific problems that contamination can cause are the following:

- Production of a toxin that can’t be removed by the purification system
- Production of an enzyme that degrades the product
- Decreased product yield due to use of substrate by contaminants
- Production of toxins that inhibit the producer strain
- Production of compounds (e.g., polysaccharides) that interfere with the operation of recovery and purification equipment

Which specific problems will exist and where in the spectrum a particular case will lie depend primarily on the product, its economic value, whether it is regulated, and how it will be used. Given all these variables and the fact that absolute sterility is an unachievable abstraction, the extents to which one should go should be considered on a case-by-case basis. As a practical matter, sterilization has to be interpreted as “effective sterilization,” meaning that the design and procedures (including sampling and detection) are suitable for the specific case. For example, sterilization of fermenters used to produce parenterals should be held to a much higher standard than sterilization of
fermenters used to produce amylases for starch hydrolysis. Unfortunately, there remains a lot of room for disagreement among well-intentioned people (see also the section “Product Safety”).

Quantification of Sterilization. Mechanical details and other practical considerations of sterilization and aseptic operation are discussed in the subsection “Sterile Piping Systems.”

Several methods have been developed to quantify sterilization. The simplest is based on the empirical observation that the death kinetics of many vegetative organisms and spores can be described by a simple first-order expression

$$\frac{dN}{dt} = -kN \quad (25)$$

where \(N\) is the number of viable organisms (or spores) at any time, \(t\), and \(k\) is the thermal death constant. \(k\) is a function of most environmental variables but is most strongly affected by temperature. This dependence is very often expressible as an Arrhenius relationship

$$k = A \exp(-E_A/RT) \quad (26)$$

where \(A\) is the preexponential factor (\(\text{min}^{-1}\)), \(E_A\) is the thermal death activation energy (cal/mol), \(T\) is the absolute temperature (K), and \(R\) is the gas constant. The effect of temperature on \(k\) is pronounced for most organisms because their \(E_A\)’s are so high (e.g., 65,000 cal/mol for Bacillus stearothermophilus spores).

Among other methods of quantifying sterilization concepts are the following:

- Probability-based theories (52). These are conceptually different from the first-order model, but give essentially the same results for the levels of kill that must be achieved practically.
- Methods based on the “decimal reduction factor,” \(D\). These are used widely (53). It is important to note, however, that the nature of temperature dependence has no rational, physical basis and differs considerably from an Arrhenius relationship: it is based primarily on an empirical observation for relatively small temperature ranges.

We use the first-order model for the rest of the discussion.

Medium Sterilization. The primary design basis for sterilization systems and procedures is usually the level of sterility that must be achieved. This is most frequently expressed in terms of the probability of sterilization failure (i.e., the probability that a single organism will survive). In most instances probabilities of \(10^{-3}\) to \(10^{-4}\) are quite adequate. This means 1 failure in 1,000 or 10,000 sterilization operations, respectively. It also is important to note that the probability of failure is numerically very close to the value of \(N\) in the first-order model.

The reader should also take note of the fact that in some quarters, sterilization criteria are stated in terms of “logs of kill.” That is to say, a fixed number of logs (e.g., 12) is taken to be adequate. This is meaningless in that it does not take into account the initial contaminant loading.

Now we consider prediction of requirements for batch sterilization of a fermentation medium as part of the fermenter design process to ensure that we’ll have adequate heating and cooling capacity. It should also be done as part of the process design to ensure adequate timing. Batch sterilization involves heating the medium in the fermenter to sterilization temperature (\(T_{\text{ster}}\)), holding the medium temperature constant at \(T_{\text{ster}}\) for an adequate time, and then cooling the medium down to fermentation temperature (see the section “Mechanical Design” for additional details). The medium is kept in the fermenter throughout.

It is assumed in what follows that (1) heating and cooling are via the jacket only, and (2) the energy and condensate contributed by the small amount of steam that flows into the vessel as a result of sterilizing piping, air filters, and so on is negligible for energy balance purposes.

The rate of kill at any instant during the sterilization process is

$$\frac{dN}{dt} = -kN = -A \exp(-E_A/RT)N \quad (27)$$

therefore,

$$\frac{N}{N_o} = A \int \exp(-E_A/RT)dt \quad (28)$$

where \(N_o\) is the number of contaminating organisms initially present. The integration of equation 26 must be performed over heating, hold, and cooling phases of sterilization. This is done by numerical integration of equation 28, coupled with the equations that describe broth temperature as a function of time.

For heating

$$T_f = T_{ST} - (T_{ST} - T_{F,0}) \exp(-U * A_{jA} * t/(M_F * C_{P,m})) \quad (29)$$

where \(T_f\) is the medium temperature (°F), \(T_{ST}\) is the steam temperature (°F), \(T_{F,0}\) is the initial medium temperature (°F), \(A_{jA}\) is the active jacket area (ft²), \(t\) is the time since beginning of heating (h), \(M_F\) is the medium mass (lb), and \(C_{P,m}\) is the medium heat capacity (Btu/(lb °F)).

For cooling

$$T_f = T_{Cj} + (T_{ster} - T_{Cj}) \exp(t (W_f/C_{P,c}/(M_F/C_{P,m})))$$

$$\cdot (\exp(UA_{jA}/(W_fC_{P,c}) - 1) \quad (30)$$

where \(T_{Cj}\) is the inlet coolant temperature (°F), \(T_{ster}\) is the sterilization temperature (°F), \(t\) is the time from beginning of cooling (h), \(W_f\) is the coolant flow (lb/h), and \(C_{P,c}\) is the coolant heat capacity (Btu/(lb °F)). Equations 29 and 30 can be derived via energy balances.

One chooses times and temperatures such that the desired kill, \(N/N_o\), is achieved. Unfortunately, one usually does not know the level of contamination or the nature of the contaminants. Indeed, it is very unlikely that contamination will remain constant from one fermentation to the
next. The usual practice, therefore, is to design for what is thought to be the “worst case.” This usually gives results that are quite satisfactory for design specifications. A “typical” worst case is a loading of \( 5 \times 10^6 \) spores/mL of the organism B. stearothermophilus. Such spores are highly resistant and have the added advantage of being used widely to validate sterilization operations (see later). It is important to note that the purpose of the calculations is to provide a rational basis for obtaining reasonable design information and ensuring a high probability that the desired level of sterilization can be achieved over the whole range of anticipated operating conditions.

Figure 3 illustrates the results of such calculations for the following case:

- Medium volume = 15,000 L
- Initial spore load = \( 5 \times 10^6 \) spores/mL
- Initial medium temperature = 70 °F
- Steam temperature = 300 °F
- Coolant temperature = 35 °F
- Coolant flow rate = 100 gpm
- Fermentation temperature = 86 °F
- Active jacket area = 225 ft²
- \( U = 150 \text{ BTU/(°F h ft}^2) \)
- \( A = 2.1 \times 10^{36}/\text{min} \)
- \( E_A = 65,000 \text{ cal/mol} \)
- \( N \) (probability of failure) = 0.0001

As seen in this example, heating and cooling phases usually contribute only a small fraction of the total sterilization kill, but they do affect turnaround time significantly. Long heating and cooling times can also have other effects:

- Cause damage to the fermentation medium to the extent that the fermentation can be compromised. The seriousness of this depends on some combination of economics and regulatory compliance.
- Cause medium changes that have negative effects on recovery and purification without causing fermentation problems. This includes the possibility of introducing foreign materials that may pass undetected into the final product.

The example also shows that cooling time is considerably longer than heating time. This results primarily from the lower temperature-driving forces during cooling. The problem is exacerbated considerably when the heat transfer coefficient is very low as is the case for viscous non-Newtonian broths. Another method used for batch sterilization is direct steam injection: live steam is injected directly into the medium as the main source of thermal energy. This decreases heating time as well as the overall steam requirement. It also increases the medium volume by about 20% (as a result of steam condensation), which can cause some problems, including the following:

- Increased cooling time.
- Medium dilution. This will be a significant problem if the initial medium cannot be made concentrated enough to account for the dilution. Some reasons include low solubility of medium components, increased viscosity, and increased reaction rates among medium components at elevated temperatures.
- Introduction of impurities. This depends primarily on the quality of the injected steam. In some cases plant steam is acceptable if boiler cleaning agents do not cause problems. At the other extreme is the requirement to use clean steam (WFI quality). It should be noted with regard to this point that some steam will be injected directly even when jacket heating is used as the main energy source; therefore, one will always be in the position of having to evaluate the effects of contaminants carried by the steam.

For cases in which sterilization times are too long or medium alterations cause too many problems, one might consider continuous sterilization. The interested reader is directed to the literature for additional information (54pp. 159–176).

Piping System Sterilization. Heating and cooling times are not usually significant issues for sterilizing piping; however, there can be some serious heat transfer problems related to piping length, diameter, and orientation. Some of these are discussed in the subsection “Vessel Design.” The reader is also referred to literature reports of some practical experiments and theoretical analyses (55,56).

Air Sterilization. Theoretical aspects of air sterilization are not discussed here. Suffice it to say that modern filters, if sized, installed, and maintained properly, will provide very reliable results. We’ll discuss some of these practical aspects in the section “Mechanical Design.” See Refs. 54pp. 176–183 and 57 for additional information.

Cleaning

Scrupulous cleaning is necessary to decrease nonbiological contamination and prevent cross-contamination of batches. Experience also has shown that reliable sterilization is difficult or impossible to achieve in the absence of rigorous cleaning.

As noted earlier, fermenter cleaning is not mentioned specifically in 21 CFR, but the basis is provided in section
211.67, which requires equipment be cleaned as per a defined plan and by trained people. Furthermore, the FDA has made clear the importance it attaches to cleaning not only in various agency publications (58), but also in more stringent enforcement. This is particularly true in multi-product facilities (a key issue for most modern biotech plants). This agency initiative and the desire to generally improve operation have caused the industry to develop more reliable cleaning systems and protocols and to move to fully automated cleaning-in-place (CIP) systems, which provide greater assurance of successful validation and long-term compliance.

Developing and designing reliable fermenter cleaning systems is not as straightforward as it might appear, and considerable controversy continues. Among the reasons for this is that while much is known about the basic science of cleaning in general, very little is known about the basic science of cleaning fermenters and little has been published. The approach taken is based primarily on experience derived from the dairy, food, and beverage industries. Such information is useful but is not directly applicable in general to pharmaceutical and biotech processes where soils are different and the cleaning requirements are far more stringent. What meager information there is concerning pharmaceutical and biotech soils has been obtained from experiments done on single soil components and/or studies done under conditions not truly representative of the process conditions (59,60). There have been no significant attempts to develop systematic analyses based on experiments done with complex mixtures typical of real fermentation soils under conditions found in real processes. As a result there are no reliable general methods, tools, or correlations on which to base the development of cleaning agents, protocols and CIP system design; decisions tend to be made based on arbitrary criteria (e.g., coupon bake on studies). In keeping with our general design philosophy, we think it important that such arbitrariness be avoided and that cleaning protocols be developed along with the fermentation. It is clear that this will help to ensure not only proper design but will also minimize cleaning validation studies and any questions concerning the presence of contaminants and/or cleaning residues in the commercial product that could not have been present in clinical trial material.

The selection of fermenter cleaning agents and protocols should be based not only on the specific soil but also on the materials of construction, surface finishes, and so on. In addition, one should consider the issues of (1) compatibility of each material with the cleaning agents and protocols, and (2) potential materials interaction during cleaning. These are decisions that usually are made during the design phase, but really should be defined much earlier. These issues are discussed in the subsection “Agitation Systems.”

Fermenter cleaning validation is beyond the scope of this article; however, the design must be done so as to minimize validation problems and ensure maximum ease in ongoing compliance. The design team must therefore be aware of the procedures that will be used for validation and recognize that this is an issue that has not yet been resolved in the general community.

Putting It Together: Preliminary Design Calculations

We now begin to see some of the interactions that can arise among the transport processes, and between them and the operating variables. This is a necessary step in developing sound designs, and a precursor to preliminary design calculations. As an example, consider a case in which we are trying to find ways to increase the OTR above some base value. One approach would be to increase the gas flow rate \( \nu \) keeping other variables constant:

- The overall driving force tends to increase (equation 10).
- \( V_\text{g} \) increases (equation 7).
- Increased \( V_\text{g} \) tends to increase \( K_a \) (equation 11).
- Increased \( \nu \) also increases \( N_A \) (equation 12).
- Increased \( N_A \) tends to decrease power input (Fig. 1).
- Decreased power input tends to decrease \( K_a \) (equation 11).

As another example, consider increasing the vessel pressure while keeping other variables constant:

- The overall driving force tends to increase (equation 10).
- The actual gas volumetric flow rate \( \nu \) tends to decrease (assuming constant molar flow rate) (equation 6).
- \( V_\text{g} \) tends to decrease (equation 7).
- The decrease in \( V_\text{g} \) tends to decrease \( K_a \) (equation 11).
- The decrease in \( \nu \) also tends to decrease \( N_A \) (equation 12).
- This tends to increase power input (Fig. 1).
- Increased power input tends to increase \( K_a \) (equation 11).

Such complex interrelationships develop regardless of what steps are taken to increase OTR. The nature of these will change depending on the approach taken and the system being considered. There will also be additional pluses and minuses related to factors such as cost, safety, cleaning, and regulatory compliance. For example, using pure oxygen to increase the driving force carries with it very distinct cost and safety problems. Many of these more practical implications are be discussed in the section “Mechanical Design.”

Now let’s carry the second approach a little further. We do this because increasing pressure is usually the easiest and seemingly least expensive way to increase OTR. The decrease in \( V_\text{g} \) caused by the increased pressure has the positive effects of decreasing foaming, aerosol formation, and gas holdup. But it has the negative effect of requiring thicker tank walls, thereby increasing heat transfer resistance in the face of an increased heat load due to increased metabolic activity. This will become a significant problem when the pressure desired is greater than that required for vessel sterilization. The larger the tank diameter, the greater the possible problem. It should also be noted that thicker walls can make it more difficult to get a good surface finish and therefore increases the possibility of cleaning problems. In addition, the higher the pressure and thicker the walls, the shorter the list of qualified vendors.
This type of reasoning process is applicable to any of the many permutations possible. We extend it after we consider some additional basic concepts, mechanical design practices, and a few more important practical constraints.

We now consider two examples of preliminary design calculations, the objectives of which are to, first, define vessel geometry to ensure that it will not violate architectural constraints and be able to satisfy general rules of thumb for good bulk mixing, impeller spacing, and so forth. Some of this will require checking calculations done later (see later) and may require some iteration.

The second objective is to establish ranges of air flow, pressure, oxygen enrichment, and power that will accomplish the following:

- Satisfy the design basis maximum OTR requirement.
- Keep $V_s$ low enough to avoid foaming, holdup, and aerosol problems. We have found a limit of 200 cm/min to be a safe guide; however, there are cases (Newtonian broths) in which considerably higher values up to 300–350 cm/min tolerable and others (very viscous, highly non-Newtonian broths) in which it is not wise to go beyond 75–100 cm/min.
- Not violate compressor and other component constraints.
- Not require heroic efforts (e.g., extremes in pressure, power, etc.).

Please note that we are trying to establish reasonably broad ranges for each of the variables within which we can satisfy the requirements with a reasonable degree of comfort (based on experience). This is important because the correlations usually available have questionable accuracy, and it allows some breathing room for process improvements and so forth.

The third objective is to define impeller type(s), sizes, and speed that will do the following:

- Provide the power required
- Provide good bulk mixing
- Satisfy impeller spacing rules of thumb
- Not violate constraints related to the sensitivity of the organism
- Make possible single-piece impeller construction and so forth (see the section "Mechanical Design" for consideration of mechanical constraints)
- Not violate any other constraints specified

The fourth objective is to define heat exchange area and coolant flow and temperature that achieve the following:

- Satisfy peak heat transfer requirements
- Not violate coolant temperature and flow constraints and such
- Avoid internal heat exchange area
- Not require subfreezing coolant or flow that will require very large pipe diameters

(Note: One should check to make sure that other heat loads [e.g., cooling after fermentation] are not greater than the load during maximum growth activity. We assume that to be the case here.)

The fifth objective is to use the results of the calculations not only to define mechanical details (see the section "Mechanical Design"), but also as a basis for recommendations for rational changes in design basis.

Please note that experimentally derived correlations for the subject cases are used in the examples (this is a design gift as rare as hen’s teeth). All were obtained at 500-L scale, and we are just within our comfort range concerning their scalability to 15,000 L (design scale). The reader is cautioned, however, that there is no basis to believe that the correlations would be applicable to other cases, nor would we recommend extrapolation to much larger scales. It also is worth noting that there are some good examples in the literature that might be applicable to the problems at hand or to others; however, one must exercise considerable caution if the conditions specified in the reference differ significantly from those for the case being analyzed. Unfortunately, it is often not possible to determine if significant differences exist.

Case 1.

<table>
<thead>
<tr>
<th>Product</th>
<th>Intracellular protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism</td>
<td>E. coli</td>
</tr>
<tr>
<td>Final broth volume</td>
<td>15,000 L</td>
</tr>
<tr>
<td>Maximum OTR</td>
<td>234 mmol/(L h)</td>
</tr>
<tr>
<td>Dissolved oxygen requirement</td>
<td>10% relative to atmospheric conditions</td>
</tr>
<tr>
<td>Broth rheological type</td>
<td>Newtonian</td>
</tr>
<tr>
<td>Maximum broth viscosity</td>
<td>5 cp</td>
</tr>
<tr>
<td>Regulation</td>
<td>CBER</td>
</tr>
<tr>
<td>Containment</td>
<td>BL2-LS</td>
</tr>
<tr>
<td>Maximum coolant flow</td>
<td>250 gpm</td>
</tr>
<tr>
<td>Minimum fluid temperature</td>
<td>35 °F</td>
</tr>
<tr>
<td>Internal coils</td>
<td>No</td>
</tr>
<tr>
<td>Compressor max press/max flow</td>
<td>50 psig/25,000 SLPM</td>
</tr>
<tr>
<td>Minimum heat transfer coefficient</td>
<td>50 Btu/(h °F ft²)</td>
</tr>
<tr>
<td>Ceiling height</td>
<td>336 in.</td>
</tr>
<tr>
<td>Floor space</td>
<td>1000 ft², unobstructed</td>
</tr>
</tbody>
</table>

The following experimental correlations are applicable:

\[ K_f a = 108(P_g/V_L)^{0.6}V_L^{0.2} \]

\[ P_g/P_{ug} = 1.0 - 18.56N_A + 215.64N_A^2 - 1082.66N_A^3 + 1859.68N_A^4 \]

\[ HU = 0.0133(P_g/V_L)^{0.4}(60V_L^{0.5}) - 0.0333 \]

These were developed under the following conditions for single Rushton turbines:

\[ 0.8 < P_g/V_L < 2.9 \text{ hp/100 gal} \]
\[ 100 < V_S < 200 \text{ cm/min} \]
\[ 0 < N_A < 0.1 \]
\[ 0.33 < D_i/D_t < 0.45 \]
Vessel geometry calculations are the same for both cases. We start by assuming a vessel diameter and wall thickness. (Note that the thickness will probably be changed a bit later to satisfy code requirements. This will not have a significant effect on height and volume calculations to follow.) In this case we assume a 96-in. outside diameter and a 0.25-in. wall thickness. By straightforward geometry and using tabulated values of volumes and inside depth of dish (IDD) for standard ASME heads, we get the following:

- Liquid height: 134.2 in.
- (Liquid height)/tank diameter: 1.41
- Total tank height: 192.5 in.
- Total volume: 21.1 L
- Percent fill (unaerated): 71.1%

This satisfies rules of thumb concerning liquid height-to-diameter ratio and fill percent. If we add about 72 in. for legs and 30 in. for opening the manway, we get a total height requirement of 294.5 in. We should be able to satisfy the maximum height limit.

The results of oxygen transfer calculations for several operating conditions are given in Table 1. From these results we conclude that the following will satisfy the OTR required and give us a reasonable amount of “breathing room.”

- Gassed power: 100 hp (includes transmission efficiency)
- Maximum operating pressure: 40 psig
- Maximum air flow: 10,000 SLPM

We do not recommend oxygen enrichment at this point, but we do suggest that the “holes be punched” to facilitate future installation of an oxygen system. The agitation system calculations are based on the assumption of using two 40-in. diameter Rushton turbines. The results are given in Table 2. A maximum speed of 150 rpm should be adequate. Note also that the impellers can be spaced to satisfy the rules of thumb for bulk mixing rules discussed previously.

Although this should give satisfactory results for such a low viscosity broth, we recommend consideration of a slightly larger turbine as the lower impeller and a hydrofoil (e.g., A-315) as the top impeller.

The results of the heat transfer calculations are given in Table 3. A minimum coolant temperature of 35 °F was used in all cases. The results show that the heat loads can be handled, but the conditions are uncomfortably tight. For example, in no case will increasing the coolant flow to the maximum make up for a 10% decrease of the heat transfer coefficient. The potential consequences of inadequate heat transfer in this case should be considered seriously before final design commitments are made. Alternatives (e.g., lower coolant temperature) should be explored.

### Case 2.

<table>
<thead>
<tr>
<th>Product</th>
<th>Antibiotic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism</td>
<td>Streptomyces sp</td>
</tr>
<tr>
<td>Final broth volume</td>
<td>15,000 L</td>
</tr>
<tr>
<td>Maximum OTR</td>
<td>50 mmol/(L h)</td>
</tr>
<tr>
<td>Dissolved oxygen requirement</td>
<td>10% relative to atmospheric conditions</td>
</tr>
<tr>
<td>Broth rheological type</td>
<td>Non-Newtonian (pseudoplastic)</td>
</tr>
<tr>
<td>Maximum broth viscosity</td>
<td>1,000-1,500 cp</td>
</tr>
<tr>
<td>Regulation</td>
<td>CDER</td>
</tr>
<tr>
<td>Containment</td>
<td>BL1-LS</td>
</tr>
<tr>
<td>Max coolant flow</td>
<td>250 gpm</td>
</tr>
<tr>
<td>Min fluid temperature</td>
<td>35 °F</td>
</tr>
<tr>
<td>Internal coils</td>
<td>No</td>
</tr>
<tr>
<td>Compressor max press/max flow</td>
<td>50 psig/25,000 SLPM</td>
</tr>
<tr>
<td>Min heat transfer coefficient</td>
<td>40 btu/(h °F ft²)</td>
</tr>
<tr>
<td>Ceiling height</td>
<td>336 in.</td>
</tr>
<tr>
<td>Floor space</td>
<td>1000 ft², unobstructed</td>
</tr>
</tbody>
</table>

The following empirical equations are applicable:

\[
K_{a}A = 8.0(P_{g}\nu_{L})^{0.2}V_{S}^{0.4}
\]

\[
P_{g}/P_{ug} = 1.0 - 26.99N_{A} + 417.37N_{A}^{2} - 2789.43N_{A}^{3} + 6643.30N_{A}^{4}
\]

### Table 1. Oxygen Transfer Calculation Results (Single Cell)

<table>
<thead>
<tr>
<th>Pressure (PSIG)</th>
<th>35</th>
<th>20</th>
<th>11,719</th>
</tr>
</thead>
<tbody>
<tr>
<td>O₂ trans eff. (%)</td>
<td>21</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>O₂ (MOL % in)</td>
<td>21</td>
<td>21</td>
<td>40</td>
</tr>
<tr>
<td>Gas flow (in, SLPM)</td>
<td>29,762</td>
<td>20,833</td>
<td>11,719</td>
</tr>
<tr>
<td>VVM</td>
<td>1.98</td>
<td>1.39</td>
<td>0.78</td>
</tr>
<tr>
<td>Average V₉ (cm/min)</td>
<td>186</td>
<td>183</td>
<td>103</td>
</tr>
<tr>
<td>hp/100 gal</td>
<td>1.3</td>
<td>2.7</td>
<td>1.0</td>
</tr>
<tr>
<td>Theor. P₉</td>
<td>52.8</td>
<td>105.6</td>
<td>41.1</td>
</tr>
<tr>
<td></td>
<td>103.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Citation System Calculation Results (Single Cell)

<table>
<thead>
<tr>
<th>Shaft speed (rpm)</th>
<th>114</th>
<th>141</th>
<th>88</th>
<th>141</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turb imp dia (in)</td>
<td>40.00</td>
<td>40.00</td>
<td>40.00</td>
<td>40.00</td>
</tr>
<tr>
<td>Turb D/Dₗ ratio</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>Ungassed hp (turb)</td>
<td>123.1</td>
<td>233.0</td>
<td>56.6</td>
<td>233.0</td>
</tr>
<tr>
<td>Nₐ, aeration numb</td>
<td>0.071</td>
<td>0.057</td>
<td>0.051</td>
<td>0.062</td>
</tr>
<tr>
<td>P₉/P₉ ratio</td>
<td>0.43</td>
<td>0.46</td>
<td>0.48</td>
<td>0.45</td>
</tr>
<tr>
<td>Gassed hp</td>
<td>53</td>
<td>106</td>
<td>42</td>
<td>104</td>
</tr>
</tbody>
</table>
Table 3. Heat Transfer Calculation Results (Single Cell)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat load (MMBtu/h)</td>
<td>1.808</td>
<td>1.943</td>
<td>1.779</td>
<td>1.938</td>
</tr>
<tr>
<td>Hold-up</td>
<td>0.199</td>
<td>0.271</td>
<td>0.123</td>
<td>0.282</td>
</tr>
<tr>
<td>Jacket area (ft²)</td>
<td>297</td>
<td>317</td>
<td>276</td>
<td>320</td>
</tr>
<tr>
<td>UJ KT (BTU/ft² h °F)</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>Coolant temp in °F</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Coolant flow (GPM)</td>
<td>190</td>
<td>210</td>
<td>235</td>
<td>200</td>
</tr>
</tbody>
</table>

\[
H_U = 0.012(P_gV_L)^{0.3}(60V_S/100)^{0.52} - 0.029
\]

These were developed under the following conditions for single Rushton turbines:

1.0 \(< P_g/V_L < 2.3 \text{ hp/100 gal}
60 < V_S < 130 \text{ cm/min}
0 < N_A < 0.1
0.36 < D_i/D_t < 0.42

The results of oxygen transfer calculations for several operating conditions are given in Table 4. From these results we conclude that the following will satisfy the OTR required:

Gassed power                      | 100 hp (includes transmission efficiency) |
Maximum operating pressure        | 45 psig                                   |
Maximum air flow                  | 25,000 SLPM                               |
Oxygen enrichment                 | 40% oxygen in inlet stream                |

Oxygen enrichment should be considered in this case despite the fact that the calculations show that it is not necessary. The reason is that the gas flow rate, and hence V_S, is quite high without enrichment. This results primarily from the very low oxygen transfer efficiency typical of very viscous non-Newtonian broths. Although it is true that we haven't broken the rule of thumb for holdup, foaming, and aerosol we should take note of the fact that high air flows in pseudoplastic broths can exacerbate impeller unloading and cause flooding more readily than for non-Newtonian broths. It also is generally true that our correlations become less reliable as conditions become more extreme and further removed from the conditions under which they were developed. Note in particular that our correlations were not developed for the high values of V_S found in the calculations. It also is clear that we should specify higher pressures than shown in the conditions under which they were developed. Note in particular that our correlations were not developed for the high values of V_S found in the calculations. Additional calculations show that 40–45 psig would bring the V_S into a comfortable range. We suggest, therefore, the following:

Gassed power                      | 100 hp (includes transmission efficiency) |
Maximum operating pressure        | 45 psig                                   |
Maximum air flow                  | 25,000 SLPM                               |
Oxygen enrichment                 | 40% oxygen in inlet stream                |

The agitation system calculations are based on the assumption of using two 40-in.-diameter Rushton turbines. The results are given in Table 5. A maximum speed of 140 rpm appears to be adequate. Note also that the impellers can be spaced to satisfy the bulk mixing rules of thumb discussed previously.

That's very nice, but we have significant reservations about the results. An alternative should be given serious consideration because of the rheological nature of the broth, the profound effect this can have on mixing, and the issues of scalability for such broths. At the very least, the lower turbine should be replaced by one having a D_i/D_t approaching 0.5, and the upper turbine should be replaced by a hydrofoil. The reader should understand that this is based on experience only: we have no solid correlations for this case. The best route to take would be to do a few pilot mixing studies in as large a vessel as possible with the real broth.

The results of the heat transfer calculations are given in Table 6. The minimum coolant temperature of 35 °F was used in all cases. The results show that the heat loads can be handled. It also appears that we are well below the maximum available coolant flow. This may be comforting, but the reader should realize that there is not a lot of breathing room for process improvement. In particular, an increase in OTR up to 60–65 mmol/(L h) would put us over the top even if there were no deterioration in heat transfer coefficient. Again, one should give this serious consideration before committing to a final design.

We trust that the discussion and examples in this section have helped the reader to understand not only the many interactions that must be considered, but also the nature of preliminary design calculations and their limi-

Table 4. Oxygen Transfer Calculation Results (Mycelial)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure (PSIG)</td>
<td>35</td>
<td>25</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Oxygen transfer eff (%)</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>O₂ mol % in</td>
<td>21</td>
<td>21</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Gas flow IN (SLPM)</td>
<td>22,222</td>
<td>22,222</td>
<td>11,667</td>
<td>11,667</td>
</tr>
<tr>
<td>VVM</td>
<td>1.48</td>
<td>1.48</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>Average V_S (cm/min)</td>
<td>154</td>
<td>172</td>
<td>102</td>
<td>140</td>
</tr>
<tr>
<td>hp/100 gal required</td>
<td>1.7</td>
<td>2.0</td>
<td>1.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Theoretical Pg</td>
<td>60.9</td>
<td>79.9</td>
<td>41.7</td>
<td>61.8</td>
</tr>
</tbody>
</table>
Table 5. Agitation System Calculation Results (Mycelial)

<table>
<thead>
<tr>
<th>Shaft speed (rpm)</th>
<th>Ungassed hp</th>
<th>Aeration number</th>
<th>Gassed hp</th>
</tr>
</thead>
<tbody>
<tr>
<td>124</td>
<td>159</td>
<td>0.049</td>
<td>61</td>
</tr>
<tr>
<td>137</td>
<td>214</td>
<td>0.055</td>
<td>80</td>
</tr>
<tr>
<td>107</td>
<td>102</td>
<td>0.042</td>
<td>42</td>
</tr>
<tr>
<td>124</td>
<td>159</td>
<td>0.049</td>
<td>62</td>
</tr>
</tbody>
</table>

Table 6. Heat Transfer Calculation Results (Mycelial)

<table>
<thead>
<tr>
<th>Heat load (MMBtu/h)</th>
<th>Gas hold-up</th>
<th>Jacket area (ft²)</th>
<th>UJ KT ([Btu/ft² h°F])</th>
<th>Coolant temp in (°F)</th>
<th>Coolant flow (GPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.533</td>
<td>0.202</td>
<td>298</td>
<td>40</td>
<td>35</td>
<td>190</td>
</tr>
<tr>
<td>0.560</td>
<td>0.230</td>
<td>306</td>
<td>40</td>
<td>35</td>
<td>210</td>
</tr>
<tr>
<td>0.463</td>
<td>0.124</td>
<td>276</td>
<td>40</td>
<td>35</td>
<td>235</td>
</tr>
<tr>
<td>0.515</td>
<td>0.182</td>
<td>292</td>
<td>40</td>
<td>35</td>
<td>200</td>
</tr>
</tbody>
</table>

The results obtained here must now be translated into a mechanical design that must also take into consideration the items to be discussed in the next section.

MECHANICAL DESIGN

Mechanical Design Basis

Much of the mechanical design follows from the results of calculations discussed in the previous section 4; however, these must be tempered by other considerations that should be included in the process and/or facility design basis. Some of these (e.g., regulatory compliance, containment) have already been discussed and are revisited in greater detail in this section. Other items that should be included and will affect fermenter design include the following:

- Extent of automation, nature of plant wide control system, etc. Such items affect not only the instrumentation (beyond our scope) and similar factors, but also valving, piping, vessel ports, and so forth. Among some of the major issues here are identification of critical (as defined by cGMP requirements) valves and other components. It must be kept in mind that failure to identify critical instrumentation and control components frequently leads to very complex valving systems in which the failure of a single switch can bring everything to a grinding halt. Unfortunately, there are too many people who do not think this through (as a system) before final design begins. Much of this results from the human tendency to rationalize and/or push "problems" downstream, the kind of thinking that results in the thought "Why worry? We'll get it whether or not we need it." Such thinking is in no one's best interest (buyer or seller).
- Staff requirements and the anticipated nature of the operating staff. This can influence the complexity and physical layout of (among other things) the piping system(s).
- Plant location. This will have some affect on overall design and component selection if for no other than service issues.
- Available utilities. This can influence such things as the designs of the cooling and aeration systems.
- Architectural constraints. These can have profound effects on the ways in which process requirements will be satisfied. Floor space and ceiling height constraints often require that a fermenter be designed in such a way that it violates some or most or even all of the rules of thumb mentioned in the previous section. It is not good but is better than building a vessel that does not fit into the plant.
- Transportation constraints. The fermenter has to be moved from the fabricator's shop to the plant—at a cost not greater than the total project budget. This may require some thumb bending or breaking. In extreme cases it may be better to fabricate the fermenter on site; however, this can carry large penalties, particularly in licensed facilities.
- Maintenance. Many design decisions can have major effects on the ease of maintenance. It is also often true that some of the design features that ease maintenance are more costly than those that do not. In most cases, the savings in capital expenditure will not come near paying for losses that will result later because of maintenance problems. This is particularly true in licensed facilities where regulators view good, facile maintenance as an integral, indispensable part of cGMP operation.
- Definition of standards. Standards for welds, finishes, and so on should be standardized for all parts of the equipment that will contact process fluid. It does not, for example, make sense to call for a high-quality vessel finish (e.g., 320 grit, EP) and at the same time accept unpolished tubing and valves in inoculation and medium addition lines.

Finally, there are a few recurring themes that are encountered in fermenter design:

- Building-in "versatility." The types of versatility desired range from wanting a fermenter that can operate well over a very wide range of conditions but with a narrow range of organisms, to wanting a convertible bioreactor that can handle microbes, mamm-
malian cells, and maybe (some day) transgenic animals. Most people understand the value of versatility, but many do not understand the attendant problems and costs. As a general rule, converting laboratory glassware directly into large, stainless steel equivalents usually costs more than any rational person should be willing to pay. Also as a general rule, versatility should decrease as a process goes from the lab to the production floor. Versatility in the laboratory is almost a requirement because change is in the nature of laboratory work. Versatility on the plant floor, however, leads to more procedures, more paper work, more testing, and more confusion, particularly when equipment modifications are required to achieve the versatility; change is not in the basic nature of most plant work. Equipment capital savings can justify the added costs of all the preceding for many cases in which the equipment is designed to handle multiple, similar fermentations on a campaign basis and without significant equipment modification. This statement becomes less true as the fermentations become less similar, as more modifications are necessary, and as regulatory scrutiny increases. Bottom line: analyze very carefully any inclination to want versatility built into plant equipment (or even pilot equipment, in some cases).

- Retrofitting existing equipment. The usual thinking is that capital and time savings can be had by refurbishing “old faithful.” Just how true this is depends not only on the condition of the existing equipment but also on the intended application(s) of the reborn version. Comments similar to the ones made for versatility apply here. The chances of success decrease as one goes from a lab to licensed production facility. There are ample, expensive corpses to prove this point.

**Vessel Design**

Please note that Figures 4 and 5 are simplified vessel and piping diagrams intended to assist the reader in the discussions to follow.

**Materials of Construction.** The major choices that must be made are (1) the type of metal to be used for the vessel and nozzles, and (2) the type of elastomer to be used for static seals (see later for rotating seals). The selections should be based on compatibility with the organism, compatibility with the product, corrosion resistance, cleanability (also related to finish), welding characteristics, and cost and durability. All of these should be determined during process development but seldom are. In almost all cases, the metal selected will be some grade of stainless steel (usually SS304, SS304L, SS316, or SS316L). The choice is usually associated with the nature and value of the product, although some of the other factors noted earlier may be considered. SS304 is usually good enough for lower-value, unlicensed products, whereas SS316L is the material of choice for high-value, licensed products. L-grade is selected when better corrosion resistance and good multipass welding characteristics are required; it adds about 15% to the cost of the vessel.

It is important in making metal selection to keep in mind that clean steam, purified water, and WFI are all highly corrosive. The reader should note that there is considerable controversy concerning metal selection and is directed to Refs. 61–63 for additional information and opinion.

Static seal (O-rings and gaskets) materials are usually limited to silicone, EPDM, Viton, or Teflon. Silicone and EPDM are the materials of choice for headplates and elsewhere, respectively. (The reader should be alert to the fact that there are several grades of both, not all of which are FDA approved.) Teflon has much better temperature resistance than either silicone or EPDM, but it doesn't stretch, and it cold flows. Viton hardens on use.

Finally, one should be aware of that some O-ring/gasket-forming processes can leave very small quantities of metals in the seals. These may not be enough to cause fermentation or product problems per se, but they can be the cause of considerable corrosion.

**Finish.** In most instances, the need for a high-quality finish is closely related to the need for high-quality cleaning: there is little question that CIP systems function better when surfaces are smoother. Smoothness is expressed on several scales, two of the more common being grit number and surface roughness, $R_a$. Grit number is related directly to the abrasive used to achieve the finish (mechanical); surface $R_a$ is the actual surface roughness as measured by means of a profilometer and expressed in microns:

<table>
<thead>
<tr>
<th>Grit</th>
<th>$R_a$ (μm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>3.2</td>
</tr>
<tr>
<td>220</td>
<td>0.8</td>
</tr>
<tr>
<td>320</td>
<td>0.4</td>
</tr>
</tbody>
</table>

The two most common ways of achieving smoothness are mechanical polishing and electropolishing. Mechanical polishing is done by means of a series of continually decreasing abrasive particle sizes (grits). Electropolishing is achieved by electrochemical removal of metal from high points on the surface. One is usually best advised to electropolish only after getting a good mechanical polish (240-grit minimum). It has long been held that the very good finishes required for vessels where cleaning is critical could be achieved only by electropolishing a 320-grit mechanical finish (although 240-grit EP is the most common practice). There is some evidence in the literature (62,63) to support this, but it is not overwhelmingly convincing. It also is important to note that recent advances in mechanical finishes have produced surfaces that easily rival the appearance of any electropolished surface (Lee Industries, personal communication, 1996).

It should be (but is not always) evident that cost increases with surface smoothness. It is also important to note that because a surface is almost perfectly smooth does not guarantee that it will be easy to clean. Cleanability will also depend on the nature of the interaction between foulants and the surface material. The only way to know for sure is to try your broth on various surfaces.
then make a specification based on rational evaluation of the experimental information. Surface material and finish, cleaning agents, and cleaning protocols must be considered together—preferably before design.

Passivation must also be considered. This is a treatment that restores oxides that protect stainless steel from corrosion but can be removed during vessel fabrication (primarily welding). The process is rather simple, but the surface chemistry is not and is not completely understood (64,65); hence, there is controversy as to which method is best and whether it is really as necessary for fermenters as it is for high-purity water systems where rouge formation is a serious problem (66). Two methods are in common use: (1) treatment with sodium hydroxide, citric acid, and nitric acid; and (2) treatment with chelating agents. The second is much milder and does not generate the noxious waste of the first. Both appear to be equally effective; hence, the chelating method is becoming more popular.

Nozzles. Ports for additions and probes must be designed to be sterilizable and cleanable. Among other things, this means they must have reliable seals and be free draining. The major debate here usually focuses on the choice between Ingold ports and those designed for sanitary clamp connection (Fig. 6). There have been sterility problems with Ingolds, most of which have been traced to the ports being out of round as a result of distortions caused by welding. Such distortion causes obvious problems with the O-ring seals. This problem can be corrected quite easily by making the nozzle thick enough to avoid distortion. Our experience has been that such a correction yields nozzles that are just as reliable and more easily cleaned than are sanitary fittings when both are mounted at the usual 15°.

Side View Ports. The first word that comes to mind is "don’t." Side view ports let the bugs see out more than they let you see in. The little that’s gained by being able to see less than 1% of the total action does not seem to justify the expense, the jacket coving, and the cleaning and sterility problems caused by these little gems. If you must have one, we suggest the circular Metaglas type (Fig. 7) attached to the vessel with sanitary clamps; the glass is bonded directly to the metal, which avoids the sealing problems encountered with other types. It does, however, have the disadvantage of having a smaller viewing area than does a standard, circular view port, and it is more expensive. We
recommend avoiding long rectangular view ports, which are very difficult to keep clean and free draining, and types in which the glass has to be sealed with gaskets and/or O-rings on both sides (Fig. 8). Keep in mind that tolerances for glass are measured in fractions of an inch, not thousandths. That, coupled with the fact that the gaskets will compress means that the probability of repeatedly achieving a good seal and good cleaning is lower than your odds of winning more than once at Atlantic City.

**Baffles.** Baffles are usually required in high-power systems to prevent swirling and vortexing, thereby increasing the power that can be delivered to the fluid. The usual practice is to use four baffles on 90° centers welded directly to the wall. Each baffle should have a width equal to 10% of the tank diameter and should have long slots cut out of the edge facing the wall so as to prevent solids build up and to make cleaning easier. Removable baffles are used in some cases; however, we discourage this practice, particularly in cases where cleaning is a critical issue, simply because unsealed joints resulting from baffle removal make cleaning more difficult.

**Jackets.** Several types (67) of jacket are used on fermenters; the choice of type is usually not critical for heat transfer purposes and is best left to the vessel fabricator. In some cases, however, a jacket type (e.g., half pipe) may be chosen to increase the vessel pressure rating. The following have proven to be useful practices:

- The jacket should extend from the probe ring (about 2 in. above the bottom tangent line) to the top tangent line and should be zoned. This allows additional active surface area to come in contact with the broth as the volume increase due to additions and to increas-
10% of the straight side jacket area (allowing for bottom drain, etc. and will add about 5% to the cost of the vessel). In cases where there is an internal coil, bottom jacketing isn’t necessary or desirable.

• Jacket coving to accommodate view ports, decreases jacket area and effectiveness and increases cost considerably. It should be avoided.

Internal Cooling Surfaces. Coils are the most common internal cooling surfaces, although there are other types (68). They can easily double the available heat transfer area and tend to be more effective than jackets; however, they can have big disadvantages:

• They make cleaning very difficult.
• They can cause additional bulk mixing problems, particularly for very viscous non-Newtonian broths.
• They will eventually leak nonsterile coolant into the broth.
• They can add up to 25% to the cost of the vessel.

We recommend that every other alternative (including slight decreases in growth rate and/or cell mass) be considered before yielding to the temptation to use coils. Keep in mind that, once installed, they will be there for the life of the vessel. If you absolutely, positively must use a coil, there are several points to remember:

• Mount it in a way that will insure minimum stress during heat-up and cooldown.
• Weld cladding over the butt welds used to join the coil pipe sections.
• Leak test (69) after construction and build in design features that will simplify leak testing on a regular basis thereafter.
• Space coil turns at least 3 in. apart. Anything closer will insure major cleaning problems.

Spargers. There has been a lot of discussion concerning the pros and cons of ring and single-orifice spargers, and the details of design of each. We have seen both work well.
and have not found any evidence for the validity of claims concerning the importance of hole size (for example) for oxygen transfer per se. The primary focus should be on gas distribution (which will depend on other aspects of the agitation system) and on aseptic operation. We usually favor single-orifice spargers.

There also have been advocates of porous (frits) spargers. The rationale presented has focused primarily on the small bubble size such spargers produce. There is some basis for these claims in cases where little mechanical energy is available for bubble breakup (e.g., in mammalian cell systems). There might also be some value in cases for which bubble coalescence is not a problem (far and few between in practical systems). They can be extremely difficult to clean, particularly for mycelial organisms. We usually suggest they not be used except in very special cases, as just noted.

Finally, it is important that the sparger be made and mounted in such a way that it can be removed easily for cleaning.

**Piping and valving**

Design and construction details of the piping system depend on process, sterility, cleaning, and containment requirements, taken together.

**Service Lines for Non-Process-Contacting Fluids.** All lines providing fluids that cannot contact process fluid surfaces (e.g., coolant lines, plant steam lines for heating only) fall into this category. Satisfactory service is provided by either copper or stainless steel piping along with a rational combination of welded and compression fittings. Durability, serviceability, cost, and corrosion resistance are major considerations. Ball valves are satisfactory on lines not turned on and off frequently (diaphragm valves tend to withstand a greater number of on–off cycles prior to failure).

**Sterile Piping Systems.** Before we discuss design of sterile piping systems, we reiterate the importance of a systems approach to integrating vessel and piping design. Independent design almost always results in a lot of aggravation, as well as higher cost and lost time.

A sterile piping system can usually be divided into five major subsystems: process, steam/condensate, air, harvest, and seal lubrication. Each has specific requirements dependent on its unique functions as well as the specific demands of the process. None of these systems is completely independent of the others, and all share common components. All also interact with the CIP piping system, which is discussed later.

Details of each of these subsystems follow a simplified description of sterilization procedures (Fig. 5). The reader is advised that the procedure is one of several commonly practiced; there are, for example, different opinions as to when live steam to the air inlet line should be turned on.

Phase 1. After the vessel is filled with set medium, steam flow is started through the jacket. The agitator is run at low speed, and particulate free steam lubricates the shaft seal. The exhaust line is open for venting to the atmosphere. All other valves are closed. During this time not only are the vessel and the set medium being heated, but air is being purged from the medium. If the air is not purged, the vessel pressure during sterilization can become dangerously high, and there will not be a reliable correlation between pressure and temperature. Phase 1 continues until the medium temperature reaches 100 °C.

Phase 2. Steam continues to flow through the jacket. The air exhaust line is closed; the only path out is via steam traps. Steam is now admitted through the air inlet line to sterilize the inlet filter; steam leaving the vessel through the air exhaust line sterilizes the exhaust line piping and the exhaust filter. Steam flow is started through subsurface ports via steam lock assemblies (Fig. 9). (See later for further discussion of steam locks.) Ports above the liquid surface are sterilized by steam escaping through them from the vessel and then to condensate lines. Steam flow also is started through the bottom drain valve. All steam flows continue until the cooldown begins. Phase 2 continues until the medium temperature reaches 121 °C.

Phase 3. Steam flow to the jacket is throttled to hold sterilization temperature constant until cooldown begins.

Phase 4. Cooldown. This is not discussed here.

**General Principles.** There are several general principles that should be applied to all sterile piping:

- Make all piping system components free draining. This requires special attention to the details of pipe pitches, valve orientations, and so forth
- Design all components and the piping layout so as to eliminate nooks and crannies where contamination (biological and nonbiological) can hide so as to escape cleaning and/or sterilization. Frequently overlooked problems include such things as deadlegs and ridges formed by welding operations.
- Make steam and condensate piping at least 3/8 in. diameter to insure proper steam flow and condensate draining.
- Make all piping lengths as short as possible without interfering with good fabrication practices (GFP), operability, and maintenance.
- Use check valves only when no other solution is possible (e.g., in overpressurized lines).
- Do not use sight glasses in condensate drain lines except in the seal lubricant drain line.
- Pay careful attention to piping orientation to avoid the possibility of air traps and inadequate heating (55, 56).
- Make sure that bottom drain valves are flush-mounted diaphragm valves capable of being steamed in the closed position (Fig. 10). They must be free draining.
- Avoid dip tubes unless there is no other way (there almost always is).

Finally, the design should take into consideration the methods that will be used for validation of sterilization operations. These will vary considerably with the product and the organism. At one extreme is virtually no validation; at the other extreme is temperature mapping of the vessel and all the trap lines, as well as the use of spore strips and/or spore ampules in the vessel and in the trap lines. In some cases, tee sections have been included in drain lines for insertion of spore strips.

Valves. Diaphragm valves have become essentially required for most classes of regulated fermentation products. There is little argument that their design provides the greatest reliability for clean, aseptic operation. They have the added advantage of integral sterile access ports, which permit very short piping runs, particularly in valve clusters. Historically, diaphragm valves have been more expensive than have ball valves, but this has changed recently, particularly for some materials (e.g., SS316L). It therefore makes sense to use diaphragms for most applications. The primary exceptions are steam lines or other applications in which diaphragm life is a problem; in those cases one should consider ball valves or plug valves, as appropriate.

Another important consideration is the extent to which one uses valve actuators and position indicators. Obviously, this will depend in large measure on the nature of the control systems employed. In any event, one should consider very seriously the problems (e.g., space and orientation constraints, reliability) associated with valve “extras” while control system decisions are being made.

Steamlock Assemblies. The primary purpose of steamlocks (also known as block and bleed) is to allow sterilization of ports and process piping at any time during a fermentation in such a way that connections can be made and broken without risk of breaching the sterile barrier and/or the containment barrier. A basic steamlock assembly is illustrated in Figure 11. In this case, vessel T1 contains presterilized nutrient that must be added at some time during the fermentation. T1 is connected to the fermenter by means of sanitary clamps (other means are possible). Steam flow is then started (open V1, V3, and V5) so as to sterilize the connecting hose and all the valving not previously sterilized. Condensate goes to drain (sanitary, if necessary) via V3 and V5. The steam and condensate valves are closed after sterilization is complete; process valves V2 and V4 can be opened anytime thereafter to make the sterile transfer. Note that the diaphragm valve sterile access ports make it much easier to sterilize the

![Figure 10. Bottom drain valve.](image-url)
steam side of the valves than is possible with other valve types. Variations on this theme are better suited to cases involving long transfer lines or having other special requirements.

**Air Inlet and Outlet.** Figure 12 is a simplified diagram of a “typical” air inlet/outlet system. The major problems for this system are related mainly to the filters. Chief among these (other than outright failure) is wetting of both filters caused by condensate during sterilization, and wetting and clogging of the exhaust filter by condensate and/or aerosol during fermentation. These problems are not all easy to resolve and should be dealt with most seriously during design. Wetting and/or clogging of the inlet filter by condensate and/or compressor oil during fermentation also can be a problem but is usually the result of not using proper quality air, having no or improper prefilters in the air inlet line, and/or poor maintenance (e.g., of the compressor). All of these are easily remedied.

The extent of condensate problems during sterilization can be remedied best by ensuring that whatever condensate does form can drain freely. This can be complicated by the fact that it is undesirable to have the sterile side of the filter connected directly to the drain line. Another approach is to steam heat the filter housing such that condensation cannot occur. This is effective but has the disadvantages of adding cost and decreasing filter life.

Other factors that can affect condensate problems are the positioning and orientation of the housing. There are differing opinions concerning these; one is best advised to consider the advice of the filter and the fermenter vendors for specific cases.

Plugging of the exhaust filter by condensate and aerosols during fermentation require special consideration. Air leaving the fermenter will be essentially saturated with water vapor at fermentation temperature. The exhaust line, filter, and so on usually are colder than the fermenter; therefore, condensation is inevitable. The amount of condensation will depend on temperature differences, air flow rate, and the nature of the surfaces of the components in the exhaust line (the maximum potential is easily calculable). Approaches that have been used to deal with these problems include (70):

- Condensers
- Heat exchangers; used before the exhaust filter to avoid condensation from the fermenter off gases and after the pressure control valve to prevent reflux from the exterior exhaust line
- Steam-heated filter housings
- Heated exhaust lines
- Coalescers
- All of the above

None of these is completely successful, and each has its proponents and detractors. Suffice it to say that each case should be considered on its own merits (e.g., fermentation temperature, duration, and air flow) and that combinations of the approaches should not be ignored.

The condensate problem is exacerbated by the aerosol problem. Aerosols will form in any aerobic fermentation and will carry liquid along with dissolved solids and particulates (including organisms) to the exhaust line. The only questions are how much and what problems they will cause. Among the problems are deposition of dissolved and particulate solids, not only on the filter, but also on heated surfaces or condenser surfaces. Among the problems deposition causes is reduction in heat exchange effectiveness, which reduces the capabilities of the previously mentioned devices to eliminate condensation. In severe cases, exhaust line blockage can lead to other problems related to pressure buildup and/or higher linear velocities in the exhaust line. The problem should be dealt with on a case-by-case basis during process development and fermenter design.

Provisions should be made for in situ integrity testing of sterile filters. This can be done by means of any of several commercial electronic testing devices that are based

![Figure 11. Steam lock assemblies for transferring sterile medium.](image-url)
Figure 12. Simplified piping diagram illustrating air inlet and outlet systems.

on liquid intrusion, diffusion, or some other well-established method for testing membrane filter integrity (71–73). One example of the piping necessary to perform the test on an exhaust filter is shown in Figure 13.

Finally, two pressure control valves are required in the case of constant pressure sterilization (continuous flow of steam into the vessel and out to atmosphere). The reason is that the steam flow will be very much lower than the air flow during fermentation; hence, the valve \( C_v \) required to maintain steam pressure accurately will be far too small to control air pressure accurately.

**Sampling Systems.** One example of a simple aseptic sampling system is illustrated in Figure 14. The sample vial, vent filter, and valve \( V_1 \) are sterilized as a unit in an autoclave. The unit then is connected to the fermenter piping via a sanitary clamp. After the connection is made, steam is introduced via \( V_3 \) and passes across the nonsterile sides of \( V_2 \) and \( V_1 \) and then to trap via \( V_4 \). Once the valves and line are sterile, the steam and condensate valves are closed. Samples may be taken any time after the sampling system cools. Samples are taken via \( V_1 \) and \( V_2 \). Lines and valves are resterilized before the sanitary connection is broken.

The valves selected for this system will depend on the class of service required. We suggest that vessel sampling valve \( V_3 \) be a flush-mounted diaphragm valve (e.g., Asepco®, NovaSeptic®) designed in such a way as to minimize dead volume and piping lengths; it is an excellent choice for service requiring very high levels of asepsis and cleanability. Less expensive valves can be used for less-demanding service, but we think that the initial extra cost will more than pay for itself. Diaphragm valves are suggested for \( V_1, V_3, \) and \( V_4 \) in cases of demanding cGMP or containment requirements. This is not required for all classes of service; indeed, it is usually satisfactory to use plug valves for steam and condensate lines even in many demanding circumstances. It should be noted, however, that you may encounter a perception problem if you use them for applications deemed to be sensitive.

There are several other designs that have been used successfully. A few have been designed specifically for systems requiring BL3-LS containment or higher. These are
discussed later. Finally, please note that addition system design and operation are very similar to those for sample systems and are not discussed here.

**Agitation Systems**

Agitation can be done by direct mechanical coupling of the shaft to the drive or by magnetic coupling. The latter has been recommended for cases where high levels of containment are required (74). There also is the perception that magnetic drives are more suitable for maintaining more stringent levels of asepsis and cleanliness than is possible with direct drive. Neither claim has any substantive basis; indeed, there is good reason to believe that existing magnetic drives may present greater cleaning difficulties because of the manner in which the driven magnet must be mounted at the bottom of the vessel. In addition, power transfer by magnetic drive is quite low.

Direct drive is the predominant current choice at almost any scale of operation. The major mechanical decisions that must be made in this case (other than power) have been discussed at length elsewhere (4,75). The major arguments for bottom drive are ease of maintenance, shorter shafts, less support structure, and lower overall height. The arguments against bottom (and for top drive) focus primarily on the potential of catastrophic spills resulting from bottom seal failure, seal grinding as a result of broth particulates working into a bottom seal, and greater cleaning difficulties. If the seals are designed and maintained properly, none of these is a problem. We have seldom seen any real difference in aseptic operability between top and bottom drives, and we are reasonably certain that the organisms don’t care. The choice probably will continue to be driven primarily by personal preference.

There are very few cases in which double mechanical seals are not (or should not) be used in fermenters. The major debate focuses on seal orientation and the details of individual seal designs. There are basically two orientations used: inline (Fig. 15) and back-to-back (Fig. 16). The details are discussed in Refs. 4 and 75. We recommend inline design because we have found it to operate more cleanly and require a simpler sterilizing/lubricant system (4,75).

Seal lubrication is usually provided by means of sterile steam condensate. It is extremely important that this condensate be free of particulates: their presence guarantees rapid seal failure, contaminated fermentations, and a hyperactive maintenance program. It is also important to note that during sterilization live steam flows through the seal housing. There are some who insist on keeping the steam flowing throughout the fermentation. (Obviously, they have no faith in the seals.) The one thing this will guarantee is much more rapid wearing of the seals (perhaps supporting the lack of faith in the seals).

One must also decide on the means for controlling lubricant flow rate. Most use a valve for this purpose. We suggest an orifice sized to deliver the proper flow. This avoids the cost and maintenance of a valve and insures fiddle-proof operation. It does, however, require the use of particulate-free condensate, but then so does proper operation of the seals. We also recommend including a sight

---

**Figure 13.** Piping diagram for in situ air outlet integrity test.

**Figure 14.** Aseptic sampling system.
glass in the lubricant drain line as well as a seal leak detector (Fig. 17). Other more complex detection systems might be considered for specific circumstances (e.g., severe containment requirements).

Finally, preventing shaft vibration is another important factor in agitation system design. Shaft vibration is a safety hazard. It will also cause premature seal failure and other costly mechanical damage. Details have been presented elsewhere (4).

Cleaning Systems

As noted earlier, specific requirements for cleaning depend on several factors including the nature of the fermentation broth. There is a major difference between most microbial broths and most mycelial broths. Single-cell microbial broths, with the exception of those containing a lot of undissolved particulates and high viscosity components (e.g., xanthan broths), tend to be free draining and readily amenable to cleaning based primarily on the physicochemical action of the cleaning agents. Many fungi and mycelial bacteria, on the other hand, tend to cling to fermenter internals and may require mechanical action (e.g., high-velocity jets) in addition to cleaning agents.

The following guidelines are applicable for most systems. These principles must be applied in light of the actual cleaning agents and protocols to be used:

- Eliminate internals and nooks and crannies to the greatest extent possible in the vessel and throughout the piping system. This practice is consistent with design for aseptic operation.
- Drill and position sprayballs to insure complete coverage of all surfaces inside the vessel. This usually requires an empirical approach. Coverage can be tested by means of the riboflavin test. The reader is cautioned, however, that complete coverage is a necessary but not sufficient condition for cleaning. Also, please note the following:
  - Spray balls designed for sanitary operation are self-draining and self-cleaning. They can be sterilized in situ. There is, however, considerable debate concerning whether they should be removed prior to fermentation.
  - High-velocity, rotating devices that may be necessary when large clumps of sticky residue must be removed (e.g., as with a fungus) are not designed to be inherently self-draining, self-draining, or sterilizable. They should not be mounted permanently.
- Eliminate deadlegs in piping and deadspaces in valves, fittings, and other system components. This is also consistent with design for aseptic operation.
- Specify the same materials and finishes for process and CIP piping as are specified for the fermenter. Ob-
viously, these must be compatible with the cleansers and conditions used.

- Avoid threaded joints. A completely welded system is best, but compression fittings can be satisfactory when cost is a major consideration, particularly for nonregulated products.
- Make CIP piping as simple as possible. For example, make dual use of process and other piping as much as possible.
- Avoid complex, expensive transfer panels wherever possible. Use swing elbows wherever practical.
- Design and construct the system to facilitate validation and on-going testing for removal of contaminants, including the cleaning agents. This design should provide for swabbing, obtaining rinse samples, or whatever else the cleaning protocol requires.

A portion of one approach to integral CIP piping is illustrated in Figure 18.

There are many cases for which a portable CIP system is preferable to an integral system. For such cases, the fermenter and the portable unit should be designed to allow simple mechanical attachment of the necessary hoses between the two units and with utilities, and a straightforward means for interfacing the instrumentation, logging data, and controlling the systems of the two units.

Finally, we note that classical CIP systems rely on continuous cleanser flow and the maintenance of a shallow puddle in the bottom of the fermenter. To achieve this they rely on special pumping devices such as eductors (74). Aside from the design, control, and other operating problems this causes, it has been observed in some facilities that the stable pool leads to the formation of “cleaning rings” at the bottom of the fermenter. These rings may not be real problems (additional evidence is still required), but they do cause perception problems. We suggest the use of a pulsed-flow system to overcome not only the need for special pumping devices but also the “cleaning ring” problem.
Such pulsed systems have been found to accomplish both objectives in practice.

Additional details concerning fluid velocity, vortex elimination, and many other aspects of CIP system design and operation are described elsewhere (59).

**Containment**

Our scope is limited to containment of fermenters; however, we reiterate here the importance of making certain that the equipment, the facility, and the operating protocols be considered simultaneously to ensure compatibility with regard to containment (just as in the cases of sterility and cleaning). As noted earlier, most new facilities for regulated products are being built to satisfy BL2-LS facility requirements and (in most cases) BL3-LS fermenter containment requirements. With few exceptions (about which there is no agreement) there is no difference with regard to the equipment containment requirements at the two levels—at present.

The guidelines require that BL2-LS/BL3-LS fermenters be designed and constructed to prevent organism release (primary containment) from the vessel and from any of the subsystems noted earlier; hence, they must be built as closed systems. (Note that it also is possible to enclose a fermenter in a biosafety "cabinet," but this is not usually practical or necessary.) Some areas that have been subject to greater scrutiny than has been required for cGMP compliance are as follows. In evaluating the discussion, the reader should bear in mind that there is not much information readily available concerning containment reliability data. Some progress is being made via the Industrial Biosafety Project in the United Kingdom.

- Static seals. It has been suggested that double static seals (e.g., in headplates) be used for BL2-LS and that double static seals with a steam trace between the seals be used for BL3-LS (76). There has been little to demonstrate that these are either needed or desirable. Indeed, there have been substantive arguments made against using such devices (76). There is as yet no consensus.
- Rotating seals. Arguments have been made in favor of using magnetic drives at anything higher than BL2-LS. This was discussed earlier. There are many who argue that only top drive be considered for BL2-LS and above. Their reasons have mainly to do with avoiding the possibility of large spills in the event of catastrophic seal failure. Such a circumstance is possible but highly unlikely. Leak detectors (which would normally be installed for cGMP requirements) would almost certainly indicate a leak long before there was the remotest possibility of catastrophic failure. One also should consider that a leak in top drive seal might be more difficult to detect and could lead to release of a more insidious nature. Again there is no consensus. Finally, it has been suggested that a low-pressure sensor or flow sensor be used to detect lubricant flow failure. This could help to avoid seal failure (77).
- Sampling systems. One system recommended (78) for contained sampling is illustrated in Figure 19. Note that it is similar to the system in Figure 14 recommended for cGMP compliance. It is a bit more complicated and may provide some incremental benefit; however, this might be outweighed by the greater likelihood of operator error. There have also been more complex systems suggested that involve more intricate valving and/or biosafety cabinets. It is apparent that this is an area that needs considerably more development work.
- Air exhaust system. It has been suggested that two in-series sterile filters or a sterile filter followed by an incinerator be used at BL2-LS and higher. All other aspects are covered by the earlier discussion of exhaust lines.
- Pressure relief. This is an area of some controversy because of the need to satisfy physical and biological safety requirements simultaneously. It has been suggested by many that relief venting be done via a large kill tank protected by a HEPA filter; however, this could be in conflict with physical safety codes that require there be no devices in the relief path (79). We do not have anything approaching universal agreement here; however, it does seem fairly apparent that limiting the supply pressure would help to minimize the risk. There does appear to be agreement on the use of rupture disks rather than pressure relief valves; this is consistent with accepted practice for
Figure 18. Simplified CIP piping diagram.

cGMP compliance. Rupture disks are cleaner and are not prone to sticking.

Finally, HAZOP evaluations for fermenter containment have been recommended. Such analyses would evaluate potential hazards that might be caused by a wide range of incidents (e.g., fire) not limited to normal operation.

CONCLUSION

Our intent has been to present a balanced view of stirred-tank fermenter design and construction for microbial and mycelial organisms. There are, however, several important aspects that space and scope considerations prevent us from considering in the depth they warrant. In
particular, we would like to have included more information related to overall process considerations, instrumentation, and control; however, we believe that these are addressed elsewhere in this volume. We trust that we have conveyed the sense that fermenter design is very much an art and that we have convinced you to consider each case separately.

**BIBLIOGRAPHY**

INTRODUCTION

Among the most capable and attractive means of microparticulate (greater than 0.1 μ) separation used in biopharmaceutical processing today is the use of highly porous powdered media in dynamic systems. Powdered media, often called filter aids (Fig. 1), provide versatility, high solids-loading capacity, high product recovery, low cost, and ease of scale-up in any filtration process. Moreover, recent technical advances have stimulated a much greater breadth of applications that use this media filtration in biopharmaceutical processes, with a concomitant exponential increase in product innovation.

PROCESS CLARIFICATION

The first unit operation in downstream processing is clarification. It involves the removal of cells, cell debris, or precipitated components from a fermentation broth or process supernatant. Four types of solid-liquid separations are common to the industry: (1) The removal of whole cells in fermentation broths where the product is expressed into the supernatant; (2) The removal of cell debris from process broths after cell disruption and extraction to release the product of interest; (3) The removal of precipitated contaminants (usually proteins) from a process fluid, for example, in the removal of misfolded forms of recombinant proteins; and (4) The selective precipitation of the target protein and the capture of this precipitate from the process supernatant.

Most processes use either centrifugation or filtration for solid-liquid separations. Centrifugation, although widely accepted, has some disadvantages in biopharmaceutical applications. The challenge in many applications is to remove both cellular and subcellular debris without applying excessive shear forces on the solids. Centrifuges depend on differences in density and on the centrifugal force applied for solid-liquid separation. However, centrifugation becomes less practical when the product and waste possess similar densities. In addition, centrifuges can be difficult to maintain, clean, and sterilize. Finally, it is important to avoid the generation of aerosols when dealing with biological fluids, and this is often difficult to avoid when using centrifuges for this unit operation.

Depth filters containing diatomite, perlite, and cellulose are suited to low solids applications, because these products tend to blind quickly when subjected to moderate to high levels of solids. Their capacity can be increased by the addition of filter aid (as body feed) into the unfiltered broth, thus extending the lifetime of the pad.

The use of powdered media offers many advantages in biopharmaceutical applications. These systems are much more flexible and dynamic because powdered media addition can be adjusted based on changes in incoming solids level. In addition, they are extremely efficient at removing suspended solids and are particularly effective at clarifying supernatants heavily contaminated with colloids and small particles. The media can be used in all four applications described above, including the selective precipitation of the target protein and the capture of this precipitate from the process supernatant. Once captured on the filter media, the protein or product of interest can be redissolved by adjusting the pH or salt concentration in the wash buffer and recirculating this solution through the filter media bed. After resolubilization, the product can be eluted from the media with approximately two bed volumes of buffer.

Capital costs for powdered-media-based systems are relatively low when compared with centrifugation or other filtration systems. Because the media are disposable, media cleaning and lifetime studies do not need to be performed, hence the cost of validating the filtration process is substantially reduced. In addition, the ability to postwash the filter cake helps to maximize product recovery, a process that is extremely expensive or impractical with centrifuges. Finally, with improvements in both powder
containment and filter equipment design, complete containment is now easily achieved both in the media preparation area and in GMP suites.

The combination of these improvements in powder containment, filter design, and powdered media technology position filter aids as a very cost-effective and economical approach to clarification.

POROUS MEDIA IN DYNAMIC PROCESS FILTRATIONS

Diatomite, perlite, and cellulose are the most widely used porous media in dynamic process filtrations, with a high percentage of applications using diatomite.

Diatomite

Diatomite is often used in combination with cellulose; however, diatomite is often the key ingredient of static or fixed-bed filters such as filter sheets and filter pads. Diatomite products are especially characterized by an inherently intricate and highly porous structure composed primarily of silica. These products are obtained from diatomaceous earth, a sediment greatly enriched in biogenic silica in the form of siliceous frustules of diatoms, a diverse array of microscopic, single-celled golden-brown algae of the class Bacillariophyceae. Surprisingly, these frustules are sufficiently durable to retain much of their ultrastructure nearly intact through long periods of geologic time when preserved in conditions that maintain chemical equilibrium.

Perlite

Perlite is a naturally occurring volcanic glass that thermally expands upon processing. After milling, porous, com-
licated structures result (Fig. 2) that are also useful in
dynamic process filtrations of biopharmaceuticals. Be-
cause its structure is not as intricate as that of diatomite,
perlite is better suited to the separation of coarse micro-
particulates from liquids having high solids loading. Like
diatomite, perlite is also useful as a functional filtration
component of filter sheets and filter pads.

Cellulose
Cellulose, like perlite, possesses a less intricate structure
than diatomite. As a result, the use of cellulose is generally
limited to coarse filtrations or in specialized applications
where a fibrous precoat on the septum is required.

The rest of this chapter will focus attention on the use
of diatomite in solid-liquid separations.

FUNDAMENTAL PRINCIPLES OF DIATOMITE FILTRATION

A discussion on solid-liquid separations can cover a wide
range of techniques and applications, but each method is
often defined by the relative size of solids being removed
or collected from a given process stream. This chapter will

Figure 2. Scanning electron micrograph of perlite.
focus on microparticulate filtration, where the solids are typically greater than 0.1 μ. The key objective is to remove unwanted solids from a process stream, so that the refined solution (filtrate) is adequately clarified and qualifies for further downstream processing.

**Solids Composition**

In general, solids can either be rigid or compressible in nature. Solids that form rigid cakes have some degree of inherent permeability, dictated by particle size distribution and packing arrangement, but most solids in biopharmaceutical applications are typically compressible. These often gelatinous, highly compressible solids retain a degree of permeability if low flow rates and low differential pressures are used for filtration. Unfortunately, these characteristics are contrary to typical processing requirements where this unit operation needs to be completed quickly and economically.

**Dynamic Filtration**

The problem in static fixed-bed filtration is that unwanted solids have very limited permeability when they collect or accumulate onto the filter septum (pad, paper, or fabric or metallic or plastic woven wire screen). If these solids build up on a septum, then eventually these solids will lack sufficient permeability for fluid drainage and filtration terminates. This behavior highlights a mistake frequently applied in filtration practice; often, higher differential pressure does not guarantee faster filtration rates precisely because these solids collapse to form an impermeable cake.

Introduction of a filter medium changes the composition of the accumulated cake, and therefore its filtration behavior. Powdered media (filter aids) are essential to this technology because they provide two functions: (1) as a precoat applied before the start of a filtration cycle and (2) as body feed added to the unfiltered feed throughout the filtration cycle. The precoat layer protects the filter septum, preventing penetration and blinding by the unwanted solids. This layer also facilitates septum cleaning and provides for a high-quality filtrate from the beginning of the filtration cycle.

The addition of powdered media (as body feed) to the process feed increases permeability in the accumulating filter cake, restricts solids movement, provides channels for filtrate recovery, and extends cycle length (Fig. 3). The extended permeability afforded by these media slows the rise in differential pressure for constant flow processes and retards the drop in flow for constant pressure operations. At the conclusion of a filtration cycle, product recovery can be maximized by rinsing or purging the accumulated cake in place. It is extremely important to maximize the body feed in a process. By adding too little, the solids quickly blind the filter cake and the filtration is terminated. Adding too much is not only wasteful but leads to a rapid increase in filter-cake thickness, with a corresponding increase in resistance to flow and a subsequent reduction in cycle length (1).

**Filtration Theory Overview**

The following section is not intended to be a comprehensive treatment of filtration theory, but rather an overview to describe how the permeability of an accumulated filter cake (solids and filter aid) ultimately affects filtration performance (rate). A thorough discussion of filtration mechanics can be found in numerous reviews (2–4).

Any discussion on porous media filtration can start with a simplified form of the Darcy equation (equation 1), which is used to describe laminar (nonturbulent) fluid flow through a homogeneous, porous medium.

\[ k = \frac{Q \times \Delta x \times \mu}{A \times \Delta P} \] (1)

The relative permeability (k) is measured in darcy units, and it is described in terms of the instantaneous volumetric flow rate (Q) and viscosity (μ) of the fluid, the thickness (Δx) and cross-sectional area (A) of the porous medium (accumulating filter cake), and the pressure drop or differential pressure (ΔP) associated with flow. The flow rate (Q) can also be represented by the differential change in volume over time (dV/dt). A value of 1 darcy corresponds to the permeability through a filter media 1-cm thick that allows 1 cm³ of fluid with a viscosity of 1 cP to pass through an area of 1 cm² in 1 s under a pressure differential of 1 atm (i.e., 101.325 kPa). Although fluid viscosity varies with temperature and shear forces, values for it and the other variables can be obtained from tables and experimental observations. Consequently, the permeability (k) is an empirical value (constant) used to characterize the filter cake composition and not a property that can be readily altered by changing the process flow rate or differential pressure. In practice, permeability may appear to be influenced by
flow rate or pressure, but those responses are strictly con-
sequences of composition, and they are often inconsistent
with the behavior described by equation 1. The type and
relative usage rate of filter aid coupled with the nature and
concentration of solids define composition, which in turn
controls permeability and filtration performance.

The constancy of k explains why (1) the differential
pressure increases as filter cake accumulates (d x) for con-
stant flow (Q0) processes (equation 2)

\[ k = \frac{Q_0 \times \mu}{A} \text{ (constant)} \times \frac{dx}{dP} \] (2)

and (2) flow decreases as cake accumulates for constant
pressure (dP0) processes (equation 3)

\[ k = \frac{\mu}{A \times dP_0} \text{ (constant)} \times Q \times dx \] (3)

Permeability measurements are not essential in the ap-
lication of dynamic filtration technology. In fact, once the
unfiltered feed is introduced to the filter, the overall per-
meability (as measured in darcies) of the filter cake can be
reduced by orders of magnitude. Ultimately, the manipu-
lization of permeability (k), through filter aid selection and
usage, describes the optimization of this solid–liquid sep-
aration process.

As an example, initial (laboratory scale) testing devel-
opps enough information to determine the permeability (kapp)
in a constant pressure application. In fact, a key lab-
atory test is to determine the elapsed time (t0) it takes to
produce the maximum cake thickness (dx; limited by the
physical capacity of the filtration equipment, typically
3 cm) at optimum filter aid usage and terminal pressure
(dPf). Hence, the Darcy equation can be put to better use
for process scale-up (equation 4):

\[ [k_{\text{app}}] = \frac{Q_{0 \text{ lab}} \times dx \times \mu}{A_{\text{lab}} \times dP_f} = \frac{Q_{0 \text{ plant}} \times dx \times \mu}{A_{\text{plant}} \times dP_f} \] (4)

Notice that differential pressure (dPf) and filter cake
thickness (dx) need not be affected by the scale of oper-
tions. Differences in allowable cake thickness (dx) must
be accommodated by adjusting filtration area (Aplant) re-
quirements. A physical capacity test also determines the
total volume of process feed that can be filtered on the
given area. In the simplest case of equal cake thickness,
scale-up becomes a matter of sizing filter area to accom-
modate production rate (equation 5).

\[ \frac{Q_{0 \text{ lab}}}{A_{\text{lab}}} = \frac{Q_{0 \text{ plant}}}{A_{\text{plant}}} \] (5)

GRADE SELECTION AND OPTIMIZATION

An optimized filtration process maximizes throughput and
product recovery and minimizes pressure drop while main-
taining the desired filtrate clarity, all in a reasonable time.
There are five main parameters one needs to consider
when selecting an appropriate grade of powder media for a
given process.

Product Stability

Careful attention should be paid to soluble metals when
selecting a grade. High concentrations of soluble metals in
reagents can lead to contamination of biopharmaceutical
products, for example, aluminum contamination of albu-
min (5–7). Also, high concentrations of soluble metals are
known to oxidize proteins and enzymes and can activate
proteases in fermentation broths (8–10). Therefore, the use
of high-purity reagents for filtration will help reduce or
eliminate product contamination and degradation issues
further downstream.

The industry demand for high purity and high perfor-
ance reagents has led to the development of a new gen-
eration of powdered media. These products, sold under the
trade name Celpure® (Advanced Minerals Corporation,
Calif.) offer greater filtration capacity with a correspond-
ing reduction in powdered media consumption and reduc-
tion in disposal costs (Fig. 4).

Filtration processes with Celpure grades typically use
less media because of a combination of the higher solids
loading capacity and improved flow properties of the me-
dia. This results in a reduction in overall processing times
compared with conventional grades of diatomaceous earth,
such as the widely used Celite® grades (Fig. 5). Finally, the
products are extremely pure and have very low levels of
extractable metals such as aluminum and iron (Table 1)
and correspondingly low electrical conductivities.

Filtrate Clarity

Most processes have a clarity specification that needs to be
met or exceeded. Achieving stringent clarity specifications
can extend the life of downstream filters and protect chro-
matography columns. The filtrate clarity achieved is dic-
tated by the grade selected and the nature of the turbidity
removed. Once the grade is chosen, its level of usage (body
feed addition) combined with the available differential
pressure to induce flow will control the volume of unfiltered
feed that can be processed by a given filtration area.

Product Throughput

Product throughput and filtrate clarity are tightly linked
when it comes to grade selection. The goal is to select a
grade that achieves the desired clarity and maximizes
throughput. By selecting a grade that is too fine, the clarity
specification can be exceeded, but the throughput rate may
be extremely low with correspondingly high differential
pressures.

Operating Pressure

Actual operating pressure can be limited by shear sensi-
tivity of the product or feedstream by equipment con-
straints or by overall plant design. Many processes have
mechanical pressure limits of approximately 30 to 45 psi
or 2 to 3 bar.
Product Recovery

This can be sometimes overlooked in initial filtration studies. At the conclusion of any filtration cycle, the accumulated solids should be washed with a product-compatible buffer to maximize product recovery. Upon completion of the filtration cycle (including rinse), it is important to confirm that the product of interest does not interact with the filtration media. Product recovery issues can often be improved by selecting a more permeable grade of filter aid. As you increase the porosity of a filter aid, the available surface area decreases and any nonspecific interaction decreases. All these factors are tightly linked; therefore, we recommend a systematic approach to grade selection and optimization.

SYSTEMATIC METHODS DEVELOPMENT APPROACH TO GRADE SELECTION

As stated above, the overall goal is to achieve optimum clarity and maximum throughput for a process while minimizing pressure and product losses. The following is a systematic approach to grade selection. Figure 6 shows a typical experimental setup for laboratory-scale constant rate filtration experiments. To perform constant pressure filtrations, a pressurized vessel is required to deliver unfiltered feed to the filtration housing.

Determine the Percent of Solids in Your Feedstream

Any information on distribution and size of feed solids is useful in grade selection. If this information is not available, a recommended procedure is outlined below.

Determine percent solids by weight in your fermentation broth. Filter a small aliquot onto a preweighed filter disk, dry in an oven, and reweigh to determine percent solids in unfiltered feed.

1. Use a Whatman 934-AH filter paper as a septum with the filtration housing. This is an open-pore septum and enables you to precoat without having to custom make a wire screen.

2. Resuspend 1.5 g of media in 100 to 150 mL of buffer (similar in pH to your sample solution; 1.5 g of most grades gives a precoat of approximately 2 mm on 20 cm² of surface area). Start with the coarsest grade, Celpure® 3000 or Celite® 545 AW. By starting with the coarsest grade, you will always achieve maxi-
Figure 5. Equal weight comparison of Celpure and Celite grades in the clarification of a fermentation broth with 0.5:1.0 body-feed-to-solids ratio. Runs were terminated at minimum (equal) flow for reasonable production.

Table 1. Typical Celpure® Properties (Applicable to All Grades)

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soluble aluminum (mg Al/kg)</td>
<td></td>
</tr>
<tr>
<td>Extraction in fermentation broth, pH 4.5</td>
<td>≤3</td>
</tr>
<tr>
<td>Soluble iron (mg Fe/kg)</td>
<td></td>
</tr>
<tr>
<td>Extraction in fermentation broth, pH 4.5</td>
<td>≤3</td>
</tr>
<tr>
<td>Conductivity in H₂O (μS-cm)</td>
<td>≤12.5</td>
</tr>
</tbody>
</table>

4. Once the precoat has been formed, reduce the volume of buffer in the chamber until 2 to 3 mm above surface of precoat.
5. Add media to the unfiltered feed solution. As a starting point, match the percent of solids in your feed with powdered media addition (2% solids by weight; add 2% by weight of powdered media). Add product/media solution to filter chamber at a flux rate of 1 to 5 L/m²/min. On a 20-cm² surface area, this is equivalent to 2 to 10 mL/min.
6. Measure initial flow rate and monitor both flow rate and pressure throughout the course of the filtration. Allow the pressure to rise to 30 psi or 2 bar (or maximum system pressure), and maintain this pressure for the duration of the filtration by reducing flow or until flow rate reaches 20% of initial flow.
7. At the end of the filtration cycle, wash the filter bed with two bed volumes of compatible wash buffer. This will maximize the product recovery for a process. Finally, measure the total volume processed, the time of filtration, filtrate clarity, and the bed height in the filter housing. Save a sample for product recovery analysis.

If clarity is not acceptable, then repeat steps 1 to 7 with a less permeable grade of filter aid. Powdered media are rated based on their Darcy permeability. Celpure 3000 and Celite 545 AW are considered the coarsest grades and Celpure 65 and Celite Filter-Cel® the finest.

For constant pressure filtrations, follow above procedure through step 4. Then switch to a pressurized vessel containing unfiltered broth mixed with appropriate grade of powdered media. Ramp the pressure up to the desired output and monitor decay in flow rate. Repeat experiments with different grades as outlined above.

**Body Feed Optimization**

Once grade selection is completed, the next step is to optimize the body feed addition. The level of body feed addition controls the permeability and volume of accumulating cake. Higher pressure does not mean greater throughput, especially when body feed usage coupled with the handling characteristics of the solids dictate filter cake

![Figure 6. Experimental setup for constant rate experiments.](image-url)
compressibility. The dilatant nature of diatomite resists cake compression and thereby retains permeability. Feed properties (pH, viscosity, etc.) also affect filtration performance; therefore, these factors must be considered in process development and optimization studies.

Past work with biological solutions at high solids levels (greater than 5% by weight) suggests that constant flux (filtration) rates above 10 L/m²/min exhibited steep rises in differential pressure because of compression and subsequent loss of cake permeability. Modest flux rates (5.0 L/m²/min) extended filtration cycles, often producing better economy. Although dilution using a compatible solution can reduce shear and also make filtration easier, the goal to expedite processing implies that it will be used only after other options have failed.

As described earlier, an excellent starting point is to match the percent solids in the process stream on a weight basis with percent media also on a weight basis. If this ratio of 1:1 is not sufficient, i.e., the filtration terminates quickly because of a rapid rise in pressure and subsequent blinding of filter cake with solids, then consider increasing your body feed addition. Repeat the above steps using body feed additions of 1.5:1 or 2:1.

If a body feed addition of 1:1 allows you to process your batch with little or no rise in differential pressure, then you can optimize the body feed by reducing the ratio of powdered media to solids. Reduce the body feed until an acceptable throughput rate is achieved (Fig 7). This approach will allow you to size an appropriate filter and minimize the size of filter needed to process a given batch.

**Process Scale-Up and Filter Selection**

Candle elements, pressure filters, and filter presses are all used in biopharmaceutical applications. Your choice of filter will depend on a number of factors, including plant design, equipment compatibility with existing systems, clean-in-place and steam-in-place considerations, scale of operation, and overall capital costs. All these factors need to be considered in selecting an appropriate filter. Once the filter design has been selected, the data generated in the methods development stage can be used to predict filter size. A smaller filter can be considered if the cycle time is short, and multiple cycles (including time for cleaning and setup) can be performed, in a typical 8 hour shift, to clarify the batch.

**SUMMARY**

The use of advanced filtration media in dynamic filtration systems is a flexible, efficient, and attractive means for the solid–liquid separation of microparticulates from various process streams. Powdered media provide versatility, high solids-loading capacity, high product recovery, low cost, and ease of scale-up in any filtration process. These systems are cost effective and extremely flexible and efficient at removing suspended solids and clarifying supernatants rich in colloids and small particles. Newer products such as Celpure™ demonstrate improvements in purity and performance and further address the needs of the biopharmaceutical industry.

In addition to the advances that have been made in powdered media, improved systems for media handling and containment are now available from vendors that readily meet the needs of the industry. Finally, a wide variety of economical filter housings of various designs, including candle elements, pressure filters, and filter presses, are manufactured to biopharmaceutical specifications.

**BIBLIOGRAPHY**


See also **CELL SEPARATION, SEDIMENTATION; MICROENCAPSULATION.**
INTRODUCTION

In the bioprocessing industry there exist a number of applications for air (or gas) filtration (1–3). The type of filter used for the application will depend on the specific removal requirements. The filtration required can be to the 0.2-l (sterilizing) level or to a coarser filtration level for particulate removal. Some air applications for which sterile filtration can be a requirement are fermenter inlet air, fermenter vent gas, vents on water-for-injection tanks, and vacuum break filters during lyophilization. Course filtration can 90 be used for particulate removal. In many cases, the course filter acts as a prefilter to the sterilizing grade filter. Other applications use coalescers as prefilters for the removal of liquid droplets, such as oil or water.

TYPES OF AIR FILTERS

Packed towers were the first air filters used by the industry for air sterilization. Packed towers consist of beds or pads of a fibrous material, such as paper, cotton wool, glass wool, or mineral slag wool. Membrane filters typically are constructed of a porous, hydrophobic membrane material, such as PVDF (polyvinylidenefluoride) or PTFE (polytetrafluoroethylene).

When filters are used to produce sterile air, the filters can be referred to as sterilizing grade, 0.2- filters. Microbial retention tests can be used to verify that the filters produce sterile air. The standard test organism is Brevundimona (Pseudomonas) diminuta (ATCC 19146).

The removal efficiency of filtration media used in packed towers or membrane filters can be dependent on the following mechanisms:

1. Direct interception by the fibers
2. Inertial impaction
3. Brownian motion or diffusional interception
4. Electrostatic attraction between the fibers and particles

Brownian motion (also referred to as diffusional interception) applies to small particles at low face velocities. When air molecules are in a state of random motion, small particles suspended in the airstream can be struck by moving air molecules and displaced. The movement of particles resulting from molecular collisions is known as Brownian motion. This phenomenon can increase the probability of capture of the particle by the fibers in a packed tower or by the diffusional interception in a membrane.

The filtration mechanism for direct interception is a sieving action that mechanically retains the particles. The filter material acts as a screen that stops particles that are larger than the pores. Direct interception is independent of face velocity and mostly involves particles with large diameters.

Inertial impaction refers to the deviation of a particle from its streamline during flow because of inertia. This deviation can result in the retention of the particle by the fibers in the packed tower or by the membrane in a membrane cartridge filter. As the face velocity increases, the probability of inertial impaction increases.

An electrostatic attraction can exist between the fibrous material in the depth filter, or the membrane material, and the particles in the air stream. This attraction can enhance the ability of the filter to remove particles.

A general description of packed towers as well as membrane filters used for air filtration is provided in the next section.

Packed Towers

Description of Packed Towers. Packed towers for air filtration are comprised of beds of fibrous material (1). The diameter of the fibrous material is between 0.5 and 15 , and the space between the fibers can be many times this range, depending on how the tower is packed.

Figure 1 illustrates the basic design of a packed tower air filter (5). The filter consists of a steel structure filled with loose fibrous packing. The air inlet is on the bottom and the outlet is on the top of the filter. The packing is supported by a grid or perforated plate. For proper opera-
Filtration, Air

Support grid
Stainless steel casing
Fibrous filter packing
Air inlet
Air outlet

Figure 1. Packed air tower design.

It is necessary to ensure that the appropriate packing density for the application is obtained. If the tower has been packed properly, there will not be movement of the fibers during use. Fiber movement during use can lead to channeling of air, which will lead to an inefficient packed tower, because only a portion of the bed will be acting as a filter. Fiber repositioning can also lead to the dislodging of trapped microorganisms.

Once the filter has been packed with the appropriate fibrous material, a support grid or plate is fitted to the top of the bed to ensure that the bed remains compressed. After the first steam sterilization of a packed tower air filter, the packing will tend to settle further. Additional fibers can be added to the bed after the first steam sterilization to maintain the correct packing density.

Bonded fiber mats have been developed to use in place of loose fibers. Resins are used to bond the fibers, and the mats must be tested to determine if they are compatible with steam sterilization. When mats are used, it is necessary to have a good seal between the mat and the tower wall, so that channeling does not occur. Fiber pads often contain mineral wool with an average fiber diameter of 4 to 5 μm.

Thin sheets of small diameter fibers are another option for packed towers. The sheets are placed on top of each other with a mesh or grid in between each sheet for support. The edges of the sheets are sealed between flanges.

Sterilization of Packed Tower Air Filters. Before using a packed tower for the sterilization of fermenter inlet air, the filter itself must be sterilized. There are two techniques that can be used: steam sterilization and dry heat sterilization (6).

Passing steam at a pressure of 15 psig through a packed tower for 2 h is typically adequate for sterilization. Actual conditions required for a specific installation may vary. The presence of air in the packed tower during the sterilization can prevent complete sterilization of the packing. A drain at the bottom of the packed tower is used to purge the steam and to drain any residual condensate in the filter. It is necessary to remove condensate from the bed because wetted fibers are less efficient for microorganism removal and may also decrease the retention efficiency of the packed tower well below its design value, especially if channeling through the wetted media occurs.

Some fiber material as well as material used to bond fibers can be degraded by steam sterilization. An alternative to steam sterilization is dry heat sterilization, which will avoid the possibility of steam degradation and fiber wetting. This can be accomplished by using a heating device at the inlet of the tower and passing air at a temperature of 160 to 200 °C through the bed for 2 h. During dry heat sterilization, the filter is isolated from the rest of the process. Fiber material can be stable up to 800 °C, thus there is no chance of damage to the bed from this sterilization technique.

Operating Considerations for Air Sterilization by a Packed Tower. The microorganism retention efficiency of packed towers is dependent on the inlet velocity of the air. The relationship between air velocity and filtration efficiency has been determined experimentally in a number of studies (7). In the example shown in Figure 2 (7), the particle retention efficiency may change by a factor of 10 as a result of a relatively small change in the inlet air velocity. The results also show that the most difficult particles to remove are in the size range of small microorganisms. Therefore, it is quite possible to encounter air flow conditions in a packed tower that can reduce the statistical probability for the complete retention of all microorganisms.

The hydrophilic nature of the fibrous material used in packed towers (e.g., glass wool) can contribute to a reduction of the microorganism removal efficiency of the packed tower. Water vapor can enter the system with the air discharge of the compressor. The air and water vapor mixture is initially at an elevated temperature. As the gas stream is cooled, water droplets may condense and wet the hydrophilic glass fibers. The wetted fibers are less efficient for microorganism removal and may decrease the retention efficiency of the packed tower. A drain at the bottom of the packed tower is used to purge the steam and to drain any residual condensate in the filter. It is necessary to remove condensate from the bed because wetted fibers are less efficient for microorganism removal and may also decrease the retention efficiency of the packed tower well below its design value, especially if channeling through the wetted media occurs.

Some fiber material as well as material used to bond fibers can be degraded by steam sterilization. An alternative to steam sterilization is dry heat sterilization, which will avoid the possibility of steam degradation and fiber wetting. This can be accomplished by using a heating device at the inlet of the tower and passing air at a temperature of 160 to 200 °C through the bed for 2 h. During dry heat sterilization, the filter is isolated from the rest of the process. Fiber material can be stable up to 800 °C, thus there is no chance of damage to the bed from this sterilization technique.
efficiency of the packed tower well below its design value, if channeling through the wetted media occurs. Also, organic components present in the compressor exhaust can provide a nutrient source for the retained microorganisms and increase the possibility of bacterial growth and eventual penetration through the depth filter medium. Although a number of approaches have been tried to overcome the wetting problem, such as heat tracing to maintain an elevated temperature, these are expensive to operate and have not consistently resolved the problem. The pressure differential across the tower may also increase significantly with wetting.

Although it is obvious that operating deviations such as these do reduce the reliability of a packed tower for air sterilization, there is no quantitative procedure to determine if the filtration efficiency is adequate to ensure a sterile inlet air condition. The lack of such a quantitative procedure adds an element of uncertainty that affects both the operation and maintenance decisions. Skilled operators are necessary to monitor the packed tower to ensure that it is packed and operating efficiently. The actual source of contamination of a production batch is often difficult to identify, and the air filtration system is always suspect during a contamination outbreak. Without a technique for testing the efficiency of a packed tower, it is difficult to ascertain whether the packed tower is the cause of the contamination outbreak.

**Design of Fibrous Air Filters.** In order to model a fibrous air filter, several assumptions must be made. It is assumed that (1) once a particle is trapped by a fiber, it then will remain trapped; (2) at a particular depth across the filter, the particle concentration does not vary; and (3) the removal efficiency at a given depth is equivalent across the filter.

The following equation describes how the concentration of particles varies with the depth position in the filter (6):

\[
\frac{dN}{dt} = -Kx
\]

where \(N\) is the particle concentration, \(x\) is the depth, and \(K\) is a constant.

Solving the equation between a depth of 0 and \(x\) and a particle concentration of \(N_0\) particles entering the filter and particle concentration of \(N\) particles leaving the filter yields:

\[
\ln\left(\frac{N}{N_0}\right) = -Kx
\]

The relationship between depth and the logarithm of the ratio of particles removed to particles incident is known as the log-penetration relationship (6). This relationship has been used in sizing depth filters.

The constant \(K\) in equation 2 will vary with the type of packing and is dependent on the linear velocity through the packed bed. If the relationship between the constant \(K\) and linear velocity through the bed is known, equation 2 can be used to size a packed bed for a given log reduction of particles.

Another consideration for the design of a packed bed is the pressure drop across the bed. Typically, the \(\Delta P\) is linear with the linear air velocity for a given depth (6). The pressure drop across the packed tower can be dependent on the type of medium, the packing density, the air density, and the linear air velocity through the filter. As an example of an equation for the pressure drop across the packed bed is given in Richards (6):

\[
\Delta P = \frac{2\rho v^2 \alpha C}{\pi D_f}
\]

where \(\Delta P\) is the pressure drop, \(v\) is the linear air velocity, \(\alpha\) is the ratio of filter density over fiber density, \(x\) is the filter bed depth, \(C\) is the drag coefficient, \(D_f\) is the fiber diameter, and \(\rho\) is the air density. The relationship in equation 3 indicates that the pressure drop is proportional to the square of the linear air velocity. At relatively low linear air velocities (2 to 3 ft/s), the relationship is linear.

**Membrane Filter Cartridges.**

Membrane filter cartridges are available as either prefilter (particulate contaminant rating) or sterilizing filter (bacterial contaminant rating) configurations (1–3). Prefilters in air service can be used for particulate removal or aerosol removal. Prefilters are positioned upstream of the final (sterilizing) filters to protect the final filter from premature plugging, thereby prolonging the life of the final filter. The following are brief descriptions of the membrane filters (sterilizing and prefilter types) that can be used for air filtration.

**Sterilizing Grade Membrane Filters for Air Service.** Membrane filters used for fermenter and bioreactor sterile inlet air and exhaust gas vents, sterile pressure gas, sterile nitrogen blankets, storage tank sterile vents, formulation tank sterile vents, and sterile air for aseptic packaging usually contain a membrane made of hydrophobic materials such as PVDF or PTFE.

Hydrophobic membrane filters are desired in these sterile gas filtration applications because hydrophobic filters do not spontaneously wet with water. When a hydrophilic filter is wetted with water, it will not pass air until the water-wet bubble point of the filter is exceeded, and this water-wet bubble point can be greater than 50 psi. The inherent hydrophobicity of membrane filters used for fermenter air sterilization allows these filters to be able to remove bacteria completely from inlet air, even when exposed to moisture (8). The filter cartridge is expected to remove bacteria and bacteriophage from air streams with 100% efficiency.

The hydrophobic membrane filter material is pleated together with a layer of material (for example, nonwoven polypropylene) on the upstream and downstream side of the membrane. These layers provide mechanical support to the membrane and proper drainage of the fluid. The pleated membrane pack is formed into a cylinder. The longitudinal side seal of the pleated membrane filter pack should be an integral, homogeneous melt seal without any additional or extraneous materials, glues, or resins.
A rigid, perforated inner core is present to provide support against operating pressure. An outer cage placed on the upstream side of the membrane filter pack is provided for additional support and protection during handling. The cage provides retention of structural integrity against accidental reverse pressure (usually up to about 50 psi). Polypropylene and PTFE are examples of materials that can be used for the cage, core, and support material in a membrane filter.

End caps are attached by melt sealing to imbed the medium in the plastic. All hardware should be specified as produced from virgin materials.

Membrane filters are available in a variety of shapes and sizes. The most typical configuration for pharmaceutical air applications are 10-in. elements. These elements can be flame welded together end to end to form multilength configurations (typically up to 40 in.). Multiple elements can be used in a filter housing, as required. A membrane filter is illustrated in Figure 3.

Membrane filter elements may be integrity tested to ensure that the air is sterile. The integrity tests have been correlated with bacteria removal efficiency by direct microbial challenge testing.

**Sterilization of Membrane Filters.** Hydrophobic membrane process filter cartridges can be subjected to multiple sterilization cycles and must be designed to be repeatedly steam sterilized in either direction of flow or repeatedly autoclaved. The vendor should provide information on filter cartridge sterilization procedures and operating data on process limitations (time and temperature). These filters are typically capable of withstanding multiple in situ steam sterilizations (for example, at least 165 cumulative hours at up to 142 °C). If needed, the membrane filter can be autoclaved and aseptically installed into the process application.

Hydrophobic membrane filters can be in situ steam sterilized in 30 min of exposure time (or longer if the process requires a longer time) at a temperature of 125 °C. Because the filters are hydrophobic, no additional drying time is required. It is necessary for membrane filters as well as for packed towers to drain the entrained condensate on the inlet side of the filter. A membrane filter housing can be isolated under pressure at the end of an aeration period and therefore does not require sterilization for each cycle.

**Prefilters.** For some applications, it is necessary to use a prefilter to remove particulate material or liquid aerosols. It is often desirable to use a prefilter to protect the sterilizing grade filter. Typically, the use of a prefilter will reduce the overall filtration economics.

**Particulate Removal.** Membrane filters composed of materials such as polypropylene or cellulose can effectively be used to remove particulate material from an air stream. The filter micron ratings range from the order of 1 to the order of 100 μ. The appropriate filter can be selected for the particulate contaminant removal. The following are descriptions of examples of membrane filters that can be used as prefilters in pharmaceutical air filtration, including porous stainless steel filters, cellulose pleated filters, and polypropylene pleated filters.

1. Porous stainless steel filters cartridges. Porous stainless steel medium is made by sintering small particles of stainless steel or other high alloy powder together to form a porous metal medium. Porous stainless steel can be formed as a flat sheet or, when used in filter elements, as a seamless cylinder. This special manufacturing process produces a high dirt-carrying capacity medium that is temperature and corrosion resistant. The recommended alloy is type 316LB, which has a higher silicon content than type 316L and provides a stronger, more ductile product with better flow properties. Elements have absolute ratings of 0.4 to 11 μ in gas service applications. Porous stainless steel filter cartridges are chemically or mechanically cleanable, offering economy of reuse. Porous stainless steel filters are used for steam filtration and for air sparging.

2. Cellulose pleated filter cartridges. Pleated cellulose filter cartridges (8 μ rating) are applicable as prefilters for inlet air for fermentors and bioreactors. These filter cartridges can be constructed of pure cellulose medium, without resin binders. The cellulose membrane can be pleated into a high area cylinder and has a longitudinal side seal with an appropriate polypropylene (approved for food contact usage by the FDA). Cellulose media cartridges are assembled with hardware components consisting of a perforated inner support core, an outer support cage, and end caps melt sealed to imbed the medium in the plastic. All hardware components should be of pure
polypropylene, without filler or reinforcement, to ensure a minimum of soluble extractables.

3. Polypropylene pleated filter cartridges. Polypropylene pleated depth filters (8-µm rating) are applicable as prefilters for prefiltration of exhaust gases. These process filter cartridges are constructed using non-migrating continuous strands of nonwoven polypropylene filaments. The medium should have a constant pore size section (downstream) for absolute rate filtration and a continuously graded upstream section for effective prefiltration. The thin sheet of polypropylene media is pleated and formed into a cylinder with a longitudinal side seal of melt seal polypropylene. The cylinder is then melt sealed to injection molded polypropylene end caps to ensure no fluid bypass. Polypropylene hardware components consisting of an inner support core and an external protective outer cage are incorporated.

Liquid Aerosol Removal. A coalescer can be used for the removal of liquid aerosols containing water or oil droplets. This is desirable as a prefiltration for a sterilizing air filter, because the liquid aerosol could prevent the flow of air through the sterilizing grade filter.

Coalescers operate efficiently if they are able to separate the liquid and the gas in the liquid aerosol. The three basic steps that are required are (1) aerosol capture (2) unloading or draining of the liquid, and (3) separation of the liquid and gas.

Figure 4 is an illustration of a liquid–gas coalescer. The coalescer has a gravity separator, which allows for the re- or draining of the liquid, and (3) separation of the liquid and gas.

Filter Housings. Cartridge filters used for air filtration are placed in sanitary, air or gas service filter housings. Cartridge filter housings usually consists of a head and a bowl. The filter is attached to the head, and the bowl is clamped to the head to provide a complete enclosure of the filter.

The size of the housing should be adequate for the flow and differential pressure requirements. Filter housings for bioprocessing applications are typically constructed of stainless steel (e.g., 304, 316, 316L, etc.) or carbon steel, with 316-series stainless steel internal hardware and cartridge seating surfaces. Housings typically have quick-release mechanisms such as V-band clamps or fast-action swing bolts to facilitate filter change-outs.

Design operating pressure of all filter housings should be specified as minimum psig and rated for full vacuum service. Design maximum operating temperature of the housing should also be specified. The housings or pressure vessels that are within the scope of the ASME Boiler and Pressure Vessel Code, Section VIII, Division 1, should be designed and U stamped per the code. TIG weld construction should be used in sanitary-style housings to minimize weld porosity and ensure high-quality, leak tight joints, with all internal welds ground smooth and flush. All weld procedures and welders should be qualified to ASME/BSI Section IX.

Housings should be capable of in situ steam sterilization in accordance with the manufacturer’s recommended procedures, and housing or system design should provide for condensate drainage. Gasket material and O-ring elastomers must also be capable of withstanding repeated steam sterilization cycles, along with being compatible with process fluids.

Industrial style housings (used for prefilters) provide cartridge mounting on a tie rod and sealing to the tie rod assembly by use of a seal nut at the top of the assembly. Tube sheet adapters should be seal welded to the tube sheet to prevent fluid bypass. Filter cartridges are thereby sealed in the housing independent of any cover assembly, ensuring positive sealing and no fluid bypass. Filter cartridges should be seated on the tube sheet adaptor assemblies above the tube sheet to ensure complete drainage of nonfiltered fluid before cartridge replacement. This prevents potential contamination of downstream surfaces during change-out of filter elements.

QUALIFICATION TESTS FOR MEMBRANE FILTERS USED FOR AIR FILTRATION

Sterilizing Grade Membrane Filters

Organism Retention Tests. Microorganism retention tests can be conducted to verify that membrane filters produce sterile air (1,2). Liquid challenge tests with Brevundimonas (Pseudomonas) diminuta (ATCC 19146), measuring 0.3 × 0.6 to 0.8 µm, is a standard challenge test for the validation of sterilizing grade filters (0.2 µm) in the pharmaceutical industry. Aerosol challenge tests with P.
Liquid droplets in air flow

Small liquid droplets coalesce to form large drops

Figure 5. Coalescence of droplets.

diminuta should approximate extreme air flow conditions. Aerosol challenges with $T_1$ bacteriophage (0.05 $\times$ 0.1 $\mu$m), for example, can provide a test of a filter’s retention efficiency of extremely small organisms, so small that the bacteriophage in a liquid suspension will penetrate a 0.2-$\mu$m sterilizing grade filter.

The retention efficiency of a given filter is less when a liquid challenge is used instead of an aerosol challenge. Thus, a liquid challenge test is a more stringent test of a filter’s retention capability. A liquid challenge test can also provide retention information for process conditions such as extreme moisture after sterilization or air entrained with water drops. An example of the technique used to perform a liquid bacterial challenge on sterilizing grade membrane cartridges was published by Pall Corporation (9).

The liquid test involves challenging a test filter with a known quantity of $P$. diminuta, no less than $1 \times 10^7$ organisms per 2 square centimeter of filter area. The challenge sample is passed through the filter suspended in sterile water at a defined flow rate and time. All of the effluent from the test filter is passed through an analysis membrane. After the challenge is completed, the system is flushed for 5 to 10 minutes with sterile water. The analysis membranes are removed and placed on Mueller Hinton Agar at 32°C for 48 h. After incubation, the plate is examined for the presence or absence of microbial colonies. Figure 6 illustrates a liquid bacterial challenge test stand.

The aerosol challenge test system can consist of a nebulizer loaded with the challenge microorganism suspension, a separate line for dry air makeup, and split-stream impingers to sample the aerosol challenge with and without the test filter. A schematic of a test stand that can be used for the aerosol challenge procedure described is given in Figure 7. Detailed descriptions of the aerosol test protocols for $P$. diminuta and $T_1$ bacteriophage Ref. 9.

During the aerosol challenge, an aerosol is generated with a nebulizer. The aerosol is introduced into the test filter at a given flow rate. The filter effluent is collected in dual liquid impingers. Controls are performed simultaneously via a split stream by using a two-channel timer to direct air flow, on an alternating basis, from the test side filter impingers to the unfiltered control side impingers for recovery.

The impingers contain sterile buffer and after the challenge is completed, the buffer can be analyzed for the test organism. If $P$. diminuta is the test organism, then the buffer is analyzed by putting the buffer solutions through an analysis membrane and placing the membrane on Mueller Hinton Agar for 48 h before titering. If $T_1$ bacteriophage is the test organism, then samples of the buffer are diluted with nutrient broth and mixed with liquid nutrient agar (1.5% agar concentration; 48°C) and Escherichia coli in the log phase of growth. After vortex mixing all three components, the mix is poured over nutrient agar plates and incubated for 3 to 5 h at 37°C, so that the plaques can be counted.

Integrity Tests for Membrane Filter Cartridges. During the sterilization of fermenter air, it is necessary to achieve the highest possible assurance of filter integrity and removal efficiency. The installation of integrity-testable filters and the performance of routine integrity testing by the user are essential to demonstrate that the system is performing to specification. Tests that qualify the retention characteristics of a membrane filter can be defined as destructive or nondestructive tests.

Destructive tests are performed using an appropriate contaminant to meet a specific claim for retention of the contaminant. The test procedure must be sensitive enough to detect the passage of contaminants of interest. For sterilizing grade 0.2-$\mu$m membrane filters, the industry standard test organism (i.e., contaminant) is $P$. diminuta (ATCC 19146). The organism and minimum challenge level (10$^7$ CFU/cm$^2$ filter area) are specified in the ASTM standard F 383-83 (10) and referenced in the FDA guideline on sterile drug process produced by aseptic processing (11).
Figure 6. Liquid bacterial challenge test stand.

Figure 7. Aerosol challenge test stand.
Because most filter users would not want to perform a destructive test in a process environment and because the filter is not usable after the destructive test, nondestructive tests related to the retention results of the destructive test are used instead. From the relationship developed between a nondestructive and a destructive test, membrane filter performance can be safely and conveniently verified in the production environment. The relationship between a nondestructive integrity test and the assurance of bacterial retention constitutes a filter validation study and is extremely important for microbial retentive filters used in critical fluid processes.

There are four nondestructive integrity tests that can be used for sterilizing grade hydrophobic membrane filters used for air filtration: (1) bubble point test, (2) forward flow test, (3) pressure hold test, and (4) water intrusion test. These tests are described in Filtration, Cartridge.

For a nondestructive test to be useful, the results of the test must be able to predict the ability of the filter to remove bacterial. The manufacturing quality control of the membrane filters tested to obtain this prediction of bacterial removal for the filters must be used to establish quality control procedures that are consistently maintained in the manufacturing of the filter. The combination of the nondestructive integrity test, the correlated bacterial challenge, and manufacturing quality control allows membrane filters to be used reliably for sterile filtration.

**Tests for Prefilters**

Many filter manufacturers use a nominal micron rating for removal efficiency. A nominal rating is an arbitrary micron rating assigned by the filter manufacturer. Such ratings are subject to a lack of reproducibility.

An alternative method for rating filters is the Oklahoma State University (OSU) F-2 Test. This rating method (ISO 4572, ANSI B93.31) has received wide acceptance for use on lubricating and hydraulic fluids. Pall Corporation, for example, uses this method for oils extensively and has adapted it for use in water with contaminants ranging from 0.5 to 25 μm.

The test is based on continuous on-line particle counts of different particle sizes, both in the influent and the effluent. The β ratio at a specific particle size is defined as βₜ₀: the number of particles of a given size (X) and larger in the influent, divided by the number of particles of the same size (X) and larger in the effluent, where X is the particle size in microns. The percent removal efficiency can be calculated from the β-value. The percent removal efficiency is [(βₜ₀ - 1)/βₜ₀] 100 (12).

As an example, Pall prefilters are given a micron rating that corresponds to a 100% removal efficiency or the value in microns at which the OSU F-2 Test gives a β-value of more than 5000.

**AIR FILTRATION APPLICATIONS**

Fermentation and downstream processing are two major operation categories in bioprocessing. During both fermentation and downstream processing, sterilizing-grade 0.2-μm hydrophobic membrane filters can be required for processing air (or gas) streams. Sterilizing-grade 0.2-μm hydrophobic membrane filters are used during fermentation for the sterilization of fermentation inlet air and for the filtration of fermentation exhaust gas. During downstream processing, sterilizing grade filters can be found in use as sterile tank vents. In a final purification, membrane filters can be used for vacuum breaking in processes such as lyophilization.

As discussed above, hydrophobic membrane filters are desired in sterile gas filtration applications because hydrophobic filters do not spontaneously wet with water.

Sterile air filtration applications discussed in this section include fermenter inlet air, exhaust gas (vent), high temperature air filtration, and lyophilizer vacuum break filtration. Special attention will be given to the specific requirements for a sterilizing filter when used for these applications. After the discussion of the applications is a general discussion of recommendations for sterilizing-grade filter usage.

**Fermentation**

During a fermentation process a specific cell (yeast, bacteria, or mammalian) is grown to provide a desired product (1–3). Products can include cells, antibiotics, amino acids, or recombinant proteins. There can be a variety of sizes for the fermenter, ranging from very small (100 L or less) cell culture reactor to very large scale antibiotic production (100,000 L). In these applications, there is often the need to maintain sterility in both liquid and gas (air or nitrogen) feeds to support growth of the desired cells.

Air filtration applications for fermentation are illustrated in Figure 8. Those applications specific to fermentation are described here, and filtration of utilities used in fermentation such as steam, air, and water are discussed in the “Utilities” section.

**Prefiltration of Fermentation Air**

Compressors are often used to generate air flow for the manufacturing facility. There are two types of compressors, oil free and oil lubricated. In older facilities where oil-lubricated air compressors are commonly used, prefiltration of inlet air is necessary for removal of oil droplets. A coalescing filter can provide greater than 99.9% removal of oil and water droplets in the 0.01- to 0.5-μm range and larger. This also acts as an excellent prefilter for the hydrophobic membrane pleated filters that are commonly used for sterilizing the inlet air to the fermenter. The typical gas flow rate per 10-in. filter module is 200 to 400 standard cubic feet per minute (SCFM).

For oil-free compressors, a prefilter acts to remove dirt in the air system, extending the service life of the final filter. For use with fermentation air, a cellulose pleated filter with an absolute rating of 8.0 μm is normally the filter of choice. Alternately, polypropylene (2.5-μm rated) pleated filters also serve as excellent prefilters for this application. The typical gas flow rate per 10’ filter module is 75 SCFM.

**Sterile Air Filtration for Pilot and Production Fermenters and Feed Tanks**

One of the largest applications for sterile
Air filtration is the sterile air used for an aerobic fermenter during a typical production cycle. Typically, 1 vol of air per volume of broth per minute is used. Thus, for a 100,000 L fermenter on line for 48 h a total of 1.01 x 10^9 cubic feet of air requires sterilization.

The contaminants present in compressed air can include dust, lubricating oil, hydrocarbons, water, rust, and microorganisms including molds, bacteria, and viruses. Microorganisms in air are often associated with carrier particles, such as dust. Water and oil can be present as bulk liquid, vapor, or in aerosol. The air distribution system can give rise to contaminants such as rust and water. The concentration and size distribution of particles in compressed air are variable. The size range is generally between 0.001 and 30 µ, with a concentration between 10^-2 to 10^-4 g/m³ (4,6).

Bacteria and bacteriophage, when present in air feeds, can enter fermentation tanks or bioreactors and contaminate the product. Bacteriophage or other viruses can destroy the producing cells and reduce yields.

The process requirements to supply this sterile air can be quite restrictive. The air sterilization process must (1) process a large volume of compressed air, (2) provide a high degree of reliability, and (3) operate economically. Several methods have been considered for the sterilization of fermenter inlet air. These include filtration, heat, irradiation, washing with sterilizing chemicals, and electrostatic precipitation. Washing and electrostatic precipitation are not effective for the removal of microorganisms. Heat and irradiation are not economical. Filtration is the only technique that meets all the requirements for sterilizing fermenter inlet air.

Packed towers, an early filtration approach, were used widely in the industry. Since the early 1980s, filtration technology has advanced (13), and there has been an ongoing trend since the early 1980s to replace depth filters with hydrophobic membrane cartridge filters (e.g., Hoffman-La Roche [14] and E. R. Squibb and Sons [8]).

The recommended filters for sterilization of air feeds to fermenters and bioreactors are the hydrophobic membrane pleated filters. The hydrophobic (water repelling) nature of these membranes can provide for bacteria and bacteriophage removal with 100% efficiency under moist or dry operating conditions. This is an important benefit over fiberglass towers and cartridges. Filters for sterile air feeds should have a 0.2-µm absolute bacterial rating in liquids.
and a 0.01-μm particulate rating in air service. The typical gas flow rate per 10-in. filter module is 75 to 100 SCFM.

For some fermentations, the requirement may be for the filtration of fermenter air at an elevated temperature. If an application involves hot air and a longer service life is desired, then a filter that can withstand the elevated temperature is required. High Temperature Emflon® filters manufactured by Pall Corporation can be used in continuous service at a temperature up to 120 °C. These filters have a 0.2-μm microbial rating in liquid service and a particulate removal rating of 0.03-μm in gas service. The filter membrane is made of inherently hydrophobic PTFE, and the cage, core, and end caps are specifically designed for high temperature applications.

Sparging. Sparging acts to disperse air evenly in the fermenter or bioreactor containing the growth media and product. The product of choice for this application is a porous stainless steel sparging element, which can provide an exceptionally uniform and fine aeration gas dispersion. These elements are fabricated with one face of porous metal and one face of solid metal. If both surfaces of the sparging elements were porous, bubbles from the under surface may coalesce with bubbles from the top surface. Porous stainless steel sparging elements should be positioned horizontally in the fermentation tank, with the porous stainless steel facing upward. Fine grades of porous stainless steel (e.g., 3.0 μm absolute liquid rated) are ideal even for shear-sensitive mammalian cell cultures because of their high gas transfer and low shear aeration capability. Elements are typically available in standard and custom designs.

Exhaust Air-Off Gas Filtration. The purpose of a vent filter on a sterile fermentation tank is twofold: to prevent contamination of the tank and to provide containment of the material inside the tank. Prevention of contamination in the tank is desirable for processes that involve long fermentation cycles or require a sensitive fermentation medium (e.g., tissue culture medium). Genetic engineering techniques as well as fermentation of pathogenic organisms (such as organisms used for the manufacture of vaccines) have made it necessary to protect the environment and prevent the escape of microorganisms from the fermentation tank. The exhaust filtration system for a recombinant or mammalian cell fermenter or bioreactor must yield sterile air to the environment and provide a sterile barrier to prevent ingress of contaminants. Additionally, it must be in situ steam sterilizable and typically has a clean differential pressure less than 1 psid.

The removal efficiencies for simple depth filters (as described earlier) are typically poor under wet or variable flow conditions. Therefore, membrane filters are recommended for vent filtration applications.

The fermenter or bioreactor exhaust gas line can be contaminated with microorganisms or cells, growth medium components expelled from the fermenter or bioreactor as droplets or as solid particles, and aerosol condensate droplets formed during cooling of the gas in the exhaust system. These aerosol droplets, when present, can potentially block the final filter and must be removed before reaching the final filter. Mechanical separation devices, such as cyclones, condensers, and demisters, may not achieve effective aerosol removal below 5 μm (15). Removal efficiencies and pressure drops also vary significantly with flow rate in such equipment. Recent studies have shown that aerosols in exhaust lines are predominantly in the very fine 1-to-5 μm range (16).

The contaminants present will depend on the fermentation conditions, the growth medium, and the design of the exhaust gas system. The basic requirements for a vent filter are ability to provide sterility, a low pressure drop, and in situ steam sterilizable.

The recommended exhaust filtration system design entails two stages using a polypropylene pleated depth-filter cartridge as a prefilter to a 0.2-μm absolute rated hydrophobic membrane pleated filter cartridge. The purpose of the prefilter (typically 1.2-μm absolute rated) is to remove aerosolized particles and liquid droplets containing cells or growth medium from the fermentation off-gas or exhaust air. This serves to extend the service life of the final sterilizing filter. If the medium contains only fully dissolved components, such as with a sterile filtered cell culture medium, and if the fermentation is run at low temperatures (<30 °C) and low aeration rate (1 to 1.5 volume of air per volume of media per minute [VVM]), the prefilter may be optional. The typical gas flow rate per 10-in. filter module is 40 SCFM.

Like the final sterilizing filter, the pleated polypropylene prefilter should be multiple steam sterilizable. As additional benefit of the prefilter is to retard foam-outs from reaching the final sterilizing filter.

Sterilizing-grade 0.2-μm absolute rated hydrophobic membrane pleated final filters, with PVDF or PTFE membranes, can prevent organisms from entering or leaving the controlled reaction zone, even in the presence of water droplets and saturated gas (8,17). Steam sterilizability and integrity test values correlated to microbial retention studies under worst-case liquid challenge conditions provide the highest degree of assurance performance. Redundant systems using a second 0.2-μm rated sterilizing filter in series are recommended for high-risk recombinant organisms.

Condensate control is usually the most critical consideration for this application. In cases in which there is condensate accumulation and if the fermenter is operated with overpressure in the fermenter head, the amount of condensate accumulation can be reduced if a pressure control valve is placed at the fermenter exit, upstream of the exhaust gas filter. An alternative technique for the prevention of condensate accumulation is to use a heating section in the exhaust gas pipe upstream of the filter installation. This can be also be accomplished by specifying steam jacketing on exhaust filter housings. In this case, the exhaust gas temperature at the terminal filter must lie above the temperature of the exhaust gas at the fermenter exit. The heater must be properly sized based on the process parameters.

Downstream Processing

Starting with the cells and conditioned broth medium from the fermenter or bioreactor, the objective of downstream
processing can be to produce a highly purified, biologically active protein product, free of contaminants such as endotoxins, bacteria, particles or other biologically active molecules. This phase of bioprocessing typically comprises a series of unit operations including cell and cell debris separation, fluid clarification and polishing, concentration and purification, and membrane filtration sterilization of the purified product.

Cartridge filters are used in many stages of downstream processing, which involves filtration of both the harvest fluid and product intermediates as well as filtration of air and gases required throughout the process. Air filtration applications include vacuum break filters for lyophilizers, sterile nitrogen blankets, tank vents, and sterile air for container cleaning.

Absolute rated cartridge filters eliminate contaminants and impurities from air, nitrogen, and other gases used in downstream processing to prevent contamination of product and further protect concentration and purification equipment. Vent filtration ensures containment and freedom from product contamination during fluid transfer operations and protects processing equipment during sterilization cycles.

In fermentation, cartridge filters are used to maintain the sterility of the makeup water, feeds, additives, media in holding tanks, and in fermenter or bioreactor exhaust. Cartridge filters are typically used in downstream processing for the filtration of air, gases, and venting applications when it is necessary to vent tanks during fluid transfers; pressurize tanks using inert gases such as nitrogen and argon; protect vacuum lines, sterile vent holding tanks, and lyophilizers; and for gas purging, blanketing, drying, and when sterilizing equipment by in situ steaming or autoclaving.

The recommended filters for nonsterile particulate removal applications are polypropylene pleated filters. Hydrophobic membrane pleated filters such as PVDF or PTFE are recommended for aseptic processing. The absolute removal rating for the latter filters should be 0.2 μm determined under liquid flow conditions. The typical gas flow rate per 10-in. filter module is 75 to 100 SCFM.

In downstream processing applications, Cell and cell debris separation and clarification processes can be broken into a primary separation, secondary separation, and a cell concentrate section. During some primary separations, a cyclone can be used for particulate removal; a sterilizing air filter can be used as a vent on the cyclone. A variety of holding and receiving tanks can be used during the separation and clarification process; these tanks can be fitted with sterile vent filters.

During secondary separation, a nitrogen blanket may be needed; the nitrogen gas can be sterile filtered with a hydrophobic filter. Downstream processing can also involve the concentration and purification of clarified harvest fluid. Applications for air filtration include sterilizing-grade vent filters for solvent or buffer tanks and for holding or buffer tanks needed for ultrafiltration and chromatography.

The final pharmaceutical product will often need to be packaged. During filling processes, a sterile nitrogen blanket and thus a sterilizing grade hydrophobic filter may be needed. The final product is can be placed into a container; sterile air or nitrogen may be needed for container cleaning. A vent filter can be required on holding tanks. The typical gas flow rate per 10-in. filter module for tank vent applications is 75 to 100 SCFM.

Several specific applications for final processing are addressed in the next section.

Vacuum Break Filters. Sterilizing grade filters are used in freeze dryer installations to filter the gases used to maintain the chamber pressure and to break vacuum during operation and in sterilizers for vacuum break purposes.

Blow-Fill-Seal Equipment. Blow-fill-seal equipment can be used for the aseptic filling of pharmaceutical products. The container is formed and sealed aseptically. Air filtration is required to ensure sterility in this unit operation. A typical arrangement is illustrated in Figure 9.

For buffer-tank air hydrophobic membrane filters are used to supply sterile air to a buffer tank on the blow-fill-seal machine. This blanket air is used to drive the sterile solution through a pneumatically controlled dosing system. The air used in the buffer tank is referred to as the gas cushion or buffer tank air.

Hydrophobic membrane filters provide sterile air used to form the hot moldable plastic tube (parison). The air used to form the parison is known as the parison support air. The parison is subsequently blown molded into the shape of a ampule strip or a bottle. Typical requirements for the filters used include that the filters be steam sterilizable, integrity testable, and the proper size to prevent restriction of gas flow. Hydrophobic membranes are used to prevent wetting out and to maintain high flow rates even in moist conditions.

Utilities

There are a number of peripheral unit operations required during a sterile process. The air filtration applications for these applications are described in this section.

Water-for-Injection Tank Vent. Sterilizing grade membrane filters can be used in vent applications in which the fluid in the tank is at an elevated temperature. One such application is the vent used to prevent contamination in a water-for-injection tank. The water is at 80 °C or higher. When a sterilizing-grade air filter is used for this type of vent service, a steam-jacketed housing is typically used. It is only necessary to maintain the temperature of the filter cartridge at a temperature slightly above the dew point of the vapor. The steam introduced into the jacket should be at ambient pressure. Continuous operation of the jacket at a significantly higher steam pressure and temperature can reduce the service life of the filter caused by accelerated aging of the hardware by oxidation.

Steam Filtration. Process equipment and final filters are frequently sterilized by direct steam flow in situ, during the normal line sterilizing cycle. This eliminates the need for making aseptic connections and risking recontamination. Filtered steam is required for this sterilizing-in-place
of filters, piping, vessels, and filling equipment. Steam is also required for general equipment cleaning and sterilizing. The steam often contains significant amounts of pipe scale and other corrosion products. This particulate material should be removed in the interest of overall cleanliness and to avoid burdening the prefilters and final filters.

Particulate contamination in process steam is efficiently retained by porous stainless-steel filters with an absolute gas rating of 1.2 \( \mu \text{m} \). Porous stainless-steel filter assemblies are typically sized at steam flow rates of 30 to 40 ACFM per square foot of filter medium.

**GENERAL CONSIDERATIONS FOR OPERATION**

**Integrity Test Considerations and Guidelines**

Field experience with the integrity testing of sterilizing grade filters shows that various combinations of integrity test procedures are in use and that different testing schedules, both pre- and post-use, are followed. Perhaps the best approach is to perform the test or combination of tests that provides the highest degree of accuracy commensurate with the economics and the practicalities of the process.

Since the requirements for the use and testing of sterile gas filters can depend on location and application, several different types of filter tests can be used. Filter users typically use the forward flow or pressure decay tests, which require the use of a low surface-tension solvent, such as 60:40 isopropyl alcohol/water or 25:75 t-butyl alcohol/water. This is a well-proved, widely accepted approach directly correlatable to bacterial challenge (9).

Air filters are typically integrity tested when they are installed and retested periodically based on service conditions and operating requirements (e.g., once a month). Change-out schedules based on filter life studies have also been used in conjunction with integrity testing. Because filters in these applications are often sterilized in situ and can be damaged during in situ steam sterilization by reverse pressurization if the sterilization procedure is not properly controlled, filter life study data or a periodic integrity test regimen can be used only if the filter sterilization procedure is validated and in control.

In certain cases (such as in facilities in which the use of a nonflammable fluid is required, or when disposal of organic solvents is a concern), some filter users in aseptic processes have recently considered the use of water-based tests for hydrophobic filters.

There are a number of integrity tests possible for a hydrophobic air filter. The selection of the appropriate test and the appropriate test schedule depends on the specific application.

After an integrity test has been completed, it is typically desirable to remove the wetting fluid from the filter. This can be accomplished by blowing clean, dry (\(-40^\circ\text{C dew point}\)) air or nitrogen through the filter. It is necessary to qualify this procedure, because every system is different.

**Filter Service Life**

Filter cartridge change-out is usually based on actual experience, with a safety factor. Filters should be inspected on a monthly basis for oxidation. This should be supplemented by monitoring the pressure drop across the filters during operation to determine if the filters are plugging and routine integrity testing to confirm filter integrity during the service life of the filter. Alternatively, filter life studies, with an appropriate safety factor, could also be used to
set a change-out schedule. Actual conditions for each application should be used during filter life studies.

For elevated temperature applications, conditions leading to oxidation need to be considered. Aging of the membrane filters by oxidation depends on the status of the system. Oxidation does not occur when the cartridge is being steamed, because there should be no air present in a properly operating steam-in-place system. If a cartridge is exposed to air at an elevated temperature, oxidation of the material in the filter, such as polypropylene hardware, will be accelerated. Oxidation will also occur when the filter is in a stagnant situation, that is, it has no air flow going through it. The flow of air through a filter can moderate the temperature environment, whereas under stagnant conditions the temperature of the filter will rise to the temperature of the housing. Stagnant conditions can exist when the tank is not being used or when the tank is empty. To prolong service life, the steam jacket should be turned off when there is no air flow through the filter for extended periods of time, when operating conditions permit.

Steam Sterilization Guidelines for Sterilizing Grade Membrane Filters

Membrane filters can be sterilized by chemical sterilants (such as ethylene oxide, hydrogen peroxide in vapor form, propylene oxide, formaldehyde, and glutaraldehyde), radiant energy sterilization (such as γ-irradiation) or steam sterilization. The most common method of sterilization is steam sterilization.

Steam sterilization of a membrane filter can be accomplished either by an autoclave or by in situ steam sterilization. In situ steam sterilization can be effectively accomplished by a variety of different process arrangements. Steam sterilization is often the most critical portion of the process, and it is important that the procedures followed lead to sterilization of the system and do not impart any damage to the membrane filters. Some general procedural recommendations are included for prevention of either forward or reverse pressurization damage.

Reverse pressurization conditions can be prevented by using a noncondensing gas, such as air or nitrogen, at the end of a steam cycle. If this step is not followed, and the steam valve is shut off without the introduction of air or nitrogen, the filter housing will act like an isolated system. As such, the temperature will be different on the upstream (outside of the housing) and downstream sides of the filter. Due to the temperature difference, steam will condense at different rates on the upstream and downstream sides of the membrane. This can lead to pressure differences of 5 to 15 psi in the reverse direction. Even if this reverse pressure condition exists for a short period of time, the filter can suffer permanent damage.

Condensation during the steam cycle can lead to long cycle times and filter damage by excessive forward pressurization. If during steam sterilization the filter membrane is wet with condensate and in order to overcome the resistance of the wetted membrane the differential pressure in the forward direction is greater than required maximum (typically 5 psi), then the filter can be damaged by excessive forward pressurization. In this case, the filter will appear crushed. Precautions should be taken to minimize condensate accumulation. Drains for condensate removal should be strategically placed and should be cracked during the steam cycle.

CONCLUSION

This chapter has provided information on a variety of air filtration applications: fermenter inlet air, fermenter vent gas, vents on water-for-injection tanks, and vacuum break filters during lyophilization. In addition to a description of the air filtration applications, guidelines were provided on the usage of the filters in these applications.

BIBLIOGRAPHY


INTRODUCTION

Bioprocess technology operations have included membrane cartridge filtration for more than 20 years. Cartridge filters have been used for the removal of submicron particles, such as bacteria, and for the removal of particulate material, usually greater than 1 μm. Many applications involve the removal of bacteria in sterile filtration applications; particulate removal cartridge filters can be used as prefilters to protect the sterilizing grade filters. Some examples of applications for liquid filtration include fermentation media, deionized water, sera, and sanitizing chemicals. The purpose of this article is to provide information on operational considerations for cartridge filters and to explain the usage of cartridge filters in bioprocessing.

DESCRIPTION OF MEMBRANE CARTRIDGE FILTRATION

Cartridge filters used in bioprocessing applications are used for the removal of bacteria (sterilizing grade filtration) or for the removal of particulate material (prefiltration). The mechanisms at work during filtration are provided, as well as, a general description of sterilizing grade filters and prefilters. Prefilters are often used upstream of the final (sterilizing) filters to protect the final filter from premature plugging, thereby prolonging the life of the final filter.

Filtration Mechanisms

There are four mechanisms that can govern the particle removal efficiency of membrane filters: (1) direct interception, (2) inertial impaction, (3) Brownian motion or diffusional interception, and (4) electrostatic attraction between the membrane and particles (4).

Direct interception involves a sieving action that mechanically retains the particles on the filter surface. The filter acts as a screen that stops particles that are larger than the pores (openings) in the membrane. Direct interception is independent of face velocity and mostly involves particles with relatively large diameters.

Inertial impaction refers to the deviation of a particle from its streamline during flow because of inertia, resulting in the retention of the particle by the membrane. As the face velocity increases, the probability of inertial impaction increases.

Brownian motion or diffusional interception applies to small particles at low face velocities in a gas stream. When air molecules are in a state of random motion, small particles are struck by moving air molecules and displaced. The movement of particles resulting from molecular collisions is known as Brownian motion. This phenomena can increase the probability of capture of particles by diffusional interception within the membrane.

An electrostatic attraction can exist or be induced between the membrane material and particles in the fluid stream. This attraction can enhance the removal of particles and their retention by the filter membrane.

The pores (open area) in most membrane filter materials are typically not straight through the membrane. In membrane filters, the membrane can be considered to be constructed of multiple screens that provide a tortuous path for the fluid through the membrane.

In bioprocessing applications, there are many processes that require bacterial removal (sterile filtration). The fil-
ters used to provide bacterial removal are generally 0.2-\(\mu\) sterilizing grade filters. (The tests used to ensure that the filters do remove the bacteria are described below.)

Removal of bacteria by a 0.2-\(\mu\) sterilizing grade filter is not only by a sieving action on the surface, but also by size-exclusion entrapment in the membrane structure. As shown in Figure 1, a membrane filter structure is spongelike. The ability of such a membrane to remove bacterial contaminants is related to its thickness as well as to the size of its pores. For example, it has been demonstrated that whereas a single 150-\(\mu\) thick layer of 0.8-\(\mu\) cellulose ester membrane can provide only a titer reduction of 70 for Brevundimonas (Pseudomonas) diminuta, 10 layers can provide a titer reduction of 2.4 \(\times 10^9\) (5).

Because membrane filters are constructed of porous materials, they will have a certain amount of open area. The open area allows for fluid flow through the membrane; this area, expressed as a percent, is known as the porosity of the membrane.

**Sterilizing Grade Membrane Filters**

**Construction.** Membrane filters used for sterile liquid filtration are typically constructed of polymeric microporous hydrophilic materials, such as nylon, cellulose acetate, modified polyvinylidenefluoride (PVDF), or other polymers. Most liquids that require sterile filtration are water based, and hydrophilic membranes will spontaneously become wet with water. Absolute rated hydrophilic membrane filter cartridges have been used for sterile filtration of parenterals, diagnostic reagents, purified water and water-for-injection, dry gases, organic solvents, buffers, and biological fluids such as serum, plasma, tissue culture media, nondilute protein solutions, and fermentation harvest fluids (1–3).

Membrane filters used for the sterile filtration of gas streams, such as nitrogen blankets, storage tank sterile vents, formulation tank sterile vents, sterile air for aseptic packaging, sterile filtration of fermenter air, or vent gases, typically contain a membrane made of hydrophobic materials such as PVDF or PTFE (polytetrafluoroethylene). Hydrophobic membrane filters are desired in these sterile gas filtration applications because hydrophobic filters do not spontaneously wet with water. When a hydrophilic filter is wetted with water, it will not pass air until the water-wet bubble point of the filter is exceeded. This water-wet bubble point can be greater than 50 psi (see FILTRATION, AIR).

Membrane materials can be modified to optimize their performance for certain types of applications. Positive-charged hydrophilic membranes contain cationic (positive charged) functional groups that impart a positive \(\zeta\)-potential when immersed in an aqueous solution and provide enhanced retention of particles smaller than the absolute rating. Most particles have a negative charge, and the positive charge provides a mechanism, in addition to the tortuous path of the membrane, by which the particles are retained. Applications involving charged membrane filters are discussed in the Refs. 6, 7, and 8.

Membrane surfaces can also be modified to expose hydroxyl groups, which tend to prevent the retention of proteins. This is a benefit for dilute protein solutions (typically less than 1 mg/mL), in which the protein is the desired product. Such fluids include serum-free tissue culture media and protein additives, protein-based therapeutics and diagnostics, dilute protein containing diluents and buffers, recombinant proteins, hormones and growth factors, protein chromatography feeds and eluates, vaccines, and other dilute biologicals. Filtration applications involving low protein-binding membrane filters are described in Refs. 9 and 10.

The membranes used for sterile filtration bioprocessing applications are usually pleated and formed into a cylinder. Figure 2 illustrates the pleated cartridge construction.

The filter membrane is typically cast onto a support material (e.g., nonwoven polyester or polypropylene substrate) that provides high tensile strength while retaining flexibility. Layers of the membrane are corrugated, or pleated, along with upstream and downstream layers of courser material (e.g., nonwoven polyester or polypropylene). These layers provide support and drainage for the membrane. The sides of the corrugated membrane pack are sealed, often by a heat seal.

The corrugated filter membrane pack is fitted around the core. The purpose of the rigid, inner core is to provide support for the filter element against pressure in the forward direction. An external cage is provided for additional support and protection during handling. The flow path for the cartridge filter is from the outside (cage) to the inside (core). The material used for the core, cage, and support layers will depend on the intended application for the filter. Polypropylene, polyester, and PTFE are examples of ma-

![Figure 1. Membrane filter structure.Courtesy of Pall Corporation (20).](image-url)
materials that can be used for the cage, core, and support material in a membrane filter. The membrane pack ends are sealed, usually by heat, with end caps. The end caps are constructed of the same types of materials as the core and cage. The final step in the construction of a membrane filter is to attach the proper end closure and O-ring adaptor by an appropriate welding technique.

The end closure is used for placement of the filter in a filter housing. A finned or flat blind end cap on one end is generally attached on one end and an open outlet at the other end. Sealing into the housing is accomplished with a double O-ring seal at the open end to ensure no fluid bypass. Filter elements can have locking tabs at the double O-ring adaptor base for bayonet locking of the element to ensure positive sealing in the housing base.

Most single filter elements are 10 inches in length. For the construction of a multiple length cartridge (20, 30, or 40 inches), the sections are attached by welding 10-in. subassemblies end to end. Figure 3 shows the construction types.

Requirements. For sterile filtration processes, the biological safety of the membrane filter or filter cartridge should be demonstrated by the performance of the USP (88) Class VI (121 °C) Plastics Test for Biological Reactivity (11). Another typical qualification used to select filter materials of construction is a listing for food contact in CFR Title 21 Part 177 (12).

For sterile aseptic processes, a typical requirement is sterilization by passage through a 0.2-μm sterilizing grade membrane filter, as defined by ASTM Standard F 838-83 (13).

Figure 2. Pleated cartridge configuration. Courtesy of Pall Corporation (29).

Figure 3. Cartridge filter construction. Courtesy of Pall Corporation (30).
A membrane filter as well as all system components can contribute extractables to the process stream. Many filter manufacturers document the extractables level of a particular filter in an appropriate solvent as a nonvolatile residue (NVR). These extractables are typically composed of the oligomers or additives of the plastic materials present in the filter element. It is important to note that the amount of extractables from a filter element as well as any extractable from the rest of the process system can be reduced by flushing the system before filtration.

Many filter manufacturers address the issue of effluent quality requirements by the performance of appropriate tests on filter samples from manufacturing lots. These include tests for:

- Cleanliness: Per current USP limits under Particulate Matter in Injections (788) and conformance with requirements for a non-fiber-releasing filter per Title 21 of the U.S. CFR Title 21, Part 211.72 and Part 210.3(b)(6) (14)
- Oxidizable substances: Per current USP requirements under Purified Water after flushing (15)
- pH: Per current USP requirements under Purified Water after flushing (15)
- Pyrogens: Per current USP requirements under Bacterial Endotoxins Test as determined using the limulus ameobocyte lysate (LAL) reagent with an aliquot from a soak solution (16)

Prefiltration

Cartridge filters used for prefiltration can be either pleated or nonpleated (depth) filters. Standard design depth filter cartridges are open cylinders of thick filter material (double open-ended). Prefiltration filter cartridges are typically available in 2-1/2 or 2-3/4-in. o.d., in multiple lengths of 10 in. up to 40 in. long.

Prefiltration encompasses a wide variety of fluid types and removal requirements. The following is a discussion of some important characteristic of prefilter construction and a sampling of some of the types available.

The particles or fibers in the prefilter membrane should not become dislodged, slough off, shed, or in some other way contaminate the filtrate (media migration). The filter medium should also not unload or release contaminants retained by the filter medium into the effluent as the differential pressure across the membrane increases. Data that indicate filtration efficiency as a function of increasing differential pressure should be available from the vendor to support the claim of nonunloading.

Polypropylene Depth Type Filter Cartridges. Polypropylene depth filters are made using nonmigrating continuous filaments of polypropylene filament medium without resin binders. Absolute rated polypropylene depth filter cartridges are applicable as prefiltration for serum, vaccines, diagnostics, tissue culture products, deionized water, container washing, and final product. These filters are used in fermentation for liquid feeds, makeup water, solvents, and antifoam. In downstream processing, they are used for cell and cell debris removal, buffers, cleaning agents, and sanitizing solutions. Positive charged versions of these filters are possible and have the added ability to remove organisms and particles such as bacterial endotoxins (pyrogens) that are much smaller than the absolute pore size rating.

The membrane for these depth filtration cartridges has an inner (downstream) section in which the pore diameter is constant (this section provides absolute rated filtration) and an outer section in which the pore diameter varies continuously from that of the absolute rated section to as much as 90 μm or more. Pore size variation of the filter cartridge is achieved by varying the fiber diameter while maintaining constant void volume throughout the medium. The constant void volume provides increased dirt capacity and a low clean pressure drop of the filter cartridge.

Polypropylene Pleated Filter Cartridges. Polypropylene pleated filters are applicable as prefiltration for makeup and rinse water, for serum for culture media, and in downstream processing for cell debris removal and prefiltration of solvents and buffers. These process filter cartridges can be constructed of the same type of membrane material as depth filters.

The thin sheet of polypropylene media is pleated and formed into a cylinder with a longitudinal side seal of melt seal polypropylene. The cylinder is then sealed to injection molded polypropylene end caps to ensure no fluid bypass. Polypropylene hardware components consisting of an inner support core and an external protective outer cage are incorporated. The process filter cartridge with polypropylene hardware should be rated to withstand differential pressure of 80 psi up to 50 °C (122 °F) and 60 psi up to 80 °C (176 °F).

Resin-Free Cellulose Pleated Filter Cartridges. Resin-free pleated cellulose filter cartridges are applicable as prefilters for makeup and rinse water, reverse osmosis membranes, deionized water, and inlet air for fermenters and bioreactors. These filter cartridges are constructed of pure cellulose medium without resin binders, which is pleated into a high area cylinder. The longitudinal side seal of the process filter cartridge should be a polypropylene (approved for food contact usage).

Cellulose media cartridges are assembled with hardware components consisting of a perforated inner support core, an outer support cage, and end caps melted to the medium in the plastic. All hardware components should be of pure polypropylene, without filter or reinforcement, to ensure a high quality filtrate with a minimum of soluble extractables. The process filter cartridge with polypropylene hardware should be rated to withstand differential pressure of 80 psi up to 50 °C (122 °F) and 60 psi up to 80 °C (176 °F).

Resin-Impregnated Fiberglass Pleated Filter Cartridges. Resin-impregnated fiberglass filters are applicable as prefilters for deionized water, solvents, serums, reagents, and cell debris removal from harvest fluids. The resin impregnation makes available a choice of filters with either a positive or negative -potential when used in aque-
ous service. Filter cartridges with negative $\zeta$-potential are constructed of glass fiber that has a natural negative $\zeta$-potential and reinforced with a resin binder that is also naturally negative when immersed in water. Positive $\zeta$-potential filter cartridges are constructed using a resin binder that coats the glass fibers and imparts a positive $\zeta$-potential to the medium.

The filter cartridges are available with hardware components consisting of end caps and an internal perforated support core made of either polypropylene or stainless steel and in an external polypropylene outer cage or protective net. Stainless-steel end caps are attached to the filter by an inert synthetic resin. Polypropylene end caps are melt sealed to embed the medium in the plastic. Filter cartridges with polypropylene hardware should be rated to withstand differential pressures of 80 psi up to 50 °C (122 °F) and 60 psi up to 80 °C (176 °F). With stainless-steel hardware, the filter can be rated to 75 psi up to 135 °C (275 °F).

**Filter Assemblies Used in Processes**

During filtration, the membrane filter cartridges are placed in housings that provide a means of fluid contact with the filter cartridge. The housings are typically composed of 304, 316, and 316L stainless steel; plastic housings are also available. The housing design is typically composed of a head and a bowl. The head is the portion of the housing in which the filter is attached. The bowl is clamped to the head. A housing is illustrated in Figure 4.

Membrane filters can also be obtained in complete assemblies. These disposable assemblies consist of a membrane filter cartridge welded into a plastic filter housing. Disposable assemblies or capsule filters are typically used in small-scale applications (e.g., 100-L batches or less, depending on the fluid) and are particularly useful in applications in which operator contact with the fluid is not desirable, because the entire assembly can be removed from the system and discarded without operator contact with the product-wet cartridge. Figure 5 shows disposable assemblies.

**MEMBRANE FILTER RATINGS AND TESTS**

Many filter manufacturers use a nominal micron ($\mu$m) rating for particle removal efficiency. This is defined by the American National Standards Institute (ANSI) as an “arbitrary micrometer value indicated by the filter manufacturer. Due to lack of reproducibility this rating is deprecated.” Further, nominal rating standards are arbitrary and a comparison of nominally rated filters is imprecise. In addition, nominal ratings can be misleading because the filter can allow passage of particles larger than the rating indicates.

In order to establish a more meaningful filter rating, a test is performed on the filter using an appropriate contaminant to meet a specific claim for retention of the contaminant. The test procedure must be sensitive enough to detect the passage of contaminants of interest. The testing should be performed under carefully controlled conditions using industry-accepted reference standards. These reference standards include silica suspensions, latex beads, or microorganisms.

**Particulate Removal Filters**

The removal ratings for particulate removal filters (typically used as prefilters in sterile filtration processes) can
be established using the Oklahoma State University (OSU) F-2 Test. This rating method (ISO 4572, ANSI B93.31) has received wide acceptance for use on lubricating and hydraulic fluids. The technique has been used for oils extensively and has been adapted for use in water with contaminants ranging from 0.5 to 25 μm.

The test is based on continuous on-line particle counts of different particle sizes, both in the influent and the effluent stream. The ratio at a specific particle size is defined as \( \beta_x \): the number of particles of a given size \( X \) and larger in the influent, divided by the number of particles of the same size \( X \) and larger in the effluent, where \( X \) is the particle size in micrometers. The percent removal efficiency can be calculated from the \( \beta_x \)-value. The percent removal efficiency is \( \left(1 - \frac{\beta_x}{100}\right) \times 100 \) (17).

Sterilizing-Grade Membrane Filter Ratings and Integrity Tests

Sterilizing-grade hydrophilic membrane filters require testing using a liquid bacterial suspension challenge that is sensitive enough to detect the passage of any microorganisms, for the establishment of the micron rating of the filter. A sterilizing-grade filter is defined by the FDA's Guideline on Sterile Drug Products by Aseptic Processing (18) as one that will produce sterile effluent when challenged with the test organism Brevundimonas diminuta (ATCC 19146) to the level of greater than 10⁷ CFU/cm² filter area. Such microorganism retention tests are conducted with P. diminuta (ATCC 19146), measuring 0.3 × 0.6 to 0.8 μm, to verify that membrane filters produce sterile fluid filtrates. This is a standard test for the validation of sterilizing-grade filters (0.2-μm pore size rating) in the bioprocessing industry.

In practical terms, there is a limit to how many bacteria a filter element can be challenged with before it becomes plugged. When a filter has been challenged at the level of 5 × 10¹⁰ CFU/cm² it becomes effectively clogged. This challenge level corresponds to about 2 × 10¹³ CFU for a 5-ft² membrane filter (17).

The liquid test involves challenging a test filter with a known quantity of P. diminuta, no less than 1 × 10⁷ organisms per cm² of filter area. The challenge sample is suspended, for example, in sterile water and then filtered at a defined flow rate and time through a test filter, followed by an analysis membrane. After the challenge is complete, the system is flushed for 5 to 10 min with sterile water. The analysis membranes are removed and placed on Mueller Hinton agar and incubated at 32 °C for 48 h. After incubation, the plate is examined for the presence or absence of microbial colonies. Figure 6 illustrates a liquid bacterial challenge test stand.

For some applications (e.g., mycoplasma control or removal) a membrane with a more stringent rating (0.1 μm) is required. These applications typically involve sera and tissue culture media. One organism that has been used extensively in validating 0.1-μm rated filters is Acholeplasma laidlawii (19).

A liquid bacterial challenge is a destructive test; the filter cannot be used after the challenge. Therefore, a correlation is made to a nondestructive test. It is the correlation, not the nondestructive test alone, that provides assurance that the filter performs as required. During the production of a sterile product, the filter should be tested before and after the filtration to ensure that the filter meets the specification, is properly installed and intact, and confirms the filter rating.

The three major tests used to determine the integrity of a membrane filter are the bubble point, forward flow, and pressure-hold integrity tests. These tests are based on the flow of a gas (air or nitrogen gas) through a liquid wetted membrane under applied gas pressure. A fourth test, the water intrusion test, is sometimes used to determine the integrity of hydrophobic membrane filters.

For a nondestructive test to be useful, the results of the test must be able to predict the ability of the filter to remove bacteria. Such a test should also be easy to perform and highly reproducible. The manufacturing quality control of the membrane filters tested to obtain this prediction must be used to establish quality control specifications that are consistently maintained in filter manufacturing. It is the combination of the nondestructive integrity test, the correlated bacteria challenge test, and manufacturing quality control that allows membrane filters to be reliably used for sterile filtration of bioprocessing products.

**Forward Flow, Pressure Hold, and Bubble Point Tests.** The industry-accepted, nondestructive tests used to verify sterilizing grade filter integrity are the forward flow, pressure hold, and the bubble point tests (20). These tests are performed by applying a preset air (or nitrogen gas) pressure to a filter wet with an appropriate fluid. For hydrophilic filters (filters that wet readily with water) the wetting fluid is typically water or a water-based fluid; for hydrophobic membrane filters (filters that do not spontaneously wet with water and are used in air service) 60/40 isopropyl alcohol/water is a typical wetting fluid.

The wetting fluid fills the voids in the membrane. For this to be occur uniformly, the wetting must be complete. The result of the wetting can be considered as a layer of water supported by the membrane. When gas pressure is applied to one side of the membrane, the test gas will dissolve into the liquid layer to an extent determined by the solubility and pressure of the gas (as described by Henry's law [21]). Downstream of the membrane, the pressure is lower, and the gas in the liquid is driven out of solution. The result is a net flow of gas through the membrane. The forward flow, pressure hold, and the bubble point tests all involve the observation or measurement of a gas flow through a wetted membrane. The measured flow is a com-
bination of the diffusion of the gas through the liquid layer and bulk gas flow through any open passageways.

The rate of diffusive flow is a function of the test pressure, the diffusivity of the gas in the liquid, the solubility of the gas in the liquid, the membrane void volume, and the membrane thickness. This diffusive flow is described by Fick’s first law (13), which is presented below in an integrated form:

\[ N = \frac{D H \Delta P \rho}{L} \]  

(1)

where \( N \) is the permeation rate, \( D \) is the diffusivity of the gas in the liquid, \( H \) is the solubility coefficient of the gas, \( L \) is the thickness of the liquid in the membrane, \( \Delta P \) is the differential pressure, and \( \rho \) is the void volume.

As the applied upstream gas pressure is increased, the diffusive flow increases proportionally. At some point, the pressure becomes great enough to expel the fluid from a passageway and establish a path for bulk flow of gas. As a result, the gas flow through the wetted filter begins to increase in a nonlinear manner due to the initiation of bulk gas flow through these open paths. The pressure at which this occurs can be modeled by a modified capillary rise equation:

\[ P = F \frac{4 \gamma \cos \theta}{D} \]  

(2)

where \( P \) is the pressure (dynes/cm²), \( \gamma \) is the surface tension (dynes/cm), \( \theta \) is the contact or wetting angle, \( D \) is the diameter (cm), and \( F \) is the shape factor (to account for the structure of the membrane). As the pressure is further increased, more passages are opened and the proportion of the total gas flow due to the bulk air flow increases. If the pore size distribution is narrow, then the possible passages through the membrane will provide similar resistance to flow.

For a wetted high-area pleated membrane filter, a plot of differential gas flow pressure versus gas flow will show a gradual transition from diffusive to bulk flow. An alternate plot of differential pressure versus gas flow or differential pressure will reveal a rapid transition from diffusive flow to bulk flow, as seen in Figure 7. This graph is referred to as a diffusive flow spectrum or a \( K_L \) (knee location) curve. The \( K_L \) pressure represents the transition from the diffusive flow region to the bulk gas flow region.

The bubble point test is a test designed to determine the pressure at which a continuous stream of bubbles is initially seen downstream of a wetted filter under gas pressure. There are several ways of performing this test. The first of these is the visual bubble point test. Generally, the

**Figure 6.** Liquid bacterial challenge. Adapted from “Engineering Considerations in Sterile Filtration Processes” by Holly Haughney in Sterile Pharmaceutical Products: Process Engineering Applications, edited by Kenneth E. Avis. Copyright © 1995 by Interpharm Press, Inc. All rights reserved. Adapted by permission.

**Figure 7.** Diffusive flow spectrum. Courtesy of Pall Corporation (20).
Visual bubble point test is not used except on small area disks where the whole membrane surface may be examined. To perform a visual bubble point test, gas pressure is applied to one side of a wetted membrane, with a layer of the wetting fluid maintained on the other side. The gas pressure is slowly increased. The rate of increase will affect the measured bubble point value. At some point, the pressure becomes high enough to displace the liquid from one or more passageways. This establishes the beginning of bulk flow of air through the membrane, and a thin stream of bubbles can be observed on the downstream side of the membrane. The pressure at which this stream of bubbles is observed is referred to as the first bubble point.

A manual bubble point is measured with a filter installed in a housing, with tubing running from the housing outlet to a beaker of water as shown in Figure 8. The geometry and position of the filter housing, and the diameter and length of the tubing will affect the measured bubble point value. In this case, the bubble point is considered to be the pressure at which a steady stream of bubbles is seen coming from the tubing; however, the determination of a steady stream is often subjective.

The manual bubble point (Fig. 9) will generally occur at a higher pressure than the first bubble point determined by a visual examination of the filter surface, because greater flow is necessary to obtain a clear stream of bubbles from the volume of the tubing. When determining the manual bubble point on a high area cartridge, the diffusion of air through the wetted filter at pressures below the bubble point can be significant. The complications resulting from the length and diameter of the tubing, the housing, and operator interpretation render bubble points determined in this manner not easily reproducible or comparable to bubble points determined by other techniques.

The forward flow test and the related pressure hold test were first introduced by Pall Corporation in 1973 (22). During these tests (Fig. 10), a wetted membrane is subjected to a predetermined gas pressure on the upstream side of the membrane. The gas will diffuse through the wetted membrane at a measurable rate. During the forward flow test, the diffusion of the gas and flow through any open pores is measured at the specified test pressure.

The pressure hold test (Fig. 11) is a modification of the forward flow test. The upstream side of filter housing is pressurized at the test pressure and then the filter is isolated from the pressure source. The diffusion of gas through the wetted membrane and flow through any open pores is measured as a decay in pressure over a specified period of time.

The pressure chosen as the test pressure for the forward flow or pressure hold test must be close enough to the $K_f$ to allow differentiation between different grades of membrane and high enough to present a measurable flow. It is also necessary that the test pressure chosen be such so that the background diffusive flow does not interfere with the sensitive detection of defects in the filter. The exact pressure chosen for the forward flow test must be below the $K_f$. The test pressure is set by filter manufacturers and should be used during tests performed to establish the correlation with bacterial retention.

Integrity tests can be performed with automated equipment. Automated filter integrity test instruments (e.g., TruFlow®, Integritest®, Palltronic®, and Sartocheck®) have been developed to provide accurate and reproducible filter integrity test values. The instruments traditionally were designed to perform a pressure hold test. The forward flow value can be calculated from the pressure hold value. Newer equipment designs use direct upstream air flow measurements.

The tests are controlled and monitored by a built-in microprocessor, and the equipment often can be used with a programmable logic controller (PLC) for full system automation. An example of an automated integrity test application is provided in Ref. 23.

Automated devices cannot perform a true bubble point test. Instead, a modified bubble point test is performed, which essentially consists of a series of pressure hold tests. In the automated bubble point test, upstream air pressure is increased in steps. Between each pressure increase, the instrument isolates the upstream volume and monitors the pressure decay upstream of the filter. As the pressure decay exceeds a certain value, the software uses a specific algorithm for calculating the bubble point from relative values of this series pressure decays. The instrument uses a series of pressure hold tests as the basis for the calculation of the bubble point rather than measuring the bubble point directly. Figure 12 illustrates how pressure hold
tests can be used to determine a bubble point. The actual algorithms used in the automated test devices may differ between automated filter integrity test instrument manufacturers.

**Water Intrusion Test for Hydrophobic Filters.** The water intrusion test can be used to test hydrophobic membrane filters (24–27). To perform a forward flow, pressure hold, or bubble point test on a hydrophobic filter, the membrane must be wet with an alcohol solution (e.g., 60/40 IPA/water). For an in situ integrity test, the wetting agent has to be added to the system to perform the test and then removed to restore air flow through the filter. The potential flammability of the alcohol is a consideration, and for some applications it can be necessary to validate the complete removal of the solvent to avoid contamination of the product. For these reasons, integrity testing of hydrophobic filters in situ has not been widely used, and off-line testing...
has been more common. The increasing industry demand and growing regulatory requirement for in situ testing of sterilizing-grade air filters has led to the development of water-based tests.

If the upstream side of a dry hydrophobic filter assembly is filled with water and pressurized, the hydrophobic nature of the porous membrane will prevent the bulk flow of water through the filter until the intrusion pressure is reached. At pressure below the intrusion pressure, a small but measurable flow of water through the membrane occurs.

The presence of larger pores in the filter will be detected by an increased flow due to bulk water flow through these pores (Fig. 13). This principle is the basis of the water intrusion test.

At the start of the test, the upstream side of the filter assembly is filled with water and a predetermined gas pressure is supplied. This causes a fall in the water level within the filter housing because of factors such as pleat compression and the passage of trapped air through the membrane. It is important that the changes in the upstream volume caused by these effects have stabilized so that the very small flows of water through the membrane can be accurately measured. A typical test arrangement is illustrated in Figure 14.

Adequate stabilization time (e.g., 10 mins) is also required to ensure thermal equilibrium. Water flow can be measured by determining the upstream flow of gas required to keep the pressure constant. The use of a direct flow measurement integrity test instrument can be used for water intrusion test measurements. After the test is over, the water can be drained and in many cases the filter is ready for use. Where dry air or gas is required, a drying procedure can be used.

**FILTRATION SYSTEM DESIGN AND OPERATIONAL CONSIDERATIONS**

**Filter Selection**

As detailed earlier, there are a number of different types of cartridge filter types for a wide variety of applications (1–3). In order to properly design a filtration system, all the requirements for the system and fluid must be considered. As a starting point, the removal efficiency must be defined. Other considerations can include process temperature, flow rate, pressure drop restrictions, integrity test regimen, and sterilization.

All cartridge filter materials (membrane, cage, core, support material and housing) should be compatible with the process fluid over the temperature and pressure range indicated for a given process application. In general, most bioprocessing fluids are water based and are compatible with most membrane materials, such as polyamides (nylon), PVDF, polysulfone, and cellulose acetate.

For many filtration processes, it is desirable to have a staged filtration system. A staged system will have one or more prefilters of higher micron rating protecting the final filter. The final filter in the stage is the filter that provides the desired effluent quality. The prefilters are typically courser filters that remove larger contaminants in the fluid and reduce the loading of the final filter. For a sterile filtration process, the final filter is typically a 0.2-μm membrane filter. In a staged system, the life of the final filter is enhanced by the use of the prefilters. This will usually reduce the cost of the system, because the cost of the final filter is generally greater than the cost of the prefilters.

The size of the filtration system will depend upon the volume of the fluid being filtered and the ease with which the fluid is filtered or its filterability. For production applications, the membrane filters are typically provided as 10-in. segments or filter elements. These segments can be welded together to form 20, 30, and 40-in. filter elements. The diameter of the cartridge is approximately 3 in. Filter housings can hold a single cartridge (10- to 40-in. elements) or multiple filter cartridges. For applications that
involve relatively small volumes (for example 10 L), disposable filter assemblies (capsule filters) with smaller membrane areas are available. If prefiltration is required, then the prefilters will also be a part of the filter train. For many applications, there is a one-to-one correspondence of prefilters to final filters. Optimization for a particular process may indicate that fewer, or more, prefilters are needed.

In some production environments, it is desirable to have a dual filtration system (filters arranged in parallel) to allow for a changeover from one system to the other. While one system is being used, the other system can be prepared for use (e.g., it can be cleaned and tested and the filters can be changed out). The decision to use a dual filtration system requires a balancing of the ease of operation with the additional space required for the additional system. For some sterile applications, redundant final 0.2-μm filters (filters in series) are used to provide additional assurance of sterility. If one of the filters fails, because of processing conditions, then the other filter is in place as a backup. If the backup filter is integral, then the fluid has been sterile filtered. The decision point is whether further assurance of maintaining sterility throughout the entire filtration processing time can justify a redundant system.

Flow Rate and Pressure Considerations

All components of a system will cause a resistance to flow, which results in a pressure drop (ΔP). Pressure drops, or pressure losses in a system, can be caused by piping, connections, valves, and filling heads and by the filter and its assembly. The pressure drop through a clean filter assembly is due to the sum of component pressure drops, including the filter housing, the filter hardware, and the filter membrane. The filter membrane will cause a resistance to flow, or a pressure drop, in a filtration system. As the filtration progresses and the filter membrane becomes plugged with the particulate contaminants in the fluid, an increasing pressure drop will occur. The pressure drop will be different, depending on the fluid hydrodynamics (e.g., viscosity) and the filter membrane material characteristics (such as porosity).

An adequate pressure source is needed to overcome the pressure drop across the system, including the filter assembly and to allow for the maintenance of a constant flow rate. If the initial clean pressure drop across a filtration system is close to the maximum available pressure, the flow rate will not be maintained because as the filter plugs the pressure drop will increase and there will not be enough pressure to overcome the increased pressure drop throughout the filtration. This will cause a decrease in flow. If possible, where constant flow is required, pumping capacity should be increased. An alternative remedy would be to increase the membrane filtration area so that the initial pressure drop is reduced.

Ultimately, the filter will become plugged with contaminants. This will depend on the amount of the contaminant in the fluid that is being filtered. For most pharmaceutical sterile filtration applications the load (percent solids) is typically low. Filters upstream of the sterile filter are more likely to become plugged in a properly designed filter train. As plugging occurs, the differential pressure will increase. System design should include a change-out schedule so that the differential pressure constraints for the filtration system are not exceeded. For example, in a batch system, the filter change-out should occur after the batch filtration has been completed. If multiple in-line stages of filtration are needed, then the differential pressure for each stage must be added, and the sum must be less than the maximum differential pressure allowed for the filtration process to allow for increased pressure caused by membrane fouling.

At no time during a process should the pressure drop exceed the recommendations for pressure drop (usually specified for a temperature range of the individual filter unit). At elevated temperatures the recommended pressure drop is typically lower than the recommended pressure drop for ambient conditions. A combination of elevated temperature and a high pressure drop can lead to structural filter damage. Structural damage to the filter is possible under these conditions because at elevated temperatures the polymeric components of the filter can become softer and are more likely to be deformed if the pressure drop is too high. For example, under extreme conditions, the core of the filter as well as the cage can be crushed.

Required flow rate, maximum pressure drop, and available pressure for the filtration system must be considered when the system is designed. Flow rate is typically reported as either gallons per minute (GPM) or liters per minute (LPM). The flow rate achievable through a filtration system is directly related to the applied pressure and inversely related to the resistance to flow. Thus, if the applied pressure is increased, the flow rate will increase. If the resistance to flow is increased, such as by membrane plugging, then the flow rate will decrease. It is important to note that an initial higher pressure (and flow rate) can lead to premature membrane plugging with many products.

The nature of microporous membrane filters is that they will tend to plug rapidly if they are subjected to a low flow rate (directly related to a high applied pressure) during the start-up of a filtration. For fluids with a significant particle loading, under conditions of initial high flow rate, the microporous membrane can become rapidly plugged, or fouled, and the pressure drop will increase and the throughput, or filter life, will be reduced. This is especially true for products that may contain gels, such as biological products. Filter life, or throughput, can be increased if the initial flow rate is reduced. If the initial flow rate is decreased, the pressure drop will increase at a slower rate and the throughput will be increased. Another approach is to increase filter surface area for the given flow rate. Additionally, the pressure drop across a filter assembly can be reduced by increasing the filtration area. Increasing filtration area can often be an economical approach because the increase in throughput is often greater than linear for an increase in filtration area. Figure 15 illustrates the concept of a lower flow rate and differential pressure providing a higher throughput (total volume filtered).
Fluids with higher viscosities will be more resistant to flow and the pressure drop across the filter will be greater. For fluids with high viscosities, a greater applied pressure will be required to maintain the process flow rate. The smaller the pore size rating of a filter, for a fluid of a given viscosity, the greater the resistance to flow.

**Scale-up**

Although a particle size distribution evaluation can often provide useful information about the contaminant material in a process stream, the filtration system usually cannot be selected without an actual scaled-down filterability test. Actual testing is required because the particle size distribution fails to provide information on the quantitative particle load, its effect on membrane life, and how the membrane filters will perform with the process fluid under process conditions. In addition, other factors such as gels and high molecular weight biopolymers present in some products will also affect filter life and can only be evaluated by actual test.

The scaled-down, or filterability, test will require a relatively small representative sample of the process fluid (typically 2 to 4 L). Membrane filter disks with a 47-mm diameter are typically used. If a process stream is subject to variable composition, then the worst case fluid should be defined and tested. Only membrane materials that are compatible with the fluid are tested during the filterability test, and test conditions (e.g., temperature and flow rate) should match the process conditions as nearly as feasible. Several filtration schemes, including staged filtration schemes, are usually evaluated to optimize the filtration. The filterability tests can be performed at either a constant pressure or a constant flow rate. The use of constant pressure or constant flow rate must be considered because the system size, system life, process time, and throughput can depend on these process parameters. The flow rate can be scaled linearly based on filtration area. Optimization parameters must be defined and can include effluent quality, time (or throughput) to reach terminal (maximum allowable) differential pressure at a given flow rate, or flow rate for a given applied pressure.

System scaling can be based on the following ratios for the final filter in the system, where it is necessary to solve for full scale system area:

\[
\frac{\text{Full scale system area}}{\text{Scaled down system area}} = \frac{\text{Full scale system flow rate}}{\text{Scaled down system flow rate}}
\]

\[
\frac{\text{Full scale system area}}{\text{Scaled down system area}} = \frac{\text{Full scale system throughput}}{\text{Scaled down system throughput}}
\]

Throughput (total fluid filtration volume required), flow rate, and process time requirements must be met by the filtration system.

If the system involves prefiltration stages, the prefilter area will usually be equivalent to the final filter area; however, the size of each stage in a multistage system will depend on the overall requirements for practical and economical filtration of the product. Further optimization is possible by performing a filterability test with the prefilters alone. The issues associated with the sizing of a filter system with pre- and final filters are described next.

Figure 16 illustrates the behavior of a typical optimized (ideal) filtration system. The differential pressure ($\Delta P$) is plotted as a function of volume filtered. The limiting or maximum differential pressure and the total batch volume are indicated on the graph. Ideally, the batch volume (throughput requirement) is reached before the limiting $\Delta P$ is reached. This allows a safety factor in the filtration and permits complete processing of the batch without change-out of the filter.

Figure 17 shows the effect of limiting $\Delta P$ versus volume for a filtration system in which the prefilter is too coarse and is not able to remove a sufficient number of particles in the fluid (the particles go through the prefilter instead...
of being retained by the prefilter). The prefilter is unable to adequately protect the final filter. Representative curves are provided for the prefilter, the final filter, and for the combined system. The combined system curve is a sum of the differential pressures for the final filter and the prefilter. The combined system reaches the limiting differential pressure (plugging) before the complete batch is filtered. Thus, a finer prefilter or additional final filter surface area would be required for this batch volume to be filtered at differential pressure less than the limiting differential pressure.

Figure 18 illustrates what occurs in a system in which the prefilter area is too small, but performs its protective function. In this case, the final filter did not reach the limiting ΔP, but the prefilter did. Additional prefilter area would allow the filtration to be optimized. Figure 19 contains representative curves for an optimally staged filtration system. In the optimized system, the combined system is able to completely process the batch at a differential pressure lower that the limiting (plugging) differential pressure. The differential pressures across the final filter are slightly higher that the differential pressure of the prefilter. This is the desired situation for a system that requires prefiltration.

Depending on the knowledge of the filtration process for a particular fluid, it may be desirable to perform a side-stream or a pilot-scale test under process conditions before installing the system at full scale. Further optimization of the system may be possible based on the results of the side-stream test. A side-stream test is recommended when the scaled-down test was limited in its ability to match the process conditions for any stage of the filter system.

System Design

Once the filters have been specified, the system components must be considered. A filtration system will typically contain the following filters, filter housings, gaskets or O-rings, pressure gauges, thermocouples, pipe or tubing, connections for the piping, and a pump (or gas pressure source). Some pharmaceutical systems may require an inert gas for product stability as well as pressure. The system components must be compatible with the process fluid, the full-scale process fluid temperature, the process pressure, and the sterilization method.

Most sterile filtration applications require a sanitary design. In a sanitary design the components (e.g., sanitary valves) and materials of construction are selected for their
ability to prevent the buildup of contaminants in the system. For a sanitary process, 316L stainless steel is typically used for portions of the system that will have product contact, provided it is compatible with the process fluid. Internal surfaces should be finished to an appropriate microinch \( R_s \) specification (e.g., a minimum of 20 to 25 \( \mu \text{in.} \)); external surfaces should be mechanically polished to a high quality sanitary finish. Both internal and external surfaces should be electropolished for a smooth adhesion and corrosion resistant surface.

System welds should be constructed in such a way that the weld porosity is minimized and the joints are of high quality and are clean. Any internal welds that have product contact should be ground smooth and flush to reduce the potential for the buildup of material from the process steam. Proper weld procedures need to be followed.

Only sanitary fittings should be used in the system to ensure sterilizability of the seals and cleanliness. Any ports in the system, such as those required for vents, drains, pressure gauges, and thermocouples, should have a sanitary stem or sanitary valve design. Absolutely no threaded connections should be used in the portions of a sanitary system where there is a potential for contact with the process fluid. Threaded connections can lead to system contamination because of a potential buildup of contaminants.

There are a number of considerations for the filter housing. housings are essentially pressure vessels, and as such, may be subject to the appropriate ASME boiler and pressure code. A housing should be rated for the appropriate pressure and temperature with a safety margin for the process. A full vacuum rating may also need to be considered, especially for housings used in vent service or lyophilizers. The design of the housing as well as the physical placement of the housing in the production facility, must facilitate filter change-out. The size of the housing must meet the requirements for the flow rate and differential pressure process specifications. Inlet and outlet ports should not lead to excessive pressure losses.

The system and housing design should not contain any portions through which there is no flow (dead legs). All product contact areas need to be accessible for cleaning purposes. Drains should be designed and positioned to minimize holdup, or retention, of the process fluid in the system.

The filter assembly should be optimized for use in a sanitary system, especially where aseptic processing is required. For example, the filter should have an O-ring sealing mechanism in the filter housing instead of a gasket sealing mechanism. For some small batch applications, disposable filter assemblies (in effect filter plus plastic housing) can be used. These assemblies should have proper sanitary fittings.

For most sterile filtration systems a appropriate pump (e.g., a positive displacement pump or a centrifugal pump) is used to provide pumping capacity. The components of the pump head should be of a sanitary design and compatible with the process fluid.

Pressure gauges should be installed upstream and downstream of each filter assembly in the system so that the differential pressure across the filter can be monitored. The ability to monitor \( \Delta P \) can be used to indicate filter life and that filter change-out is required. For some applications it is desirable to have temperature probes or thermocouples installed in the system. During an in situ steam sterilization of a system, the ability to monitor the temperature is needed to ensure that the temperature conditions for steam sterilization are met. For a process that requires a temperature other than ambient, the temperature of the fluid should be monitored. Temperature measurement should be performed immediately downstream of the filter installation to ensure that the filter installation is at the correct temperature.

Throughout the system, sample ports may be required to allow for sample collection during a process. It is important to ensure that the collection of a sample does not cause contamination in the system. This can be attained by using sample ports that are fitted with septa. For systems that are not hard piped (plumbed with permanent connections), the ability to make sterile (or aseptic) connections must be available.

**Sterilization**

In sterile filtration processes, the downstream side of the filter must be sterilized and must remain sterile during the process. The sterilization process must be properly validated to ensure that the sterile condition is met for a given system. Filters can be sterilized by a number of techniques, including in situ steam sterilization, autoclave sterilization, irradiation, and chemical treatment (hydrogen peroxide vapors or ethylene oxide). For any method, it is important to ensure that the sterilization is adequate for the system and that the technique does not damage the filter. The Parenteral Drug Association (PDA) published a technical monograph on the validation of steam sterilization cycles (28). The majority of sterile processes employ in situ steam sterilization or autoclave sterilization and will be the subject of this discussion.

Methods that involve steam are validated through the use of thermocouples or biological indicators to ensure that the system has been properly sterilized. The maximum steam temperature for most sterilizing grade filter cartridges is 140 °C; the maximum conditions for temperature and pressure should be obtained from the filter manufacturer. Typically, the steam pressure is held at a minimum of 15 psi at 121 °C for a minimum of 30 min. Although steam temperatures up to 140 °C can be used, the service life of the filter may be shortened, especially where multiple uses require repeat sterilization.

A decision must be made on the appropriate sterilization technique for the process. Larger systems tend to be in situ steam sterilized, while small volume systems tend to use autoclave sterilization. A system that utilizes in situ steam sterilization can be automated, which is becoming a common practice for a new system.

**Integrity Testing Design and Operation Considerations**

The issues that must be considered for integrity testing of a filtration assembly include the technique (forward flow, pressure hold, bubble point, or water intrusion for hydrophobic filters), the wetting fluid (water or product wet), and the test time (after installation, after sterilization, after
filtration) and the test location (in situ or off line). Once a plan for integrity testing is established, the system must be designed so that the integrity test regimen can be properly carried out.

Field experience with the integrity testing of sterilizing grade filters shows that various combinations of integrity test procedures are in use and that different testing schedules, both pre- and postuse, are followed. In some sterile processes, redundant filters are used for convenience in integrity testing or for increasing assurance of sterility. For others, filter usage is based on filter life studies. A common approach is to perform the test or combination of tests that meet regulatory guidelines and that provide the highest degree of accuracy commensurate with the economics and the practicalities of the process.

The highest degree of confidence in the maintenance of filter integrity can be attained by routinely integrity testing the filter system, either by the forward-flow or pressure-hold integrity test, before sterilization, after sterilization, and after filter use.

Pre- and poststerilization integrity tests are recommended, because both tests can provide information about the process conditions. However, if only one preuse integrity test is to be performed, then an integrity test after sterilization will provide more useful information, because improper sterilization procedures can lead to filter damage. Regardless of any preuse testing performed, a postuse test is always recommended. The assurance of bacterial retention during the filtration of critical fluids, such as parenterals, biological liquids, and media for fermentations, is extremely important and, at the very least, filters are integrity tested after use.

It is often desirable for a filter user to perform a product–wet integrity test after the filter has been used to filter the product. Product–wet testing allows the filter user to perform the integrity test under actual process conditions and does not require a flush step to remove the product from the filter membrane. Flushing out the product and testing with water can lead to a false failure if there is incomplete removal of the product that may have a lower surface tension than water and thus reduce the bubble point to below the test pressure required for a water-wet integrity test.

Appropriate integrity test parameters for specific applications should be obtained from the filter manufacturer in writing because the filter manufacturer can provide the appropriate test parameters for a fluid and can ensure the proper relationship to the claims for the specific filter. The filter user should confirm performance of the tests before incorporating the test parameters into their standard operating procedures. If any anomalies are noted during confirmation, then these must be resolved.

Integrity test parameters can be provided with either air or nitrogen (or other inert gas) pressurization. Carbon dioxide cannot be used as a test gas, because the diffusive flow through the filter will be extremely high and the gas is reactive. Pure oxygen is typically not acceptable as a test gas, because it is a strong oxidizer. Process conditions, including temperature and gas used for pressurization, must be considered for the issuance of integrity test values.

BIBLIOGRAPHY

FLOW CYTOMETRY

INTRODUCTION

Cytometry is a technique for measuring physical or chemical characteristics of single cells (or other biological particles). Such measurements are usually performed with the aid of a microscope, but there are severe limitations both in the number of cells that can be investigated microscopically and in the number of features that can be observed simultaneously.

In flow cytometry (which avoids these limitations), the cells pass the measuring volume in a fluid stream, preferably in a single file, and in a short time. The first flow cytometers, developed in the 1940s, were based on the Coulter principle, that is, the detection of changes in the electrical conductivity or impedance of a small saline-filled orifice as cells pass through.

The history involved leads most people to think of flow cytometers as instruments that measure optical signals from cells. These optical signals result from light being scattered by the cells (differentially, according to cell type and status) and the fluorescent light emitted by different cell compartments when excited by light of a certain spectral range. The cellular parameters detected can be characterized as intrinsic or extrinsic, depending on whether their measurement requires use of reagents or probes.

A cytometer consists of five main components:

1. Illumination optics. Usually consisting of an arc lamp (high-pressure mercury) or a laser (argon ion) combined with crossed cylindrical lenses to focus the light.

2. Flow cell. A device for focusing the cells in a linear file by hydrodynamics. The core stream with the cells travels in a laminar flow to the illumination point with the aid of a sheath stream of clean water. The main types are stream-in-air systems and variations thereof passing through a (flat-sided) quartz cuvette.

3. Collection optics. Usually consisting of a combination of optical filters for spectral separation, generally set perpendicular to the illuminating beam and in four spectral ranges (green, 510–540 nm; yellow, 560–580 nm; orange, 605–635 nm; and red, 650 nm and above).

4. Detector electronics. These consist of photodiodes or photomultiplier tubes (PMTs). The particles of inter-
est in biotechnology, which are frequently very small, (e.g., bacteria), emit only a small number of photons, typically about 1,000 in passage through the detection chamber. Therefore, an appropriate combination of high-power illumination optics and sensitive detectors is necessary for the application of flow cytometry in biotechnology.

5. Data analysis. The first flow cytometers displayed and stored the data in multichannel pulse height analyzers. Now the data are stored as standard list mode files in PCs, ready for statistical analysis and sophisticated visualization.

The result of a flow cytometric investigation is a frequency distribution, that is, a histogram. In the channels of this histogram, cells are counted according to the light intensity detected, which corresponds to the relative size or quantity of a specific cellular compartment or component, such as the DNA content. Measuring rates up to 10,000 cells s\(^{-1}\) are possible, although about 100 cells s\(^{-1}\) is more usual.

Flow cytometry is not recommended for measuring the absolute content or concentration of cell compartments. However, it is the best method for classifying cell types with different physiological properties and for selecting and quantifying subpopulations in a cell ensemble.

If subpopulations of (vital) cells are to be chemically analyzed or cultivated further, the information provided by flow cytometric measurement can be used for selecting appropriate subpopulations by an electrical or mechanical sorting process. Depending on the type of flow cell and the type of illumination involved, three basic constructions are used:

- Stream-in-air
- Contained flow chamber
- Dark-field illumination

The devices range from large constructions for research purposes to small and inexpensive laptop machines on the way to an online approach. The development of new flow cytometric methods is often accomplished using special modular constructions ("cyto mutt" comp. [1]).

A further development is the slit-scanning procedure, which allows the shape of the signal generated by the cell passing through the illumination beam to be used for gathering additional information. In this way, morphological features and the distribution of cellular compartments can be recorded.

Special demands have led to appropriate modifications. The distribution of aquatic microorganisms can be monitored in the size range of 0.1 to 2000 \(\mu\text{m}\) by such a device, for instance (2).

Flow cytometers are best used to answer questions concerning the dynamics of subpopulations in a cell population, for example, cells in a mammalian tissue that have a higher DNA content than that of G1-phase cells. Thus, in medicine, flow cytometry was first applied in oncology, where it was quickly (and frequently) used in cancer diagnosis. These results generated a solid basis for mathematical modeling of cell-cycle regulation in eukaryotic organisms.

Using immunofluorescence and multiparameter analysis, flow cytometry became a successful, and standard, method in laboratory hematology and clinical immunology in the 1980s. Several commercial devices are now available, constructed specifically for medical purposes. An excellent and comprehensive documentation of the method of flow cytometry is given by Shapiro (1).

The use of flow cytometry in biotechnology, and especially in bacteriology was for a long time not as successful as it was in medicine. After early inspiring investigations by Bailey et al. (3), Slater et al. (4), and Hutter and Eipel (5) further development seems to have been delayed. There are probably three major reasons for this:

- The high cost of the method
- The small size of the cells investigated (an Escherichia coli cell has only \(10^{-3}\) to \(10^{-4}\) times the size or DNA content, and consequently the potential fluorescence energy gain, of a mammalian cell)
- The belief of most biotechnologists that analysis of physiological state distributions do not have an important role to play in understanding the dynamics of bioprocesses or for running them safely and economically

The first two handicaps have now been overcome because powerful and inexpensive microcomputers have considerably reduced the cost of the technique. Flow cytometry has also benefited from new methods of data analysis and mining, such as neural network analysis, which allow a comprehensive evaluation of the rich and complex information gained by flow cytometry (6). Furthermore, new staining methods combined with improved illumination, detection, and amplification have made flow cytometric analysis highly reliable, even for bacteria and microplankton as small as 1 \(\mu\text{m}\) in diameter.

Nevertheless, the key to more widespread use of flow cytometry in biotechnology is more general acknowledgment of the importance of a segregated model concept. Toward this end, Fredrickson et al. (7) proved that the statistics, the distribution of cell populations, and the dynamics of bioprocesses are inseparably connected. Moreover, Munch et al. (8) demonstrated, with a sophisticated approach to flow cytometry, that knowledge concerning the distribution of states in the cell cycle is essential for deeper understanding of the growth dynamics of yeast and the metabolism of a biomass.

**BACTERIAL CELLS**

Bacteria exist as dynamic and diverse populations in nature. Even the cellular heterogeneity within a pure bacterial culture is far greater than previously assumed. To understand bacterial ecology in nature and in bioprocesses, new techniques are required for answering questions that could not be solved by traditional methods alone. Now more than ever, flow cytometry has been recognized as a valuable asset for microbiologists and is perceived to
be an important and indispensable tool for investigations on the microenvironmental level.

Bacterial flow cytometry enables visualization of cell states and allows the analyst to follow growth, death, replication, cell division, metabolism, and surface phenomena, greatly enhancing the ability to understand and control cell physiology. A great assortment of fluorescent stains is commercially available; their numbers have rapidly increased recently. It is not possible to give an overview of their application in this article, but Lloyd (9) provides a fine introduction to this topic. Summaries are given in Refs. 10 and 11, too.

Generally, however, it should be realized that these techniques need to be handled more carefully for use with bacteria than with eukaryotes to avoid unclear results from artifacts arising either from sampling or from analytical errors.

Growth and the Cell Cycle

All survival strategies should be programmed, above all, to safeguard the genome for the future. Results obtained by flow cytometry allow conclusions to be drawn and tested about how the physiological state of cells is connected with survival strategies under changing environmental conditions. Consequently, the most obvious targets for bacterial staining are the nucleic acids. Steen and Boye (12) and Skarstad et al. (13) presented a model for calculating the relative rates of DNA synthesis per cell from studies based on E. coli strains. By means of this model and flow cytometric evaluation of the DNA distribution, the generation time (\(T\)) and the initiation and termination of the replication and postreplication periods (C and D, respectively) can be estimated. Features of replication (e.g., the number of initiation origins) seem to be dependent both on the genetically determined disposition of the bacterial strain analyzed and the growth conditions leading to the initial cell mass. In unperturbed bacterial cultures with high growth rates, several replication forks may be initiated in each cell, if the growth conditions are optimal. Then, after treatment with antibiotics, E. coli cells will contain two, four, or eight fully replicated chromosomes. In contrast, bacteria growing slowly, at doubling rates of less than 1 per hour, exhibit an increase in the prereplication period as well as in the C and D periods. This effect can be amplified by further decreasing the growth rate using chemostat cultures under limited growth conditions (12,13).

Besides the considerations just mentioned, it must be stressed that during limited, and even unrestricted, growth, the behavior of individual cells differs. As pointed out by Åkerlund et al. (14), each cell goes through different discontinuous processes in its life (e.g., the cell cycle). They are therefore different both in size and metabolic activity. This is generally true, except in steady-state conditions, to some extent, where only the distribution of the cells and the cell size are time invariant.

Considering these results, it is clear that key events in cell physiology, including initiation of replication and cell division, are tightly bound to the microenvironmental conditions in the surroundings of the bacterial cell. Obtaining information about the replication behavior of biotechno-logically useful bacterial strains is, therefore, a crucial prerequisite for controlling industrial bioprocesses.

This knowledge could also be used for investigations into the population dynamics of marine bacteria. Because of fluctuating climatic influences, these organisms must constantly adapt to changing conditions and move between states of growth and starvation. In such cases, major alterations in the level of other cellular components, such as lipids and proteins, may be observed in addition to changes in the nucleic acid contents.

Bioprocessing

Progress in bioprocess engineering depends ultimately on the level of understanding and control of the physiological state of the bacterial population being exploited. Process efficiency is strongly sensitive to changes in the cell state. Characterization of these states via fluorescence monitoring, evaluation of the data obtained, and subsequent regulation of the process regime by controlling the surrounding natural or artificial conditions is one of the most ambitious tasks in the near future for biotechnology. Flow cytometry appears ideally suited for determination of such parameters as cell concentration and size as well as for visualizing intracellular performance in industrial processes promoted by bacteria. Using this technique, the state of the cellular system can be precisely monitored, allowing optimization of process efficiency in production or degradation of the target substances (15).

Although bioprocess control on the cellular level is not being widely practiced as yet, some processes are now being adjusted by flow cytometry. One main limitation seems to be the high content of particulate constituents in most industrial media, because discriminating between microorganisms and noise signals from media is not a trivial undertaking.

Microorganisms respond to an unbalanced supply of nutrients or deviations from optimum physical factors by increasing the synthesis of intermediates. This phenomenon is called overflow metabolism. The metabolites overproduced can be excreted or accumulated intracellularly. Flow cytometric methods have been established for detecting intracellular accumulation of polyhydroxybutyrate (PHB), a typical overflow metabolite in many bacterial strains (16). PHB serves as an energy and carbon reserve and has received attention as a thermoplastic and biologically degradable polymer. Furthermore, researchers are searching for novel bacterial strains that may accumulate novel types of polyhydroxyacid (PHA) possessing better physical and mechanical qualities. Some bacterial yield up to 90% of their dry weight in this material. The inclusion bodies so formed alter both the size of cells and their light-scattering behavior. The effect of different cultivation conditions on PHB formation has been investigated in Ralstonia eutropha and recombinant E. coli transformed with PHA-synthesis genes. In these cases, cell-sorting technologies can be used to isolate highly efficient strains at a high speed. Flow cytometry can also be used to quantify heterogeneity of PHB production and accumulation, as found in Methylobacterium rhodesianum cultures. These investigations are based on the idea that growth and product for-
mation (overflow metabolism) are coupled to specific cell states, which are an expression of maturation of individuals and are connected to stages in the proliferation cycle. It has been found that under growth-limiting conditions, the cells first go through the DNA replication program, thereby safeguarding the genetic information by doubling the chromosome content. Doing this, the organisms maintain the chance of restarting multiplication as a forward strategy of survival if better conditions arise. Cells only lay down PHB as an energy reserve in this kind of situation.

Introduction of foreign genes into bacterial cells enables synthesis of desirable products such as proteins, lipids, and a wide variety of other biologically active compounds. Normally, the expression of these genes alters the normal pattern of interaction and synthetic activities in the host cell. It may be observed that highly active cloned gene expression or extremely large plasmid content leads to a markedly slower growth rate in recombinant cells. Generally, however, some of the recombinant cells revert during prolonged cultivation, either (in the case of plasmids) through defective partitioning during cell division, or through changes in plasmid structure. By this process, a subpopulation of nonrecombinant cells develops that has a higher specific growth rate. In chemostat experiments, the recombinant cells are usually replaced by this phenomenon, and the plasmid-free population comes to dominate the dynamics of the reactor. It is generally believed that the production of heterologous products is metabolically harmful to the cell. Occasionally, it has been proposed that the decrease in the growth rate seen is more likely to be a result of deleterious and injurious interaction between the expressed product and some component of the cell.

Flow cytometry can be used to detect segregational instability of a plasmid-containing expression system as well as to estimate the physiological state of the cells involved. There have only been a few biotechnological studies investigating the stability of recombinant cells on near-industrial scales. By way of illustration, however, the method has been used to examine recombinant E. coli cells forming protein inclusion bodies via wide-angle light-scattering measurements.

Fluorescence-activated cell sorting has also been found to be capable of isolating gram-negative hyperproducing mutant cells of Bacillus brevis from wild-type cells. Furthermore, flow cytometry can be used to study the progress of gene expression in a wide range of bacterial cell systems. The expression of foreign genes in bacterial cells or the stability of transformed plasmids can generally be determined by the use of bioreporter genes. These systems allow flow cytometric monitoring of the gene expression even when the gene product is difficult to assay. The promoterless reporter gene is placed under the expression control of the promoter of the foreign gene. A typical example is use of the lacZ gene from E. coli, one of the most widely used bioreporter systems. The lacZ gene encodes the enzyme β-galactosidase, which cleaves a fluorogenic product from specially designed substrates. The emission of the resulting fluorescence is proportional to the amount of the enzyme in the cell, corresponding to the expression of the target gene. Consequently, the β-galactosidase-positive cells exhibit a level of fluorescence that is measurable for each cell by flow cytometry. However, a primary problem in single-cell β-galactosidase assays is leakage of the intracellular fluorescent marker, although a number of methods have been developed to prevent this movement. Other bioreporter systems encode detectable markers such as green fluorescent proteins or surface antigens for monitoring microbial performance.

There have also been attempts to use polymerase chain reaction (PCR) in DNA amplification methods for sensitive detection of specific innate or foreign DNA or RNA sequences inside intact cells. Some of the population seem to remain broken after such treatment, allowing subsequent flow cytometric assessment. For example, PCR can be used to detect plasmid-encoded gene sequences or poorly expressed mRNA in individual bacterial cells.

**Vitality**

In contrast to biotechnologically processes, bacterial cells in nature live neither in pure cultures nor in surroundings of constant temperature, pH, or nutrient availability. Many of the (microscopically) visible bacterial cells are inactive or dying. Every change in the habitat is potentially a source of stress, which results in a change in growth rates. The sum of the cellular responses to stress is a factor associated with the survival strategy.

Responses to stress take place at two distinct cellular levels: first, activation of existing enzymes and transport processes and, second, gene expression. The biochemical bacterial stress response can be precisely determined using flow cytometry, by analyzing changes in the type of duplication, cellular pH, membrane potential, and the amount and kind of storage products. The physiological links between the environment and the genetic systems with respect to survival of stress situations are often mediated by expression of proteins involved in a range of tolerance mechanisms. It is commonly accepted that specific genetic programs exist for prolonging the survival of non-growing bacteria exposed to starvation or stress. Bacterial programmed cell death seems to be mostly related to population development, because the lysed cells are usually essential as sources of nutrient compounds, allowing completion of the developmental cycle of the remaining, living population.

To date, knowledge of the connections between bacterial processes and bacterial community structure in nature is inadequate and restricted by procedural flaws. Most commonly used methods for the detection and enumeration of bacteria have serious limitations because of their incapacity to allow certain bacteria to grow. The reason may be that the bacteria concerned need symbiotic partners, anaerobic microenvironments, other uncommon conditions, or unknown growth factors. However, for a variety of reasons, quantifying the bacteria present in the investigated system, independently of growth, is a fundamental requirement for almost all bacteriological studies. Living cells must be distinguished from dormant cells (induced by starvation) as well as from viable (but not culturable) and truly dead bacteria. To char-
acterize the degree of viability of bacteria, flow cytometric detection of basic cell functions such as reproductive activity, metabolic activity, and membrane integrity is an invaluable tool (26,27). Reliable flow cytometric methods for viability assessment that circumvent the need for culture were largely developed by Porter et al. (28). Investigations associated with the state of dormancy in bacteria, in which the cells survived for extended periods of time without growth or multiplication, have been performed by Kell et al. (29). Cationic lipophilic viability dyes should be handled with care in such studies, because they are very often extruded in an energy-dependent manner. There are several indications that a transport system is involved. Excellent correlations can regularly be observed between viability of the cell and the ability to translocate the dye to the cell’s surroundings.

A wide variety of bacteria are able to metabolize xenobiotics as sources of carbon and energy for growth and to decontaminate polluted ecosystems in this way. However, the effects these toxic substances have on the artificially introduced bacterial species should be considered in such biotechnologically forced processes. Above certain, critical concentrations the chemicals become toxic, the bacteria are poisoned, the bioremediation process slows down and finally stops. It is well known that most of such substances are membrane active and burden the cell energetically. Consequently, cellular changes in the energetic state and the form of the membrane potential must be analyzed with the solute transport, ATP synthesis, and pH homeostasis. To control and optimize such processes, information is necessary about death characteristics, which depend on the physiological state of the cells (30). Flow cytometric assessment of the membrane potential allows a rapid and differentiated analysis of the ecotoxic potential of pollutants toward bacterial cells and detection of the rapidly changing physiological behavior of the cell in growth-phase-dependent variations of the membrane potential (Fig. 1). Very little information exists regarding the members of bacterial communities with respect to morphology and taxonomy that are responsible for the decontamination process. There is also little information about differences in the physiological state of specific strains under certain conditions. For this reason, methods must be developed that relate specific cellular parameters to taxonomic classes (Fig. 2). The first such investigations of this type were done by Herrmann et al. (31).

In medical science, situations also arise in which living and dead cells must be distinguished. In this way, flow cytometry can be used in the assessment of biocidal drugs, offering rapid assays for prokaryote drug susceptibility. The aim is to use the technique in the clinic for rapid susceptibility testing as an aid for choice of therapy. Common methods for investigating resistance to antibiotics and disinfectants and their effects on bacterial strains are based on estimating the minimum growth inhibitory concentration (MIC) of the compound or the decrease in colony-forming units after exposure. These conventional techniques need at least 24 to 48 h to get information about the action of antibiotics.

Flow cytometry is much faster and, furthermore, gives considerable advantages in drug research for investigating the mode of action of novel antibiotics. The susceptibility of bacterial cells to antimicrobial agents may be so profoundly influenced by age, growth, and metabolism to be heterogeneous within a population. The permeability of the bacterial membrane changes during the life cycle, for instance. Recently, it has become apparent that the permeation of lipophilic substances through the envelope of Gram-negative bacteria has great potential in the development of new chemotherapeutic agents. Generally, using flow cytometry, key drug-induced changes in light-scattering behavior, uptake of vitality stains, DNA ploidy, and protein patterns can be estimated (32,33).

Mixed Populations

Bacterial ecology requires the application of new techniques to help address problems that cannot be solved by traditional methods alone. Efforts to demonstrate the value of flow cytometry in identifying different species within complex populations have relied on immunofluorescence-based methods such as the application of antibodies (34), lectins (35), or rRNA-targeted oligonucleotide probes (36). The methods chosen must be applicable to as wide a range as possible of different organisms. These techniques have also found uses in areas such as quality control of water and foodstuffs as well as in soil and water ecology.

Today, fluorescently labeled oligonucleotide probes are most commonly used for monitoring specific strains. The probes are complementary to group-specific regions of the highly conserved multicopy 16S or 23S ribosomal RNA molecules and are bound using in situ hybridization. In this manner, the bacterial strain of interest can be detected without separating it from the surrounding microflora and contaminating debris. Furthermore, comparative sequencing and the design of nucleic acid hybridization probes have served to establish an altered phylogenetic framework.

Differentiation between artificial cultures is possible if the members have contrasting proportions or absolute amounts of guanine-cytosine or adenine-thymine. Estimations of guanine-cytosine content vary from nearly 20% up to 80% (37). Using dyes with different binding preferences to DNA and double fluorescence excitation, composition analysis seems to be workable.

Another way of differentiating mixed populations is the genetic insertion of a fluorescent system to split populations with special physiological capacities.

Food

The use of fluorescent labels is accepted as a successful and sensitive approach for detecting bacteria in food. Lactic acid bacteria are fermentative Gram-positive organisms that are widely used as starter cultures for manufacturing a great assortment of foods and drinks, such as cheese, yogurt, and sausages. The quality of a starter culture is highly important. It must be taken into account that industrial fermentations may involve stressful conditions that can affect bacterial gene expression, arrest cell multiplication, interfere with metabolic activity and survival potential, and lead, possibly, to cell death. Studying responses of lactic bacteria to stresses such as freezing and
Figure 1. Analysis of the cellular DNA content, membrane-potential-related fluorescence (MPRF), and rRNA content of Ralstonia eutropha JMP 134 in relation to the growth rate during batch cultivation on phenol. Samples were harvested at the lag-phase (a), early log-phase (b), and stationary phase (c). The cells were stained with the carbocyanine dye di-OC<sub>6</sub> (3) for MPRF, with oligonucleotide probes for analyzing the rRNA, and with 4’,6-diamidinophenylindole (DAPI) for the DNA. The proportions of cells having two chromosomes is estimated to be 3.88% (a), 87.80% (b), and 38.96% (c).

There are many situations where the total bacterial concentration is the only measurement that is important. Bacteria are a principal cause of food spoilage. Contamination by harmful organisms can cause severe human food-borne diseases, either caused by the bacteria themselves or the toxins released by the bacteria. The time required for conventional tests (plate-count techniques) can lead to substantial delays (of 24 to 72 h), which is a serious disadvantage. Direct epifluorescent filter enumeration is another commonly used technique that allows microscopic
counting of bacteria retained on a filter. Over the past 10 to 15 years, this method has been used extensively for studies on the survival of bacteria in food and for estimating biomass in drinks (39).

However, results suggest that flow cytometry also has significant potential for the detection of pathogenic microorganisms in the food industry. Using this method, pathogenic bacterial cells can be detected by applying fluorescently labeled monoclonal antibodies, for rapidly detecting Salmonella or Listeria, for example. Accurate detection has been demonstrated down to levels of below 10^6 cells/mL or even 1 cell/mL after preenrichment (40,41).

Further Applications

The function of bacterial strains in certain human or animal diseases is increasingly being assessed using flow cytometry. For instance, specific cellular characteristics of bacteria involved in skin diseases are being studied by these techniques (42). Furthermore, flow cytometry offers a rapid method for characterizing anaerobic bacteria present in human fecal suspensions using specific physical and biochemical features (43). Attempts have been made to detect bacterial toxins via fluorescently labeled antibodies bound to polystyrene microbeads in stool specimens (44). Moreover, flow cytometric investigations have been used to identify antisera produced against bacteria capable of specifically binding to surfaces of the target bacteria (45). The resurgence of tuberculosis has caused considerable effort to be focused on developing rapid methods for inhibiting growth of the causal organism by a variety of agents (46). Also, in animal disease investigations, there is growing interest in the capacity of fish-pathogenic bacteria to survive long-term starvation (in a nonculturable mode) in seawater. Knowledge of how long pathogenic bacteria preserve their infective capacity after being released to the environment is clearly of great practical importance (47).

In conjunction with other biochemical and genetical methods, flow cytometry has been used as a rapid method for testing the effects of shear stress, extreme temperatures, action of chemical compounds, sonication, and electroporation on bacterial vitality (48). Bioluminescent E. coli harboring lux genes have been used for detecting environmental pollutants, for example. Furthermore, fluorescence-activated cell sorting is often the best available method for obtaining purified samples from natural environments in order to apply molecular techniques or for extracting or enriching samples for detecting low numbers of specific bacteria, such as pathogenic or genetically engineered microorganisms.

The function of bacterial cell surfaces has been investigated in various ways, such as through the interaction of bacterial cells with electrically charged surfaces, macrophages, or other cells through electrostatic interaction. The investigated phenomena may be mediated by bacterial polysaccharides, which are habitually associated with the outer surface of the bacterium. These polysaccharides form a unique class of polyelectrolytes, possessing antigenic properties, for which a lot of fluorescent monoclonal or polyclonal antibodies are available (49).
Figure 3. Flow cytometric analysis via double fluorescent measurement of the DNA and 3-β-hydroxysterol content of Saccharomyces carlsbergensis during fermentation. The first sample (a) was harvested 20 after inoculation of the brewing reactor and is characterized by a high rate of proliferation. The second sample (b) was harvested the following day, the third (c) after 4 days, and the fourth (d) after 6 days. The fermentation process was finished on the 7th day by carbon limitation. As shown in the last three samples, most of the cells (especially in the case of c) were in the G1 phase of the cell cycle and possessed a limited amount of 3-β-hydroxysterols. The insets show the distribution of the neutral lipid content. A large increase in Nile red fluorescence could be observed after 6 days (d).
YEASTS

Of all the biotechnologically important microorganisms, the yeast Saccharomyces cerevisiae has been the most extensively investigated using flow cytometry to date. As baker’s yeast and producer (along with certain related species) of beer and wine, this yeast is of great economic significance. Nevertheless, the chief reasons for the exceptional interest for flow cytometry in this yeast is that it is a widely accepted model organism for eukaryotic cells, because of its ease of cultivation and its well-developed genomics. Investigations of other yeast genera have been very rare and have not been done systematically.

Growth and the Cell Cycle

Basic research into yeast growth and cell cycling by flow cytometry began in the late 1970s with one-dimensional analysis of cell protein content (by FITC-staining) and cell DNA content (using mithramycin staining) by Gilbert et al. (50), Slater (4), and Hutter (51). Both common batch-growth phenomena and the typical diauxic growth pattern of S. cerevisiae were investigated. These authors elucidated basic rules concerning the timing of DNA synthesis in cells growing at different rates. The first experimental evidence supporting the hypothesis that the eukaryotic G0/G1 phase is the most variable phase of the cell cycle, and that the duration of the S phase is nearly constant and is not greatly influenced by the cell’s environmental conditions, was generated by these flow cytometrically monitored experiments.

In the 1980s, Alberghina et al. (52) and others introduced two-dimensional approaches, involving a light-scattering signal and a fluorescence signal. In particular, methods for the determination of fluorescently stained single-cell protein were added by Schepfer et al. (53). Furthermore, sharper detection of the DNA content and, therefore, of the cell cycle state distribution was also made possible by the introduction of more specific staining methods such as propidium iodide (excitation at 488 nm) and DAPI (excitation in the near UV at 351 nm). The theory that DNA replication of S. cerevisiae is initiated when a cell has reached a certain critical size was proved experimentally by these investigations.

Asymmetric cell division, subcellular characteristics, and numerous features of the intracellular distribution of cellular components were also shown to be detectable by the application of slit-scanning data acquisition (54). A double flow cytometric tag allows the progress of the cell cycle to be tracked in the different cohorts of mother and newborn daughter cells during balanced exponential growth (55).

Bioprocessing

Substantial progress also came from bioprocess type investigations. Because S. cerevisiae populations tend toward autosynchronization under certain process conditions in a continuously stirred tank reactor (CSTR), the changing distribution of states in the cell cycle can be determined by flow cytometry and correlated with the physiological performance of the population. It was demonstrated in this way that the type and rate of substrate utilization and cell compartment synthesis are dependent on the stage of the cell proliferation cycle (53).

In a key paper from 1992, Munch et al. (8) proved on the basis of a comprehensive use of flow cytometric methods that understanding cell cycle behavior is essential for the characterization of growth dynamics in bioprocesses. Using Calcofluor White M2R, the chromosomal and mitochondrial DNA content in the yeast cells could be separately monitored, and the aging process in the mother cells of the budding yeast could be recorded.

Changing respiration coefficients, combined with changes in substrate fluxes and cell processes, such as the storage and mobilization of carbohydrates and the excretion of ethanol, can now be correlated with different stages in the cell cycle.

A further important cell component, with cell-cycle sparking and bulk membrane functions (3-β-hydroxysterol), became recordable by staining with FITC covalently coupled to the amino group containing polyene macrolide antibiotic, Nystatin A1 (56,57).

Brewing

Demands of the brewing industry for ever-faster production, without prejudicing the quality of the product, have made flow cytometry more and more significant in this classical branch of biotechnology. The use of cylindrical cone-shaped brewing reactors with volumes of at least 160 m³ requires increasingly abundant information concerning yeast physiology. Close monitoring of changes in the distribution of DNA, neutral lipid, and 3-β-hydroxysterol contents in Saccharomyces cells during propagation, fermentation, and storage enables time-saving process control (58).

DNA distribution monitoring has shown that only cells with a single chromosome content produce ethanol and CO₂ efficiently. On the basis of flow cytometrical investigations, the optimum point for controlling the wort and oxygen supply can be accurately determined. Double-staining techniques prove that cells with a high content of neutral lipids and 3-β-hydroxysterols have a high survival capacity. Only cells with such features are able to generate vigorous subsequent fermentation. However, additional measurements for determining the proportions of living cells or detecting bacterial infections are only sporadically performed in the brewing industry by flow cytometry.

Further Applications

Wittrup and Bailey (59) use flow cytometry to monitor the β-galactosidase activity in S. cerevisiae by measuring the fluorescence decay of the fluorogenic substrate resorufin-β-D-galactopyranoside. With this method, the expression of foreign protein, coupled to the β-galactosidase gene, could be measured, and a segregated model mapping β-galactosidase activity as a function of plasmid copy number was presented (60).

Another method for detecting this protein was a quantitative immunofluorescence procedure developed by Eitzman et al. (61). These authors demonstrated that the ac-
tivity of the enzyme is correlated with the stage in the cell cycle that individual cells have reached.

Peroxisomes are inducible organelles that may occupy large proportions of the cell volume when yeasts are growing on methanol-containing media. Hansenula polymorpha, a methylotrophic yeast, was investigated via side-scattered light (which was found to depend on cell volume, morphology, and structure) and FITC retention (which was largely dependent on the vacuole). A wild strain and a strain partially repressed by glucose were compared (62). Furthermore, it has been shown that the viability of the pathogenic Candida yeasts can be determined by measuring the membrane potential (63) and that of viable blastospores using tetrabromofluorescein (64).

CELLS OF OTHER ORGANISMS

Aquatic Organisms

In aquatic science, flow cytometry can make an important contribution to ecological and physiological studies of natural populations of microbial plankton. It has been increasingly used to analyze natural communities of aquatic microorganisms because of its sensitivity and quantification capacity. To understand the processes that influence the behavior of any planktonic species, information is required on both the variability between individuals and groups and the major environmental factors.

Flow cytometry was introduced in freshwater and marine sciences at the beginning of the 1980s. However, natural samples of aquatic particles are generally very heterogeneous; there are often immense differences in concentrations and cell sizes (from 0.2 to more than 20,000 μm in diameter) as well as in cell types. Additionally, they may be contaminated by inorganic particles, and artifacts may be generated that can make the data difficult to interpret.

However, the application of flow cytometry to planktonology is no longer hampered by shortcomings of commercial instruments. High-resolution and high-speed sorters are available for rapid and precise estimation of cell numbers, differentiation of cell types, and other pertinent problems. Neural network analysis provides an innovative approach for examining taxonomic properties and the state of the marine food web, including its involvement in the surrounding environment.

The prokaryotic fraction of natural marine communities is composed of both heterotrophic and autotrophic organisms. In spite of their very small sizes (0.2 to 2 μm), even the oceanic picoplanktonic cells are now easy to identify. Major groups of photosynthetic prokaryotes can be discriminated by flow cytometry because of their different and, compared to other algae, unusual pigment compositions and their forward light-scatter signal. The heterotrophic organisms are mostly bacteria. Their lack of pigments makes them less easy to autotrophs to analyze by fluorescence methods. Nevertheless, flow cytometry has been used to evaluate kinetic constants of nutrient uptake in relation to cell quantity and, additionally, to characterize individual parameters of the marine bacteria (65). Frequently, total counts of these bacteria include a large fraction of non-nucleoid-containing bacteria (ghosts), which may be cell residues of virus-lysed bacteria or remains of protozoan grazing. Nevertheless, simultaneously measured DNA content allows further discrimination between autotrophic and heterotrophic components of the picoplankton. In some cases, it has been demonstrated that the cell cycle of autotrophic planktonic prokaryotes progresses in phase with the daily light cycle (66), in contrast to cyanobacterial populations, which apparently do not adjust their circadian timekeeping.

Eukaryotic algae exhibit a strong natural autofluorescence emanating from chlorophyll or other pigments such as phycoobiliproteins. The simultaneous measurement of DNA and chlorophyll fluorescence, together with cell size, is necessary to distinguish subpopulations of phytoplankton. Flow cytometry allows the quantification of changes in cellular pigmentation (e.g., chlorophyll and phycoerythrin concentrations) of divergent phytoplankton subpopulations. Such alterations arise in response to nonsteady light or nutrient regimes, such as may be encountered along a depth profile.

Information on the physiological status or response induced by different, artificial treatments can be easily assessed by flow cytometry. Biomass production in phytoplankton and its dependence on environmental conditions can be detected via measurement of the activities of various enzymes such as esterases (FDA) or nitrate reductase (by immunolabeling). The recent development of taxon-specific probes should increase the applicability of flow cytometry for rapid identification of cultured phytopico- and nanoplanktonic strains, especially those that lack taxonomically useful morphological features. Moreover, rRNA probes may provide new information on species that are not amenable to culture in the laboratory (67).

Furthermore, contemporary research has shown that protozoa are important in the dynamics of the aquatic system. Grazing by heterotrophic protozoa has now been accepted as the most important factor responsible for limiting the numbers of bacterial cells in aquatic ecosystems and maintaining nutrient availability. There is evidence suggesting that bio remediation processes involving bacteria may be strongly inhibited in the presence of phagotrophic protozoa, for instance. Sometimes, microbial population growth rates are difficult to ascertain because ubiquitous grazers remove cells as quickly as they are produced. These taxonomically heterogeneous organisms (including ciliates, amoebae, and a variety of flagellates) can graze voraciously on phytoplanktonic and bacterial cells that form the foundation of the marine food web. Discrimination has been achieved between prey cells and a dinoflagellate through DNA staining and isolation of the eukaryotic nuclei. Furthermore, differentiation of phagocytic activity within populations of protozoa is measurable using fluorescently labeled microbeads (68).

Some investigations have been performed on viruses, regarded as ubiquitous, and biologically active members of marine and freshwater microbial communities, in which the viruses may be responsible for the destruction of a considerable proportion of the bacterial and cyanobacterial populations. In some cases, virus and host support each other in a symbiotic manner, in contrast to the observed
annihilation of alternative host populations within the microbial community.

Plants

Application of flow cytometry and cell sorting procedures to study higher plant cells required methods for the production of single-cell suspensions. The need to probe the performance of protoplasts was the main motivation for developing methods for analyzing plant cells. This also required a series of modifications to the flow cytometric instrumentation, because plant cells were sometimes found to be much larger than those of other living organisms.

Flow cytometry has been accepted as a rapid and trustworthy technique for genome analysis; for obtaining information about the size, ploidy, and aneuploidy changes during plant evolution; and for the differentiation and state of replication of plant cells. It provides an accurate method for determining the proportions of cells in the G1, S, and G2/M stages of the cell cycle and for calculating cell cycle times. Consequently, cell-cycle-dependent events such as the expression of distinct proteins can be observed in synchronized cell cultures. For example, the expression of a Nicotiana tabacum cyclin, a homologue to the cdc2 gene, disappears in the G1 and S phases of the cell cycle, but is present in the G2/M phase.

Flow karyotyping from mitotic cells and chromosome sorting provide opportunities for gene mapping and the construction of chromosome-enriched DNA libraries. Thus, the analysis and sorting of plant chromosomes is of considerable economic interest. Also, the occurrence and extent of polyploidization, which often accompanies differentiation of plant cells, is important for understanding the regulation of gene expression in differentiated tissues. Furthermore, in situ hybridization methods allow accurate quantification of copy numbers of repetitive DNA sequences on different chromosomes within a species.

Examination of intra- and interspecific variation of DNA content can be important in plant hybridization and genetic manipulation programs. Therefore, the method is widely used in molecular cytogenetic research. For example, it can be used to search for foreign chromosomes added to a species, as in the introduction of resistance genes into wheat from the germplasm of rye. Moreover, one of the major problems of using wheat–rye addition lines is their genetic instability. The absence of the easily lost rye chromosomes must be detected, but counting the chromosomes by methods other than flow cytometry is time consuming.

For some properties it is necessary to obtain data concerning distinct compartments within a plant cell, and a fundamental goal is appraisal of the physiological behavior of plant cell protoplasts. For this, there are a number of flow cytometric methods for assessing different intracellular components, such as reserve metabolites as well as enzyme activities. Changes in the cytosolic pH using common pH-specific probes have also been shown using these methods. Indeed, it has been proved that analysis of the quantity and quality of cellular fluorescence can yield important information concerning tissue status and derivation.

Furthermore, plant flow cytometry can be a powerful tool in the examination of symbiotic systems. For instance, Azolla filiculoides harbors phototrophic, aerobically grown cyanobacteria in vegetative cells and nitrogen-fixing anaerobic cyanobacteria in heterocysts. Some of these differentiated tissues have been investigated regarding the internal distribution of the blue-green algae and the capacities of the various tissues for nitrogen fixation using DNA estimation.

Recently, the green fluorescent protein (GFP) from the jellyfish Aequora victoria has been used as a powerful new vital reporter in diverse plant cells. In higher plant protoplast studies, it can be used to monitor gene expression, signal transduction, cotransfection, transformation, protein trafficking and localization, protein–protein interactions, and cell separation and purification. In plant cell systems, the bright green fluorescence of GFP is visible even in the presence of the red chlorophyll autofluorescence. GFP is very stable in plant cells and shows little photobleaching. Viable cells, marked with GFP, can be recovered after fluorescence-activated cell sorting.

Animal Cells

Flow cytometry is widely exploited in the study of spermatogenesis and routinely applied to estimate viability. It is also used to determine the chromosome content and sex of the offspring. It is a valuable and adaptable tool for research into the genesis of germ cells in both normal and perturbed situations, and it can be used in the control of reproduction, domestic animal breeding, and wildlife conservation. Furthermore, cytogenetic investigation of eukaryotic organisms is important for understanding evolutionary processes and essential for testing evolutionary models. Improved knowledge of the morphology of germ cells, and the motility and fertility of spermatozoa enhance the efficiency of preinsemination technology. Also, functional heterogeneity can be easily detected, as in the assessment of mitochondrial lesions, which are connected to subfertility. The abundance of live cells, altered sperm morphology, and chromatin structure and the resistance of membrane-bound enzymes and membrane permeability to environmental stress factors can also be calculated from flow cytometric data.

In contrast to other methods, the productivity of the flow cytometric separation of the X or Y chromosome-containing germ cells (using highly efficient staining and sorting methods) yields highly purified samples that can be used for sex control of domestic animals. More accurate still is slit-scan technology, which allows the whole X or Y germ cell to be spatially visualized along the direction of flow. This is important, because the intensity of fluorescence detected is affected by the flat ovoid head shape and the compactness of the chromatin. Thus, the orientation of the germ cell during passage through the laser beam, detectable by slit-scan flow cytometry, is critical for a high efficiency sorting procedure.

Another range of applications in flow cytometry lies in improving the control of antibody production. Generally, hybridoma cell lines serve as sufficient carriers. Optimizing the productivity of hybridoma cell lines depends partly
on developing suitable methods for screening and selecting highly productive cultures and on understanding the regulation of antibody production. Surface-antibody immuno-fluorescence can be used to select high yielding cells by sorting and to gauge the efficacy of the whole population by cytometric cell analysis. Furthermore, information can be obtained about the stability of expression, the yield, and the integrity of the antibodies obtained. Antigen production can be improved using cytotactically growth-arrested cells, because they do not need to consume cellular resources for biomass production (73). Besides the hybridoma lines, other cell lines (often from insects) have been systematically developed for somatic cell genetics, expression of transfected genes, and synthesis of hormone-inducible proteins. Toward this end, flow cytometry-based studies on animal cell systems can be used for developing convenient and widely applicable procedures for gene cloning. There are, for instance, protocols for DNA transfection and subsequent selection of rare transfectants by sorting. For such investigations, the green fluorescent protein (GFP) may be inserted. Several new GFP variants have been generated that are brighter or have altered excitation spectra, facilitating the monitoring of the marked cell lines.

Viruses have a particularly important biotechnological role in the expression of both foreign prokaryotic and eukaryotic genes in host cells. However, only a few virus-cell systems can be subjected to flow cytometry until now, though the method is competent for broad applications in detecting and quantifying virus-infected cells. Viruses can be assayed in animal cell cultures by situ hybridization using fluorescently marked oligonucleotides and quantitatively estimated by flow cytometry. Characterization of the accessible virus host systems offers information about their capability for providing adequate quantities of virus outgrowth. The process of productive and nonproductive virus infection in animal cell lines can also be investigated, for example, by in vivo analysis of promoter activity regarding transcription of the virus gene. The staging of the virus can be estimated using a virus reporter gene, whose product is, for instance, immunofluorescently stainable. Another possibility is to insert an insect luciferase gene, encoding a protein with characteristic light-emission properties that are measurable by flow cytometry. High sensitivity and rapidity of the reaction (in combination with the luminescence, which does not naturally occur in most of the cellular systems used) makes the method valuable for flow cytometric estimation of virus production. It is also possible to get information about the reproductive capabilities of the virus-infected and uninfected host systems. Differences can be analyzed between highly susceptible and less susceptible cells. Furthermore, virus-induced DNA synthesis may be monitored by assisted determination of BrdUrd incorporation, using monoclonal antibodies. An overview is given by McSherry (74).

Another biological field where the prospects offered by flow cytometry seem extremely encouraging, besides clinical diagnostics, is pharmacology. Before the application of a drug is allowed for human use, the mode of action at cellular and subcellular levels must be clear. The experimental conditions must be as close as possible to physiologic drug-cell interactions within living target cells. Model systems have been developed to study the operation of curative drug systems. On-line flow cytometry is the most sensitive and fastest method available, and it causes the least possible harm to living cells.

Additionally, the method is appropriate for investigating the mode of action of other biologically active substances. In phytopathological sciences, for instance, insect cell lines are very suitable target systems for studying the cytotoxic potential of microbe-produced cytostatic agents.

**BIBLIOGRAPHY**

FLOURESCENCE TECHNIQUES FOR BIOPROCESS MONITORING

Bioprocess monitoring
Fluorescence
On-line
Optical sensor
OUTLINE

Introduction
Principles of Single-Wavelength Fluorosensors
Application of Single-Wavelength Fluorosensors
New Developments in Fluorosensor Technology and Application
Two-Dimensional Fluorescence
Biogenic Fluorophors
Instrumentation
Hitachi F4500 Spectrofluorometer
BioView
Applications
Bibliography

INTRODUCTION

In recent years, various bioanalytical systems for improved bioprocess monitoring have been developed and investigated. The application of these systems in pilot facilities and production plants can provide the information needed to better understand all types of bioprocesses, including cultivations, biotransformations, and downstream operations. In addition, better documentation of the processes, especially important for the production of pharmaceuticals, can be obtained. However, most of these sensors are invasive, and thus there is the chance of impacting the cultivation process (e.g., contamination) and causing complications in the processes. For this reason, more and more research groups are involved in the development and application of noninvasive bioanalytical tools, which can be interfaced to the bioprocess without causing problems such as contamination or changes in cell metabolism.

Optical sensors offer tremendous advantages for bioprocess monitoring. They can be interfaced noninvasively to the bioprocess via glass fiber optics. The information can be transported via fiber optics without interference caused by electromagnetic fields over long distances, and signals can be obtained instantaneously, because spectro-optical systems have extremely short response times.

Fluorescence sensors have been investigated intensively during the past 15 years for different applications in biotechnology, especially biomass estimation, reactor characterization, metabolic studies, and general bioprocess monitoring. These so-called fluorosensors were developed mainly for monitoring of the reduced form of nicotine adenine dinucleotide (NADH) and its phosphorylated form (NADPH). This coenzyme is present in all living cells and is involved in different metabolic reactions; its level is extremely sensitive to environmental conditions.

PRINCIPLES OF SINGLE-WAVELENGTH FLUOROSSENSORS

The NAD(P)H pool gives information about the metabolic state of the cells. In 1957, Duyssens and Amesz first showed that relative NAD(P)H measurements could be performed in living cells using spectrofluorometric techniques. UV light of 340 to 360 nm was used to excite fluorescence of NAD(P)H in cells in a suspension. The NAD(P)H fluorescence was monitored at 460 nm while metabolic experiments were performed. The first spectrometer device for biotechnological application was designed by Harrison and Chance in 1970. It could be used in glass bioreactors. The excitation light was guided through a quartz window into the bioreactor, and the fluorescence was collected at an angle of 60° through a second quartz window. The penetration depth of the light was only a few centimeters at low cell densities. The sensitivity decreased as the cell density increased. However, it was possible to monitor aerobic and anaerobic transitions in Klebsiella aerogenes. Zabriskie and Humphrey used this technique in 1978 for biomass estimation in different cultivation experiments (e.g., Saccharomyces cerevisiae, Streptomyces spp., and Thermoactinomyces spp.).

Based on these experiments, different smaller miniaturized process fluorometers were developed that could be interfaced to a bioreactor via a standard port. In 1981, Beyeler et al. developed a miniaturized fluorosensor in an open-end detection mode. This design has the advantage of eliminating the inner filter effect by monitoring the surface fluorescence at the observation window. Three commercial versions of the open-end spectrofluorometer were produced by BioChem Technology/USA (FluoroMeasure), Mettler Toledo AG/Switzerland (Fluosensor), and BioBalance A/S, Denmark (BioBalance).

All these sensors shared two disadvantages. First, they were limited to detection of NAD(P)H fluorescence; it was only possible to excite fluorophores in the range of 340 to 360 nm and to detect the fluorescence intensities at 450 to 460 nm. And they also lacked the ability to distinguish between different fluorophores that might interfere in this region.

In order to obtain more information, a flexible two-channel fluorometer was developed by Scheper and Schlüerl. In this device, the light of the UV lamp was focused outside the optical filters and lenses onto a quartz fiber bundle. This fiber optic was used to guide the light into the bioreactor. The backward fluorescence light was collected via the same fiber bundle and split into two bundles, passing...
interference filters that selected for different wavelength bands. Thus, it was possible to monitor simultaneously different fluorescence wavelengths.

APPLICATION OF SINGLE-WAVELENGTH FLUOROSENSORs

Single-wavelength fluorosensor devices have been used to monitor different cultivation processes, including suspended as well as immobilized organisms (Table 1). Investigations in which the NAD(P)H fluorescence of these cultivations was monitored had one of two goals: on-line determination of biomass concentration or on-line detection of metabolic changes. A significant advantage of using fluorescence for biomass concentration monitoring is that only living cells are detected. To use culture fluorescence as an indicator for biomass concentration, it is necessary that the culture fluorescence per cell is constant during the cultivation process. In this case, a linear correlation between fluorescence and the biomass concentration could be expected. However, the intracellular NAD(P)H pool changes in a cell during the batch experiment. Thus, it is necessary to linearize the corresponding biomass concentration and fluorescence data to obtain a calibration plot. One problem that must be overcome is that all process conditions affecting the culture fluorescence intensity cause problems with the biomass determination. The two most common interferences are gas bubbles and the presence of compounds with fluorescence spectra that overlap that of NAD(P)H. Interestingly, fouling of observation windows was not observed. Many investigators have reported efficient biomass concentration determination via culture fluorescence monitoring. Examples include cultivations of Methylomonas mucosa, Zymomonas mobilis in synthetic medium, and Pseudomonas putida. In all these applications, suspended cell cultivations were studied, and it was shown that online biomass estimation is also possible in technical media under certain process conditions (for detailed information, see Ref. 6).

A wide range of metabolic changes have also been studied using NAD(P)H culture fluorescence (Table 2). For example, Ristroph et al. (8) studied the cultivation of Candida utilis under different conditions and developed a feeding strategy for fed-batch cultivation of this organism on the basis of NAD(P)H-dependent culture fluorescence. Einsele (9) reported on different applications of culture fluorescence monitoring and determined the mixing-time behavior of different bioreactors and the substrate uptake rate of different microorganisms during cultivation. Rendon et al. (6) studied immobilized Clostridium acetobutylicum in a fixed-bed reactor. The cells were immobilized in calcium alginate beads, and the cultivation medium was circulated through the reactor. The growth and productivity of the cells could be monitored under different cultivation conditions. Schepet et al. reported (10) on the application of this technique for the monitoring of immobilized yeast cells.

NEW DEVELOPMENTS IN FLUOROSENSOR TECHNOLOGY AND APPLICATION

Several approaches have been taken in the past 8 years to circumvent the restrictions of single-wavelength NAD(P)H-dependent culture fluorescence monitoring. In 1991, Li et al. (11) reported on a special fluorosensor with five different wavelengths for the monitoring of Saccharomyces cerevisiae and Candida utilis cultivations. They showed that other wavelengths could be used for biomass concentration estimation and for metabolic studies. Horvath et al. (12) used a fluorescence spectrometer coupled to a bioreactor via quartz glass fibers for the determination of S. cerevisiae biomass concentrations. Based on his results, the tryptophan fluorescence was better correlated to the cell mass concentration than was the NAD(P)H fluo-

<table>
<thead>
<tr>
<th>Phenomena studied</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic-anaerobic transition</td>
<td>Kluyveromyces lactis, Candida tropicalis, Escherichia coli, C. guilliermondii</td>
</tr>
<tr>
<td>Aeration rate</td>
<td>Penicillium chrysogenum, Saccharomyces cerevisiae, Candida tropicalis, Escherichia coli</td>
</tr>
<tr>
<td>Addition of carbon source to starved cells</td>
<td>Saccharomyces cerevisiae, Candida tropicalis, Escherichia coli</td>
</tr>
<tr>
<td>Diauxic growth</td>
<td>Saccharomyces cerevisiae, Candida tropicalis, Escherichia coli</td>
</tr>
<tr>
<td>Dilution rate changes</td>
<td>Saccharomyces cerevisiae, Candida tropicalis, Escherichia coli</td>
</tr>
<tr>
<td>Glycolytic oscillation</td>
<td>Saccharomyces cerevisiae, Candida tropicalis, Escherichia coli</td>
</tr>
<tr>
<td>Culture synchrony</td>
<td>Saccharomyces cerevisiae, Candida tropicalis, Escherichia coli</td>
</tr>
<tr>
<td>Metabolic shifts</td>
<td>Saccharomyces cerevisiae, Candida tropicalis, Escherichia coli</td>
</tr>
<tr>
<td>Addition of metabolic uncouplers</td>
<td>Saccharomyces cerevisiae, Candida tropicalis, Escherichia coli</td>
</tr>
</tbody>
</table>

Table 2. Different Phenomena Studied via Fluorescence Monitoring

Source: Based on Refs. 3 and 4.

Table 1. Application of Fluorescence Monitoring for Different Cultivation Processes and Organisms

<table>
<thead>
<tr>
<th>Phenomena studied</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspended</td>
<td>Saccharomyces cerevisiae, Penicillium chrysogenum, Cephalosporium acremonium, Escherichia coli (incl. recombinant), Zymomonas mobilis, Bacillus licheniformis, Clostridium acetobutylicum, Spodoptera frugiperda, Baby hamster kidney cells, Wastewater treatment</td>
</tr>
<tr>
<td>Immobilized</td>
<td>Saccharomyces cerevisiae, Zymomonas mobilis</td>
</tr>
</tbody>
</table>

Source: See Refs. 6 and 7.
FLUORESCENCE TECHNIQUES FOR BIOPROCESS MONITORING

A more recent development was described by Chenina in 1993 (13) for the cultivation of yeasts (Cyathus striatus, Eurotium cristatum, and Crinipellis stipitaria). A special fluorescence spectrometer was developed that can also be interfaced to a bioreactor via light guide cables. Monochromators were used to monitor the fluorescence of wide excitation and emission ranges.

TWO-DIMENSIONAL FLUORESCENCE

These fluorescence spectrometers can be used to produce a two-dimensional (2-D) fluorescence spectrum, which represents the emission spectra at different excitation wavelengths. Such a spectrum is described by the three parameters of excitation wavelength, emission wavelength, and fluorescence intensity, which can be shown as an excitation-emission-data matrix. In this matrix, each row represents an emission scan and each column an excitation scan. Different compounds can be identified by their position in the matrix. The most illustrative way to show a 2-D fluorescence spectrum is to use an isometric plot, but contour plots are the most useful because no fluorescence peaks are hidden (Fig. 2). Dark shadings in the fluorescence plots indicate high fluorescence intensities.

BIOGENIC FLUOROPHORS

The spectral ranges of important biogenic fluorophores, which occur inside the cell and can be detected simultaneously by 2-D fluorescence spectroscopy, are shown in Figure 3. Almost all biological processes depend on proteins, which fluoresce because of aromatic amino acids such as tryptophan, tyrosine, and phenylalanine. In another region of the 2-D fluorescence spectrum ($\lambda_{ex} = 310-480$ nm, $\lambda_{em} = 350-550$ nm), vitamins (e.g., pyridoxine [vitamin B-6], riboflavin [vitamin B-2]), and coenzymes (e.g., NADH, NADPH, FMN, FAD) can be found.

INSTRUMENTATION

For on-line monitoring and control of biotechnological processes using 2-D fluorescence, the instrument must be capable of measuring a complete 2-D fluorescence spectrum in a short time (about 1 min), and the time between two measurements should not be longer than 5 to 15 min. This is to ensure that a complete 2-D fluorescence spectrum represents the fluorescence at a certain time and that changes during a cultivation can be monitored in real time.

HITACHI F4500 SPECTROFLUOROMETER

For 2-D fluorescence measurements, fluorescence spectrometers for use in laboratories can be used. One suitable apparatus is the fluorescence spectrophotometer model F4500 (Hitachi, Japan). This spectrometer allows a very fast scan speed (30,000 nm/min) and an entire 2-D spectrum with a wavelength range of 250 to 550 nm for excitation (10-nm step size) and 260 to 600 nm for emission (20-nm step size) can be measured within 1 min. For selection of excitation and emission wavelength, grating monochromators are used. A 150-W Xe lamp is used for excitation, and a photomultiplier at a voltage of 700 V is used for detection of fluorescence.

The spectrometer is connected by a 1-m bifurcated liquid light conductor (Lumatec, Germany) to a quartz window in a 25-mm port of a bioreactor (Fig. 4). The excitation light is guided via the fiberoptic into the reactor, and the backward fluorescent light is guided via the second fiberoptic to the emission detection unit. With this setup, in situ measurements are possible without interference from light outside the bioreactor, and there is no risk of contamination because the sensor is not in contact with the medium. Furthermore, sterilization is possible without connecting the sensor to the reactor.

BIOVIEW

The BioView sensor (DELTA Light & Optics, Denmark) is a fluorescence sensor optimized for fully automated industrial measurements and applications to monitoring and control of bioprocesses. It is very robust and suitable for harsh environments (e.g., high temperature or moisture).
FLUORESCENCE TECHNIQUES FOR BIOPROCESS MONITORING

Figure 3. Biogenic fluorophores in a 2-D fluorescence spectrum. 1 = riboflavin, FAD, FMN; 2 = NAD(P)H; 3 = pyridoxal, pyridoxamine, pyridoxal-5'-phosphate; 4 = tryptophan; 5 = tyrosine; 6 = phenylalanine). Source: Adapted from reference 9.

and electromagnetic interference. For data transfer, a single-fiber modem is used, which allows a distance between sensor and computer up to several hundred meters. The optical parts of the BioView are optimized for use in biotechnological processes in that interference filters (DELTA Light & Optics, Denmark), rather than spectrometer gratings, are used by the BioView sensor for selection of excitation and emission wavelengths. Filter wheels (one for excitation and another for emission) with 16 filters each are used (Fig. 5). These filters can be designed and chosen individually. Usually, measurements in steps of 20 nm are performed in a wavelength range of 270 to 550 nm for excitation and 290 to 590 nm for emission. For the measurement of a complete 2-D fluorescence spectrum, the excitation filter wheel is set to the first filter position. With this excitation filter, the fluorescence spectrum is monitored via the emission filters that switch from filter to filter until a cycle is completed. Afterward, the excitation filter switches to the next position and the next fluorescence spectrum is measured. This procedure is continued until the excitation filter wheel has performed a complete cycle. The complete measurement takes about 1 min, depending on the number of different filter positions and measurements for each filter combination, which can be chosen individually. The connection to the bioreactor is the same as that described for the Hitachi instrument, except that a 2-m long liquid light conductor is used.

APPLICATIONS

Two-dimensional fluorescence spectroscopy offers the possibility of simultaneous analysis of several fluorophores. The location of the peak maximum in a 2-D spectrum characterizes each fluorophore, and the relative fluorescence intensity correlates with its concentration. Despite this conceptual simplicity, the interpretation of 2-D fluorescence spectra is not trivial. Interactions of the fluorophores with one another and with other nonfluorescent compounds, bubbles, and other physical influences such as pH, temperature, and viscosity, quenching effect, and changes in morphology of the cells all cause problems in the interpretation of the 2-D spectrum.

To make the system user friendly, automated interpretation by artificial intelligence was developed for 2-D fluorescence spectra. The intelligent BioView sensor is able to predict the course of a cultivation if it is taught with a training cultivation under comparable cultivation conditions. Modern chemometric systems like principal component analysis (PCA) or neural networks are able to consider all influences on the fluorescence signal and lead to a sensor with high capacity, able to measure various cultivation variables.

The changes in fluorescence intensities during cultivation processes are shown in difference spectra, which are obtained by subtraction of a 2-D spectrum taken during the cultivation from the one in the beginning. Through comparison to the spectrum of a mixture of pure biogenic fluorophores, such as riboflavin, NADH, pyridoxal, pyridoxamine, pyridoxal-5'-phosphate, and tryptophan, the peaks could be assigned to the expected fluorophores. All changes have a biological origin and offer information about the cell growth and metabolism. Even small changes in the concentration of a single fluorophore can be seen in a complex mixture of several fluorophores. The regions of changes are regarded as the areas of interest.

Measurements made nearly continuously allow us to construct the course of various culture variables during a bioprocess. One of the most important is the biomass concentration. Off-line analysis of counting cells, measuring optical density or weighing dry mass are time consuming and fail if filamentous fungi are used. On-line tools such as a turbidimeter often cause problems. The search for a reliable automated technique for on-line monitoring of biomass still continues. In a 2-D fluorescence spectrum, vari-
Continuous excitation–emission regions can be correlated to biomass concentration. This offers the advantage to change to another region if one cannot be used because of overlapping or other influences. Fluorescence of proteins (because of their high quantum yield, especially in the case of tryptophan), pyridoxine, the area of NAD(P)H, and the area of riboflavin and its derivatives (FAD or FMN) show good correlations with biomass for several organisms (e.g., Acremonium crysogenum, Claviceps purpurea, Sphingomonas yanoikuyae, Bacillus licheniformis, different strains of Escherichia coli, Enterobacter aerogenes, Saccharomyces cerevisiae, Schizosaccharomyces pombe, mammalian cells or Tetrahymena thermophila (9,14). For the filamentous fungi, Claviceps purpurea fluorescence intensity in the region of riboflavin and its derivatives ($\lambda_{ex} = 450$ nm, $\lambda_{em} = 530$ nm) allow on-line monitoring of biomass concentration, which is not possible by optical density or cell count because of the mycelial growth. Scattered light is also measured in a 2-D fluorescence spectrum. These data correlate well with data derived from conventional turbidimetric sensors.

Along with the ability to estimate biomass concentration, 2-D fluorescence sensors such as the BioView system offer on-line information about intracellular variables without interference into the process. This is achieved by monitoring changes in the excitation–emission regions associated with the redox-sensitive coenzymes NADH, FAD, and FMN. This approach can be used to study aerobic–anaerobic transitions, the tricarboxylic cycle, effects of uncoupling agents such as 2,4-dinitrophenol (DNP) on oxidative phosphorylation, diauxies, the effects of pulse additions of substrate, and oscillations of a culture.

Fluorescent products can be observed by 2-D fluorescence. The productivity of ergot alkaloids could be estimated by 2-D fluorescence spectroscopy (9). Degradation of fluorescent substrates such as polycyclic aromatic hydrocarbons can be observed. Other extracellular or physical variables such as the microenvironment of the cells are captured by 2-D fluorescence.

Two-dimensional fluorescence spectroscopy is ideal for a wide variety of organisms and installation uses and can therefore be seen as a multianalyzer that is capable of replacing several other sensors in the future.

**BIBLIOGRAPHY**

INTRODUCTION

Foods and beverages are processed to increase their safety and shelf life, to favorably modify their sensory and nutritional quality and composition, and to improve ease and reliability of preparation by consumers. Foods must be protected from contamination and stored and transported safely and without adversely affecting quality. Food process engineers employ engineering principles and methods to identify, select, design, develop, and “debug” processes, equipment, systems, and plants used to accomplish these tasks. Food process engineers also supervise the start-up and operation of food processing systems and plants; devise related control, maintenance, and sanitation systems; devise ways to eliminate hazards and minimize waste; and help develop new food products and ways to upgrade existing processes. In carrying out these tasks, engineers must account for the biological activity and chemical instability of foods and for the ability of microorganisms and pests to invade and grow in or on foods.

Many different types of food products are produced. The processes used are extremely diverse (1) and too numerous to cover in detail. Therefore, this article examines only the engineering aspects of a limited number of processes or operations. Many of these are used to preserve foods. Others are used to separate food components, and still others are used to change the shape, size, and texture of foods. Food processing frequently involves heat transfer. This article briefly examines such transfer.

Individual operations may produce effects that fall into several processing categories. Thus drying, which separates water from other food ingredients, is used mainly to preserve foods. Drying often changes the texture of foods and involves transfer of heat from hot air to moist solids. It also involves transfer of moisture (mass) within solids and from solids to air and transfers of momentum, which affects airflow.

HEAT TRANSFER IN FOOD PROCESSING

Food process engineers are frequently asked to model heat transfer, devise ways to control or improve such transfer, and calculate heating and cooling loads and sizes of equipment needed to provide desired amounts of heating or cooling. These tasks are accomplished by means of methods...
widely employed by other engineers. However, in using these methods, food process engineers must account for the physical, heat transfer, and sensory properties of foods, and for their temperature sensitivity. Food components rapidly decompose in undesirable ways at high temperatures. The highest product temperature reached during food processing is probably 250 °C, the end-of-roast temperature for very darkly roasted coffee. Maximum tolerable temperatures for many foods are much lower. Higher processing media temperatures may be used (e.g., gas temperatures ranging up to 500 °C are used in roasting coffee). For economic reasons, foods are rarely cooled to temperatures lower than −40 °C.

Foods usually are either solid or complex composites (e.g., moist multicellular tissue, porous materials, stiff doughs) that act like solids. Heat transfer in such foods is modeled by means of partial differential equations based on Fourier’s laws of conduction. Modeling also requires specification of boundary conditions based on surface temperatures, surface heat transfer coefficients, or surface rates of heat transfer. Solutions to partial differential equations for conductive heat transfer and many boundary conditions of practical interest can be found in the work of Carslaw and Jaeger (2) and other standard texts. Similar solutions for partial differential equations and boundary conditions describing diffusive mass transfer equations are provided by Crank (3). Many pairs of heat transfer and mass transfer equations and boundary conditions are formally equivalent. Therefore equation solutions for diffusive mass transfer often can be readily converted into solutions that apply for conductive heat transfer and vice versa.

Heat transfer calculations for solid foods may be complicated by a need to account for heats generated by ongoing biological processes, surface evaporation, anisotropic conduction, and internal vaporization and condensation. Often boundary conditions are spatially nonuniform, or they may vary with time during food processing. The thermal conductivities of foods may be difficult to predict. Nevertheless, most problems relating to heat transfer in solid foods can usually be solved fairly easily by means of computer-based numerical methods.

Food engineers also deal with heat transfer that is due to natural convection in liquid-filled containers, in pore spaces in stacks of heat-generating, respiring produce, and in the air that surrounds solid foods discharged from heating devices. In addition to widely encountered problems involving forced convection of heat in heat exchangers, evaporators, and condensers, they deal with convective heat exchange between streams of air or water and solid foods lined up one behind another in rows on shelves. These problems are often solved empirically. Food engineers working on aseptic preservation and ohmic heating processes are now attempting to precisely predict heat transfer to individual solids in groups of solids suspended in heated, flowing fluid.

Food engineers also deal with prediction of temperature changes during microwave and dielectric heating. In doing so they must account for attenuation and focusing of energy as electromagnetic radiation passes through foods, refraction and reflection of such radiation at food surfaces, energy distribution patterns in food-containing microwave cavities, and temperature-induced changes in the energy absorption characteristics of foods.

**PRESERVATION OF FOODS**

Drying, chilling, freezing, heating-induced inactivation of microorganisms and enzymes, application of bactericides, storage in controlled atmospheres, solute-induced water activity or pH reduction, and combinations of some of these processes are used to preserve foods.

**Drying**

Drying, the evaporative removal of water from moist solids or solidifiable solutions, may well be the earliest method of preserving foods and still serves that purpose today. It is also used to harden foods that have been shaped in a moist state, to reduce food weight, and volume, and to reduce moisture contents to facilitate processing.

Drying preserves foods because biological activity (e.g., microbial growth) radically slows or stops and chemical reactants become immobile when little water is present. Fresh fruits and vegetables usually contain 5–10 kg water/kg dry solids; stable, dried vegetables, 0.03–0.1 kg water/kg dry solids; and stable, sugar-rich dried fruits, 0.20–0.39 kg water/kg dry solids (4). The water content of freshly harvested grains is up to 0.22–0.43 kg water/kg dry solids; to provide safe storage, this must be reduced to 0.11–0.15 kg water/kg dry solids.

**Hot Air Drying.** Food drying most frequently is carried out by transferring heat to moist food particles from flowing hot air. This causes food moisture to evaporate and move away with the air. Consequently, the air moisture content (humidity) rises, and the air temperature drops proportionately. Relative humidity (RH) is the partial pressure of water in air divided by the vapor pressure of pure water at the same temperature. Water activity (a_w) is a measure of the chemical activity of water. It is also the equilibrium RH for a food at a given moisture content.

Food moisture sorption isotherms, such as those shown for apples and potatoes in Figure 1, are constant-temperature plots of a_w versus a food’s dry-basis moisture content, X. Moisture sorption isotherms shift downward moderately as temperature rises. The minimum X obtainable by drying is X_f, the sorption isotherm X at which a_w at the air temperature equals RH.

After start-up transients have decayed, food temperatures during drying depend on a_w at the food’s surface. As long as the surface a_w is greater than 0.9, the food temperature will be no more than 1 °C higher than the air wet-bulb temperature; if air conditions are almost constant, the drying rate will be almost constant. Constant-rate drying does not last very long for foods. Surface a_w depends on X_s, X at the surface. As drying removes water from the surface, water diffusing to the surface from the food’s interior partially replaces it. Because water diffuses slowly in foods, X_s very rapidly drops enough to drive a_w below 0.9. After this, a_w drops sharply as X_s decreases. The food temperature rises and progressively approaches the air tempera-
Wet particles may stick together during drying. This property can be used to provide desired agglomeration. When sticking occurs and is unwelcome, particles are broken apart by passing them between rubber-faced rolls after drying.

**Shallow-Bed Dryers.** Foods are air-dried in many different types of equipment. In conveyor dryers (Fig. 2), moist food particles are uniformly spread as a shallow bed on a perforated belt and conveyed through a series of stages where heated air is blown through the bed and belt. Different air temperatures, humidities, and flow rates are often used in the stages. To improve uniformity of drying, upflow is used in some stages and downflow in others.

In conveyor-based drying of vegetables, 60–85% of the food’s moisture content is removed in the first 1.5–2 h of drying. An additional 2–8 h usually is needed to complete the process. Air temperatures between 66 and 120 °C are used in early stages of drying. Lower air temperatures (32–52 °C) are used later, at the end of drying. Quite different conditions are used for some foods. Low temperatures are used during first stages of gelatin strip drying to prevent melting. Quick-cooking rice is dried in less than 15 min using temperatures up to 200 °C. Temperature around 175 °C may be used to puff diced potatoes at late stages of drying.

In tunnel dryers, hot air passes over shallow beds of food particles conveyed through a rectangular tunnel on racked trays mounted on carts. Airflows in the direction of cart movement or counter to that direction. Clearances between the carts and the tunnel are small, to minimize bypassing of air. Periodically, a freshly loaded cart is pushed into the tunnel and a cart containing dried material simultaneously leaves. Drying tunnels often are used in series. Co-current drying with high inlet-temperature air is followed by counter-current drying with lower temperature air. Conveyor dryers are used instead of tunnel dryers wherever feasible because tray loading is burdensome and conveyor dryers permit automatic loading and discharging. In addition, drying conditions are more readily controlled in conveyor dryers.

**Deep-Bed Drying.** Bin dryers in which air passes through a deep bed of particles may be used to complete the drying of partly dried vegetables discharged from shallow-bed dryers. Grain is often directly dried in deep beds in bins, where air is introduced through a false floor or ducts near the floor (5). Ambient air may be used if its relative humidity is less than 0.55, or it may be heated to 11 °C above ambient. In such dryers, moisture desorption waves (Fig. 3) and accompanying temperature-change waves move upward through the grain bed. Most of the time, the exit air temperature equals the initial grain temperature and the exit air RH equals the initial grain aw.

Design engineers select airflow rates that provide drying rapid enough to prevent mold growth in moist region: 0.007 m³ air/(m³ grain s) may be used for relatively dry feed; 0.08 m³ air/(m³ grain s) may be needed for very moist feed (5).

In other grain dryers, air flows across descending beds of grain that are 0.2–0.4 m thick (5,6). In these cases, air-

**Figure 1.** Moisture sorption isotherms for apples and potatoes.
flow rates of 0.7–2.7 m$^3$ air/(m$^3$ grain s) are used. Inlet air temperatures up to 82 °C are used for animal feeds. Less than 54 °C is used for grain that is to be milled and processed, and less than 45 °C is used for seed grain. Grain is cooled with cool, dry air after such drying. Grain may be partly dried with high temperature air and then transferred to a deep bin where local moisture contents equilibrate without airflow. Final drying and cooling are accomplished by blowing cool, dry air through the grain.

**Pasta Dryers.** Special dryers are used to dry pasta. Freshly extruded pasta contains roughly 0.45 kg water/kg dry solids. That moisture content is reduced roughly 0.13 kg water/kg dry solids to set pasta’s shape and preserve it. Long goods (e.g., spaghetti) are draped over moving rods and carried through the dryer system. Short goods (e.g., macaroni) are conveyed through dryers by simpler means. To prevent adjacent pieces of pasta from sticking together, the pieces are predried using 63 °C, 65% RH air. They leave the predryer and enter a large, final dryer where very humid air is used. Zones in which active drying occurs are followed by zones in which internal equilibration of moisture can take place. To prevent cracking (checking) due to stress buildup, relatively low temperatures (e.g., 50 °C) and very long drying times (up to 18 h) were and often still are used for long goods. Now, based on better modeling of moisture diffusion and stress development, temperatures up to 110 °C and drying times of 5 h are frequently used for long goods. Short goods are less susceptible to cracking and can now be dried in 2 h. Use of higher temperatures also reduces microbial growth.

**Spray Dryers.** Spray dryers are used to dry liquid foods and pumpable food slurries (7). Such foods are converted into drops with mean diameters ranging between 50 and 250 μm by nozzles or rotary atomizers and sprayed into a large chamber. The drops enter at high velocity but soon decelerate. Hot air is blown in and flows in the same direction as the descending drops. One type of spray dryer is shown in Figure 4. Inlet air temperatures as high as 270 °C may be used; often they range between 160 and 200 °C. Outlet air temperatures usually range between 80 and 110 °C. Drop temperatures quickly rise to the wet-bulb temperature of the inlet air and then progressively approach the local air temperature as drying proceeds.

Because air cools as it moves through the dryer, concurrent air-drop flow prevents drops from overheating. Dried particles leave the dryer with temperatures 10–20 °C below the exit air temperature. Large particles drop out of the airstream. Smaller particles are collected in cyclones.
Figure 4. Spray dryer in which large particles drop out of the drying air by gravity and small particles are collected in a cyclone.

Drying times range from 5 s for small drops to 50 s for large drops. Spray-drying involves exchanges of heat, momentum, and mass between the air and the drops, nonlinear diffusion of moisture within drops, and changes in drop size and shape. Complex computer programs have been developed to account for these processes (8). Spray-dried foods are rapidly cooled after drying by sucking cool, dry air into solids discharge lines or by drawing cool, dry air through a screen-supported product bed on a vibrating conveyor.

Freeze-Dryers. Freeze-dryers are used to dry beverage extracts, shrimp, soup ingredients, and military and camper rations. In these dryers, moisture sublimes from frozen food in a vacuum chamber and condenses on very cold refrigerated coils. Heat transfers radiantly from heated panels to the surface of the food and then passes by conduction through a porous dry food layer to a sublimation interface. The interface recedes into the food as drying proceeds (9,10). Vapor created by sublimation escapes from the food through pores produced by prior sublimation.

Temperatures at the product surface and sublimation interface depend on the radiant panel temperature and on the thermal conductivity, vapor permeability, and thickness of the dry layer. Depending on whether heat conduction or vapor permeation controls the drying rate, radiant panel temperatures are programmed either to prevent product surface overheating or to prevent melting at the sublimation interface (11). Small, frozen particles are sometimes used to circumvent drying rate retardation caused by low dried-layer permeability.

Noncondensibles must be removed before and during freeze-drying. If adequate noncondensibles removal is achieved, chamber pressures in the dryer approach ice’s vapor pressure at the condenser temperature. Condenser temperatures of ~40 °C are frequently used, resulting in absolute pressures around 100 μm Hg. Local chamber pressures in plant-scale freeze-dryers are effected by pressure drops due to high-velocity vapor flows between dryer trays and heating panels (12).

Other Dryers. Foods are also dried in air-lift systems, in fluidized beds, on the surface of rotating, steam-heated drums, by contact with hot air in rotating, hollow cylinders of various types, by passage under or through arrays of high-velocity air jets, and by exposure to the sun in hot dry regions. Sun-dried foods may be treated with SO2 to prevent browning and fumigated to prevent insect infestation.

\( a_w \) Reduction

Microorganism growth can be prevented by sufficiently reducing \( a_w \), the chemical activity of water (13). Thus, Clostridium botulinum will not grow at \( a_w < 0.94 \), most bacteria will not grow at \( a_w < 0.905 \), most yeasts at \( a_w < 0.88 \), and most molds at \( a_w < 0.8 \); even osmiophilic molds and yeasts will not grow at \( a_w < 0.6 \). Drying preserves foods in part by reducing \( a_w \). Moderate \( a_w \) reduction can be used in combination with other steps (e.g., pH reduction, refrigeration or freezing, and use of antimicrobial agents and fungistats to provide storage stability). Water activity values are also reduced by evaporation or by adding solutes.
that reduce the mole fraction of water in solution. Different combinations of salting and drying are used to preserve cod. Lightly salted cod is stable when its moisture content is 12.3%; heavily salted cod is stable with a moisture content of 52.4%.

Chilling
Chilled, nonfrozen storage is widely used to extend the storage life of fresh produce, meats, fish, and dairy products (14). Recommended temperatures and humidities and approximate storage lives for the commercial storage of fruits and vegetables are listed in U.S. government publications (15) and by the American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE) (16). These sources also provide heat capacities and heats of respiration for many of the commodities listed. Typical rates of heat influx are also given to permit computation of refrigeration loads for storage facilities. Storage temperatures close to 0 °C and humidities ranging between 85 and 95% are frequently recommended; cold-sensitive produce is stored at temperatures between 7 and 13 °C. Storage life at these conditions varies widely, ranging from as short as 2 days to as long as 12 months; storage life is less than 1 month for roughly 60% of the foods surveyed. Design and operating considerations for food storage refrigeration systems are treated by ASHRAE (16). Precooling before storage is carried out by blowing chilled air on products, by immersing products in chilled water containing chlorine or an approved phenol compound or spraying similar chilled water on them, or by mixing products with ice or slush. Leafy products that conduct heat poorly are often cooled by evaporating a small part of their water content in a vacuum chamber. To minimize weight loss, such products may be prewetted before vacuum cooling.

Controlled Atmospheres. Chilling is combined with use of controlled atmospheres that contain as little as 2–5% O2 and up to 5% CO2 to increase the storage life of apples, pears, and cabbage (17). CO2-rich controlled atmospheres are also used during long-distance shipment of meat. The rooms or shipping containers used must be sealed and air-tight. Exceeding tolerable CO2 levels or falling below tolerable O2 levels damages apples. These levels significantly differ for different varieties of apples. Respiration can be used to reduce O2 and raise CO2 levels. Any excess CO2 produced is absorbed by hydrated lime, ethanolamine (which can be regenerated by heating), water, or less frequently activated charcoal or molecular sieves. To more rapidly adjust O2 and CO2 levels, catalytic or noncatalytic combustion of fuel and flushing with nitrogen are used. Multiscale O2 and CO2 sensors and loggers have been used to ensure that improper local environments do not develop in controlled-atmosphere storage rooms. The use of selectively permeable wrapping can generate modified atmospheres in packages and improve shelf life during food distribution. Vacuum packing is also used.

Freezing
Freezing, the cooling-induced conversion of most of a food's water content into ice, is used to preserve virtually all types of moist food (18,19). It is also used as a step in freeze-concentration and freeze-drying and to produce ice cream and sherbet. Freezing preserves foods by reducing reactant mobility, reducing reactant mobility, and reducing a.* The water activity of a frozen food depends solely on its temperature and is independent of the nature of the food involved, as long as it is suitably moist: a* = 0.908 at −10 °C, 0.824 at −20 °C, and 0.748 at −30 °C.

Solid foods are frequently frozen by exposing them to suitably cold streams of refrigerated air. This may be done by passing cold air through a shallow, fluidized bed of small particles or by blowing cold air over larger pieces of food carried on perforated conveyors or mounted on trays in racks. Meat carcasses are frozen by passing cold air around them while they hang by the hind legs on hooks. Foods have also been frozen by direct contacting with brine-ice mixtures or by contact with an inert, evaporating refrigerant or cryogenic gas. Packaged foods are often frozen by compressing them between refrigerated plates.

Ice cream and sherbets are partially frozen in scraped-surface freezers. Finish freezing (hardening) is carried out in plate freezers. Ice cream pops are frozen in molds conveyed through baths of refrigerated brine. Descriptions of most types of commercial freezing equipment, their specifications, and operating conditions are provided by Postolski and Gruda (20).

Most water in moist foods will freeze, but part (0.17–0.31 kg water/kg solids, depending on the particular food) is bound to food solutes or solids in nonsolvent form and will not freeze, even at very low temperatures. Food solutes depress food freezing points. Equilibrium initial freezing points for moist foods range between −0.4 and −3.2 °C, and usually lie between −1 and −2 °C. Foods frequently supercool 5 or 6 centigrade degrees below their equilibrium initial freezing point before ice nucleates and freezing starts. Then the temperature rapidly rises to the initial freezing point. As freezing proceeds, solutes concentrate in residual unfrozen solution, further depressing the freezing point. Thus, food freezing occurs over a range of temperatures rather than at a single temperature.

Fractional conversions of freezable water to ice, fractions of water that remain unfrozen, and thermal properties of partly frozen foods can be predicted with the aid of the freezing point depression equation (21). These predicted values can be used in calculating freezing times for foods and temperature profiles for foods undergoing slow to moderately rapid freezing (i.e., freezing in which local phase equilibrium can be assumed without serious error) (22–24). Freezing involves incorporation of water in ice crystals and diffusion of water to active growth surfaces on ice crystals. These processes occur at rates slow enough to cause significant local deviations from phase equilibrium during very fast freezing. They also largely determine ice crystal size and shape (but not the total amount of ice formed) during slower freezing.

Ice usually grows as dendrites (treelike crystals) when foods freeze. Spacings between ice dendrites tend to be inversely proportional to the square root of the initial freezing rate. Specific surface areas and surface free energies of dendrites are much larger than those of rounded crystals.
Therefore ice dendrites tend to convert into rounded ice. Conversion occurs rapidly in agitated slushes but is quite slow in unagitated frozen food, particularly food that is kept well below its initial freezing point. Specific surface energies are larger for small crystals than for large crystals. Therefore small ice crystals tend to sacrificially melt and cause growth of larger ice crystals. If storage temperatures are not low enough or are allowed to cycle excessively, ice crystal size will increase appreciably after long storage, affecting texture adversely.

When unblanched fruits and vegetables are frozen, ice initially nucleates outside cells. If freezing is slow, much water transfers outward from cell interiors and forms extracellular ice before nucleation occurs inside cells. This affects texture adversely. Good practice calls for rapid freezing. Then, intracellular nucleation occurs quickly, most water freezes inside cells, and adverse textural effects are minimized. Preferential growth of ice in extracellular space usually does not occur during slow freezing of blanched vegetables. Fast freezing of meat and fish helps minimize losses of water-holding power and decreases drip following thawing.

**Thermal Processing**

Sterilization, pasteurization, and blanching are processes in which foods are heated and held at suitable temperatures long enough to kill the microorganisms and inactivate the enzymes they contain. Commercial sterilization is designed to kill substantially all microorganisms that can grow or spores that can germinate at storage conditions. Because sterilized foods are packed in sealed containers that contain virtually no air, anaerobic organisms are of greatest concern. Destruction of C. botulinum spores is often used to evaluate adequacy of sterilization for low-acid foods (i.e., foods with pH > 4.5). Fractional survival levels must be less than $1 \times 10^{-12}$ for such spores. Adequacy of sterilization for low-acid foods may also be based on destruction of a hardier putrefactive anaerobe called (PA) 3679. Adequacy of sterilization for foods with pH values less than 4.5 is based on destruction of one or another of the facultative anaerobes, Bacillus coagulans, B. mascerans, or B. polymaxa. Fractional survival levels for these organisms should be less than $1 \times 10^{-5}$ or $1 \times 10^{-6}$. Sterilization of low-acid foods involves maximum temperatures in the 116–149°C range. Maximum temperatures up to 110°C are used for more acidic foods.

Sterilization of canned products is carried out batchwise by steam heating in vertical or horizontal retorts. The retorts must be vented for at least 5 min with steam flow to remove air before steam pressurization and starting a heating cycle. Foods in jars are often sterilized in air-pressure hot water in retorts. Sterilization based on steam heating is also carried out continuously in retorts with sections separated by pressure locks, in rotary sterilizers (where a rotor causes cans to move through the unit along spiral guide rails), and in hydrostatic sterilizers (where entering and leaving cans are conveyed through a long, steam-heated zone confined between vertical water legs that balance the steam pressure). Figure 5 shows an example of a hydrostatic sterilizer.

Headspace is left when food containers are filled. The container walls or tops have elements that flex. These elements and headspace prevent thermal expansion from damaging containers during sterilization. Headspace air and dissolved gases are partly removed from cans, jars, and their contents before sterilization by steam flushing or vacuum evacuation before capping. Thermal exhaustion (i.e., preheating material to expel dissolved gases) is sometimes used instead of or in conjunction with these processes.

Thermally processed products are cooled under air pressure with chlorinated cold water immediately after heating. Air pressure is used to prevent bulging of containers or unseating of caps caused by excess residual pressures inside containers. Cooling is continued until products are just warm enough to fully evaporate water clinging to the container surfaces. Sterilization, container filling, and capping procedures for specific products differ in detail but are described in standard references (25).

**Pasteurization**. Pasteurization, the thermal treatment at temperatures less than 100°C, is used to inactivate enzymes, virtually all pathogenic microbes, and many food spoilage organisms. It is used in combination with protective packaging and other treatments (e.g., chilling, pH, and $a_w$ reduction) that greatly retard subsequent microbial...
growth. The target organism or enzyme on which adequate
inactivation is based depends on the product involved. For
milk it is Coxiella burnetti, which causes Q fever. For wine
or beer it is wild yeasts. For high-acid fruits it is molds and
yeasts. For citrus juices it is pectinesterase. Inactivation
of pectinesterase also provides more than adequate micro-
bial inactivation. Acceptable holding time-temperature
combinations may be specified by government regulations.
In the United States, at least 30 min at 62.8 °C or at least
15 s at 71.7 °C is required for pasteurization of milk. Most
frequently, the latter conditions are used. Other, equally
effective time-temperature combinations can be approved.
Ultrahigh temperature pasteurization carried out at 88.3
°C for 1 s or 100 °C for 0.01 s is used sometimes. Liquid
mixes used to produce ice cream and frozen desserts are
pasteurized for at least 24 s at 79.4 °C.
Milk is pasteurized before it is put in containers. This
is done by regeneratively and directly heating milk to a
specified temperature in a plate heat exchanger, holding it
at that temperature for a required time and then cooling
it in the same exchanger. Integrated systems that also pro-
vide cream removal by centrifugation and homogenization
are often used (see Fig. 6). Other arrangements for pas-
teurizing milk are described by Kessler (26), who also pro-
vides correlations for overall heat transfer coefficients ver-
sus fluid velocities or Reynolds numbers for different
plate-heat exchanger flow arrangements. Simpler plate-
heat exchanger setups are used to pasteurize fruit juices.

Other foods and beverages are pasteurized in contain-
ers. Products in glass jars are conveyed on metal-link belts,
heated gradually, held at temperature, and cooled gradu-
ally in stages to prevent thermal shock to the glass. Long
hot water baths or sprays followed by similar cold water
baths or sprays are used for jarred pickles, canned fruit,
and bottled beer. Heating by passage through steam at at-
mospheric pressure is sometimes used in pasteurizing
products packed in glass.

Blanching. Blanching, heating in water somewhat be-
low its boiling point or in steam at atmospheric pressure, is
used to inactivate enzymes that adversely affect food stor-
ability and quality. Sterilization and pasteurization pro-
cesses usually inactivate enzymes as well as microorgan-
isms; blanching is used primarily before drying and
freezing for vegetables and some fruits. Blanching also
causes expulsion of gases from foods, softens foods, facili-
tates cutting and peeling, sets some colors, and provides
some microbial inactivation. Unfortunately it also leaches
nutrients from foods, which in turn increases the biological
oxygen demand (BOD) of effluent blanch water. Less leach-
ing occurs when steam blanching is used and still less
when very humid hot air is used.

Aseptic Processing. Aseptic processing involves heating
liquids, semiliquid products, or liquids containing small
particles to sterilization or pasteurization temperatures in
heat exchangers or scraped-surface heaters or by injection
or infusion of culinary steam and then passing the prod-
ucts through a holding tube. They are then cooled in heat
exchangers or, for products heated by steam addition, by
vacuum evaporation, and put in sterilized containers. The
containers may be cans, jars, or drums that have been
heated with superheated culinary steam. They are capped
with steam-sterilized covers in chambers filled with super-

---

**Figure 6.** System for standardizing, homogenizing, and pasteurizing milk. Reprinted with permission of Kessler Verlag.
Inactivation Kinetics. Thermally based microbial destruction and enzyme inactivation are treated as first-order reactions; that is, destruction or inactivation rates are proportional to the residual concentrations of living microorganisms or active enzymes involved. Inactivation rates are strongly affected by pH and food composition. If \( k(T_p) \) is the destruction rate constant for a particular microbe or enzyme in a food expressed as a function of product temperature \( T_p \) and \( T_p \) is a function of time \( t \), we can write

\[
\ln \left( \frac{N}{N_0} \right) = - \int_{t=0}^{t} k(T_p) \, dt
\]

where \( N \) is the residual number of microbes per unit volume or enzyme concentration at \( t = 0 \) and the spot where \( T_p \) and \( k(T_p) \) are evaluated and \( N_0 \) is the initial value of \( N \). Figure 7 depicts how \( k(T_p) \), \( T_p \), and \( N/N_0 \) change during thermally based sterilization. To achieve acceptable destruction or inactivation, the highest \( N/N_0 \) at any locale in a product for the most critical microbe or enzyme in question must be equal to or less than a required value. For low-acid foods, \( N/N_0 \) must be less than \( 10^{-12} \) for \( C. \) botulinum spores. For nonmobile canned foods, the greatest \( N/N_0 \) is assumed to occur at the center of a can’s contents. Convective occurs in foods containing free liquid, and the greatest \( N/N_0 \) is assumed to occur at the slowest heating point. This usually lies along a can’s axis and near its bottom, but its locus should be experimentally determined.

In the food industry, adequacy of thermal processing (27) is often evaluated by determining whether inactivation is equal to or greater than that provided by treatment for a specified time at standard temperature, 121.1 °C (250 °F). Temperature-induced changes in thermal inactivation rates are accounted for in terms of \( Z \), the temperature reduction that will cause \( k(T_p) \) to decrease by a factor of 10. A standard \( Z \) of 10°C (18°F) is often used for \( C. \) botulinum. Depending on the food and corresponding \( Z \) involved, acceptable treatment of low-acid foods should provide \( C. \) botulinum destruction equal to that provided by 1.2–2.4 min isothermal exposure at 121.1 °C.

In evaluating adequacy of thermal processing, one must account for changes in product temperature that occur during both heating and cooling. Product temperatures predicted from heat transfer analysis are often used in trial-and-error fashion to determine suitable heating and cooling times. These calculations are useful for estimating required processing time, but determination of adequacy of processing is usually based on measured product temperature-time data. This is particularly true for new products, where adequacy of sterilization must be further verified by microbial inactivation tests in inoculated packs.

Formula methods and nomographs based on characteristics of semilog plots of unaccomplished temperature change versus time were once widely used to determine adequacy of sterilization and needed thermal processing time. Now, simple computer programs are used for the same purpose.

Sterilization causes partial destruction of some nutrients and often adversely affects food flavor and color. Destructions of nutrients and some colors and flavors are first-order or pseudo-first-order processes for which kinetic parameters are sometimes available (28). New flavors, colors, and textures that affect product quality also develop by processes with more complex kinetics. Effects of product temperature on nutrient, flavor, and color destruction are often framed in terms of activation energies and the Arrhenius equation instead of \( Z \). In evaluating nutrient destruction and quality reduction, fractional retention of the nutrient or attribute being evaluated is averaged over the whole volume of the container. Sterilization methods and temperature-time conditions that provide adequate pathogen or spoilage agent inactivation but minimize adverse effects should be used. Rates for reactions that destroy nutrients and adversely affect sensory properties increase less sharply than microbial destruction rates as temperature increases. Therefore high-temperature, short-time (HTST) processing often improves nutrient and flavor retention. Heat-resistant enzymes may not be adequately inactivated when microbial destruction is

**Figure 7.** \( T_p \), \( k(T_p) \), and \( N/N_0 \) versus \( t \) during sterilization of a canned solid food by thermal processing: \( T_s \) and \( T_{cw} \) are, respectively, the steam temperature and the cooling water temperature in the retort, \( T_p \) is the temperature at the center of the food, and \( t_c \) is the time at which heating ends and cooling begins. Reprinted with permission of John Wiley & Sons.
adequate if very high temperatures are used for very short times.

Outer regions of containers are often overprocessed and suffer excessive quality degradation when the points that heat most slowly are just adequately processed. Overprocessing can be reduced by providing more uniform heating (e.g., by using aseptic processing and microwave and ohmic heating). Other preservation processes that avoid or greatly limit heating (e.g., radiation preservation, microfiltration, and high-pressure treatment) have been extensively investigated or, in some cases, are already in use. Their advantages and limitations were examined in a recent review (29).

Concentration Processes

**Evaporators.** Water and solvents can be removed from liquid foods by means of evaporators. In multiple-effect evaporators, the most commonly used type, vapor generated by steam or vapor heating in an effect provides heating that causes evaporation in a subsequent effect operating at a lower pressure. In N-effect evaporators, roughly N kg of vapor is generated for each kilogram of steam used. In some cases, a steam ejector is used to compress part of the vapor generated in a second or third effect, and the compressed vapor and motive steam are used to heat the first effect, providing more than N kg of vapor per kilogram of steam used (30). Vapor recompression may also be carried out by large centrifugal compressors. Electrical energy is used to drive the compressor; virtually no steam is used directly. Evaporation, particularly multiple-effect and vapor recompression evaporation, is more thermally efficient than drying. Therefore evaporators are often used to concentrate liquid foods before drying. Evaporation is also used to concentrate juices, syrups, sauces, fermentation residues, stillage bottoms, and absorption refrigeration solutions. It is also used to induce crystallization and produce glassy supersaturated sugar solutions used to make hard candies.

Vacuum evaporation, often combined with single-pass operation in each effect to provide low product holdup, is used to minimize thermal damage to foods (31). In single-pass evaporators, steam condenses on the outer walls of vertical heat transfer tubes 6–12 m long mounted in a steam chest. Water evaporates from liquid flowing in film form down the inner walls of the tubes. Concentrated liquid and vapor leave through the bottom of the tubes. Entrained liquid is separated from the vapor in a cyclone. A single-pass, multieffect system is depicted in Figure 8. High-velocity recirculation in liquid-filled tubes is used to improve heat transfer in evaporators used to concentrate viscous materials such as tomato paste and gelatin and pectin solutions. Liquid superheats as it passes through tubes in such evaporators and flashes as it leaves the tubes. Cyclones are used to separate entrained liquid from the vapor produced by flashing. Scraped-surface, thin-film evaporators are used for very viscous products (e.g., hard candy melts) and for heat-sensitive products. Natural convection and boiling cause circulation of liquid in evaporators containing short, steam-heated tubes mounted in a wide-diameter assembly (calandria). Vapor separates from the foaming, boiling liquid in a large, vapor-filled chamber above the calandria. Heat transfer coefficients in these evaporators peak at a certain liquid level in the calandria. Multieffect systems containing such evaporators are frequently used to concentrate sugar solutions in sugar refineries. Part of the vapor produced may be withdrawn from effects and used for process heating elsewhere in the refinery.

Most food aromas have high relative volatility with respect to water and almost completely vaporize during initial stages of evaporation. Such aromas can be recovered in concentrated form from first-stage vapor streams and used for add-back purposes.

**Membrane Processes.** Pressure-driven flow through permselective membranes is used to gently remove water or aqueous solutions of low molecular weight solutes from liquid food products. Reverse osmosis (i.e., selective removal of water) is used to raise the solids content of maple sap from roughly 2% to 10–12%, to recover water and solutes from waste processing streams, and to concentrate dilute aqueous caffeine extracts. Ultrafiltration (selective removal of water and low molecular weight solutes) is used to concentrate proteins in milk, milk whey, and soy whey and to recover protein from potato fruit water. Protein-rich concentrates obtained from heated milk by ultrafiltration are used to make cheese with reduced loss of milk solutes and reduced use of rennet and starter cultures. Diafiltration (ultrafiltration accompanied by water addition) is used to de-ash gelatin. Pervaporation has been used experimentally to produce aroma-rich concentrates by selectively evaporating water through permselective membranes. Mass balances for retained species and equations governing permeation rates are used in designing membrane-based processes. Design calculations must not neglect the effects of concentration polarization on permeation rates and selectivity.

Freeze Concentration. Freezing is sometimes used to gently and selectively remove water from fruit juices, beverage extracts, beer, and wine. Scraped-surface coolers partially freeze the solution being processed, producing small ice crystals. Ostwald ripening (i.e., growth of large crystals induced by sacrificial melting of smaller crystals) is then used to produce large ice crystals, which are separated from residual solution in wash columns (32) or centrifuges. Thermal efficiency and upper concentration limits have been increased, and costs reduced, by means of countercurrent freeze concentration (33), but the use of freeze concentration remains limited because of its high cost.

**Solid-Liquid Extraction**

Solid-liquid extraction is used to recover sugar from sugar beets and sugarcane, oil from oilseeds, juice solutes from fruits, beverage extracts from coffee and tea, protein-rich materials from soybeans, fermentable solutes from grains, and hydrocolloids, gums, and natural food colors from various sources. It is also used to remove unwanted constituents (e.g., caffeine from coffee, bitter compounds from olives, cyanogenic compounds from cassava, sarcoplasmic
Effect 1 2 3 4a 4b

Figure 8. Single-pass, falling-film, multieffect evaporator. The fourth and fifth stages in terms of liquid flow together make up the fourth effect in terms of vapor utilization. bc, barometric condenser; e, ejector; cy, cyclone; s, steam; w, cold water; c, condensate; v, vapor.

protein, and undesired flavors from minced fish, and salt from pickles) (34,35). Extraction also occurs when flavoring agents leach out of charred barrels during the aging of wines and distilled spirits, when solutes leach out of vegetables and fruits during blanching and fluming, and when plasticizers leach out of plastic containers into liquid foods and beverages.

Vegetable or animal matter from which solutes are to be extracted is cut or otherwise divided into small particles. Oilseed grits are flaked or extruded as steam-puffed cylindrical particles. This coalesces isolated oil bodies and creates fissures or pores that allow solvent or extract to reach the oil. During extraction, solutes diffuse out of the solid through internal liquid-filled paths and pass into solvent or extract that contracts the solid. Dry materials (e.g., coffee and tea) must imbibe solvent or extract before diffusion can proceed. In rare cases (e.g., extraction of vanilla from vanilla beans with alcohol), imbibition is very slow and is a rate-controlling step. Plasma membranes are denatured to facilitate water-based extraction of solutes from cellular material. Then, pores in cell walls become the main diffusion bottleneck. In-particle diffusivities for water-soluble extractibles are 0.1–0.5 times as large as in water alone. Required extraction times are inversely proportional to the in-solid solute diffusivity and proportional to the particle size squared.

The extract carries away the solute, following a contacting path that promotes high solute recovery. Extraction efficiency and extract concentration depend, in part, on the stripping factor, the solute distribution coefficient multiplied by the ratio of extract to moist solids. High concentration extracts can be obtained by using a low stripping factor (i.e., low solvent flow rates or small amounts of solvent), but efficient extraction can be obtained only if the stripping factor is greater than 1.0. Stripping factors of 1.2–1.4 are often used in water-based extractions. Higher stripping factors (e.g., 2.4) are used for hexane-based extraction of oil from oilseeds.

Small-sized particles favor fast extraction but cause excessive flow pressure drop and displacement instability. Large particles extract slowly. Particle thicknesses or diameters in the range of 3–5 mm are a good compromise. Where possible, particle shapes that prevent blinding of surfaces by interparticle contact are used.

Commercial extraction is often carried out in very large, continuous or semicontinuous countercurrent extractors where extract percolates downward through a rising bed of solids (see Fig. 9) or in units where a bed of solids moves horizontally while extract repeatedly percolates through it in a flow arrangement that effectively provides countercurrent contact. In other cases, a cyclically loaded and discharged set of interconnected columns is used to provide nearly countercurrent contacting. Most of these processes can be analyzed and designed by means of solutions of partial differential equations describing diffusive mass transfer in solids (3,34) or mass transfer analogs of partial differential equations describing conductive heat transfer in solids (2). Solutes are sometimes produced from insoluble precursors by reactions whose kinetics must be accounted for in designing the extraction process. Extraction processes involving foods often are subject to imperfect contacting (e.g., channeling, unsteady displacement, and axial
Figure 9. DDS extractor, a type often used to extract sugar from sugar beets. Reprinted with permission of De Danske Sukkerfabrikker.

dispersion). These imperfections must be accounted for in design calculations.

Supercritical carbon dioxide is now used as a solvent in decaffeinating coffee and tea and in recovering solutes from foods. Supercritical carbon dioxide is a particularly attractive solvent because it is nontoxic and can be readily removed from foods. Further, its dissolving capacity and selectivity for solutes such as caffeine can be adjusted by adjusting temperature and pressure. Figure 10 shows a schematic of a large supercritical extraction system used to decaffeinate green coffee beans.

Solid-Liquid Separations

Filtration. Filtration is used to remove solid particles from or to clarify juices, extracts, vegetable and fish oils, fermented beverages, redistilled, cooking oil, flume water, milk, and soy milk. It is also used to separate potato starch from potato fruit water, high-melting fats from vegetable oils in fractionation processes, crystals from mother liquors, and chemically precipitated impurities from sugar juice.

The engineering principles that govern filtrations involving foods are very similar to those that apply to the filtration of chemicals (36), but superior sanitation must be maintained. A fluid that contains particles is passed through a cloth or other porous medium that retains particles. The particles deposit on the medium, forming a porous cake that thickens as deposition continues. As long as total pressure drop remains constant, total flow resistance increases linearly as the cake thickens. Depending on the feed pumping system used, flow resistance increases cause either decreases in flow rates and solids deposition rates, increased fluid pressure drop, or a combined increase in pressure drop and decrease in flow rate. Filtration is stopped when these changes become excessive, and the filter cake is manually or mechanically removed.

Extents of these changes and optimal scheduling of cake removal can be predicted by means of calculations that relate cake depth and flow pressure drop to the initial solids content of the fluid, the amount of filtrate discharged, and the experimentally determined relationships between applied pressure, cake density, specific filtration resistance, and applied pressure. Some food-based filter cakes are highly compressible, and their specific filtration resistance increases markedly as pressure increases. Filter aids (insert, highly porous solids) are sometimes added to filter feeds to reduce cake filtration resistance and compressibility and to improve solids retention. Cake filtration resistance tends to be inversely proportional to the particle size squared. Therefore chemical agents that cause particles to flocculate (stick together in clumps) are often added to filter feeds.

Deep porous media that capture and retain particles within their pores are used for some feeds. Media flow resistance increases as filtration proceeds. Consequently flow pressure drops increase and/or flow rates decrease. The medium is replaced or cleaned by backwashing when these changes become excessive. Unlike cake filtration, where total filtration resistances at constant pressure increase linearly as solids removal progresses, filtration resistance for in-media solids capture is more than linearly proportional to the amount of solid removed. Therefore in-media capture is used only for fluids with light solids loads.

Sterile microfiltration based on use of membrane filters with 0.2-μm pores is used to remove bacteria from beers, thereby eliminating or reducing needs for pasteurization. Ceramic microfilters that can be sterilized and cleaned by back-flushing are used to clarify fruit juices. They are very expensive but will last for 10 years or more if care is taken to avoid heat shock.

Beds of particles used in solid-liquid extraction and adsorption and ion-exchange processes act like deep filter cakes. Pressure drops during flows through such beds often can be predicted with the aid of the Ergun equations or Kozeny-Carman equation. In other cases, flow pressure drops are calculated by means of empirically determined permeabilities or specific filtration resistances.
Centrifugation. Particles forced to rotate about an axis at high angular velocity in a centrifuge experience a radially acting force proportional to the radial distance from the axis, the angular velocity squared, and the density difference between the particle and surrounding fluid. At rotational speeds used in commercial centrifuges, particles denser than the fluid move outward and particles lighter than the fluid move inward thousands to tens of thousands times faster than they would sink and rise, respectively, under the influence of gravity. Therefore, centrifuges are used to separate and collect dispersed particles and droplets too small to be cleanly and quickly separated by other means. Centrifuges are used, for example, to separate dispersed oil, microorganisms, and fine biological particles and precipitated curds from aqueous solutions (23) and to remove yeasts from fermented beverages and waste treatment sludges from dischargeable water. Density differences driving centrifugal settling are usually much smaller for food processing than for chemical processing; the particles involved are more compressible, and, again higher levels of sanitation are required.

The particles or droplets involved move outward or inward until they strike a collection surface or material previously deposited on that surface. Deposited material then slides along the collection surface or is mechanically conveyed along it to a conduit or port that permits it to leave the centrifuge. Settled solids may leave through peripherally located nozzles as a dense slurry or leave in even denser form through intermittently opened peripherally located slits. Settled solids may also be knifed or augered out of centrifuges. Clarified liquid and separated oil or oil-rich creams leave through ports located closer to the rotational axis of the centrifuge. Centrifuges are sized to provide holdup time sufficient to permit complete settling or a desired level of settling. Large numbers of thin, closely spaced, concentric conical disks are mounted in centrifuges to reduce settling distances and times for extremely fine materials (e.g., butterfat droplets, and microorganisms). This greatly increases usable throughput rates for feeds containing such fine material.

Filtering centrifuges have a finely perforated wall or a perforated wall covered with filter cloth. Centrifugal force causes solids to deposit as a cake on the wall or cloth and then forces liquid through the cake. Wash liquor may also be forced through the cake. Such centrifuges are used to recover crystals (e.g., sugar and citric acid) from concentrated mother liquors, to separate freeze-concentrated vinegar and beverage extracts from ice, and to recover juice from milled fruit.

Centrifugal force generated by rotational flow is produced by tangentially injecting solids-laden liquid or gas into hydroclones or gas cyclones. These units have a cylindrical shell attached to a conical bottom, a central outflow port at their top, and a bottom port through which solids or solids-rich underflows leave. The injected liquid or gas first spirals downward in an outer vortex and then spirals upward in an inner vortex. The vortex flow tends to cause particles to move outward toward the wall. In gas cyclones virtually all the gas leaves through the top port. Gas cyclones are sized to provide enough flow residence time to permit wall impact by all particles greater than a desired size. Gas cyclones are used to separate fine spray-dried material from discharged drying air, pneumatically elutriated materials from carrier gas, chaff from coffee roaster discharge gases, and pneumatically conveyed material from conveying air.

In hydroclones, most entering liquid leaves through the upper port as overflow, but enough leaves through the lower port to provide a flowable slurry, the underflow. The centrifugal force generated by vortex flow is partly counterbalanced by drag force generated by radial flow of liquid from the outer to the inner vortex. Consequently, small particles or low-density particles may migrate inward and be carried away with the overflow. Larger or denser particles leave with the overflow. Large banks of miniature hydroclones used in parallel are employed to separate starches from protein- and solute-rich solutions during the wet milling of corn.

Expression. Pressing of fluid-rich biological matter is used to recover fruit and sugar juice, vegetable and fish oil, pectin released by cooking apple pomace, and protein-rich juice from leftovers. It is also used to expel whey and to fuse curds together during cheese manufacture and to dewater spent food processing wastes, by-products, and processing intermediates (37). Expression is carried out in a wide variety of equipment (38). In screw presses, feed is trans-
ported down the length of a perforated barrel by a rotating single screw or by twin screws. The feed compacts, and fluid is expelled through barrel perforations because the screw’s root diameter increases, its pitch or rotary speed decreases, or the barrel diameter decreases, or because constricted solids outflow from the barrel causes a progressive buildup of pressure.

In roll mills, a grooved, rotating roll is hydraulically pressed against two similar counterrotating rolls that lie beneath it. As the rolls rotate, they drag feed (usually shredded sugarcane) through nips between the top and bottom rolls. This compacts the feed and expels juice. Some of the juice flows back through the advancing feed; some flows laterally into the grooves and then backward. The process is analyzed by relating local backflow and flow pressure drop to the cumulative extent and rate of compaction and the local specific filtration resistance of the compacted feed.

In belt presses, feed is deposited on a moving cloth belt and pressed between that belt and an opposing, moving cloth belt driven over a series of solid or perforated rolls. In some cases the belts follow a serpentine path. Expelled fluid flows through the cloths and is collected in troughs beneath the rolls. Figure 11 shows a serpentine belt press of a type used to press juice out of milled apples. In presses of other types, feeds confined in a perforated cage or in stacked cloth pouches are pressed by a diaphragm or ram. Fluid is also expelled from cakes compacted by centrifugal force in decanter centrifuges.

Engineering treatment of expression is based on numerical solution of partial differential equations that describe how flow in press cakes is driven by fluid pressure gradients or related solid stress gradients and how rates of local cake volume change as a function of flow rate gradients (38, 39). Empirically determined relationships between solid stress, specific filtration resistance, and the solids-based specific volume of the cake are used in solving these equations. Methods have been developed for determining these relationships from constant-pressure or constant-rate pressing tests carried out in suitably instrumented equipment (40).

Some press feeds (e.g., waste treatment sludges and fruit with thin cell walls and weak intercellular bonds) compact excessively as pressure increases. When such feeds are pressed, fluid pressures drop precipitously and solid stress rises sharply over very thin layers near outflow surfaces. Pressing can be improved for such feeds by adding an inert material (e.g., rice hulls) that stiffens the press cake. Other feeds (e.g., spent coffee grounds and oilseeds) have thick cell walls that adhere strongly together. Fluid pressure builds up inside the solid when such feeds are pressed. The extent of in-solids pressure buildup depends on the rate of pressing and the extent of compaction. Pressing of such materials can be improved by treatments that promote cell wall rupture.

**PEELING, SKINNING, AND HULLING**

Peels, skins, fatty layers, hulls, and shells frequently must be removed from foods. Skins are often manually cut from meat and fish with the assistance of special tools. Wholly mechanical systems are also used. These grab a carcass, carcass part, or fish at a selected spot and drag it past suitably positioned knives and saws. Fat-depth-sensing devices are sometimes used to guide cutting.

Moisture contents of shell-encased and hull-encased foods are often adjusted to increase shell and hull friability and separability. Hulls and shells are then cracked by impaction in rapidly rotating pin mills or by passing the food involved through pairs of rollers or plates that turn at different speeds and produce a pressing, rolling action. Thin shell and hull fragments are then pneumatically separated from more compact usable food particles.
Manual peeling of vegetables and fruit is costly; abrasion peeling removes too much peel and is wasteful. Therefore, treatments that break down bonds between cells or decompose cells in or near skins are used. These include short time heating with high pressure steam, short-time immersion in baths of lye, and passing product through a flame. The loosened skin is then brushed off or washed off by water jets, a step that also removes residues of chemical peeling agents. The process can be analyzed in terms of short-term heat or mass transfer in surface layers and the kinetics of proteopetin or cell wall breakdown. To sharply limit intercellular bond breakdown, internal temperature gradients near the product surface must be very large and exposure times very short. Even so, undesirable textural and color changes often occur in regions close to the peeled surface.

**CLEANING**

Washing-based cleaning of food surfaces and cleaning and disinfection of equipment are widely used food processing operations that have been analyzed from an engineering point of view (41–43). Efficient cleaning and rinsing can be achieved only when smooth, highly polished, crevice- and pocket-free equipment is used. Equipment conforming to the sanitary design standards of the Technical Committee of Dairy and Food Industry Supply Association and made of type 304 or type 316 stainless steel is frequently used. Parts may also be made of plastics and synthetic materials approved by the U.S. Food and Drug Administration or Department of Agriculture.

Equipment cleaning frequently involves an initial water rinse, a caustic detergent wash, a second water rinse, and a final water rinse. Conditions that promote corrosion and etching (e.g., use of chlorides) must be avoided. Disinfection is provided by cleaning or rinsing at high temperatures, by steam heating of equipment, or by use of chemical agents (e.g., hypochlorites). Cleaning formerly involved manual hosing down and brushing of disassembled pipelines and equipment and one-time use of cleaning agents. Manual cleaning procedures are still frequently used in meat processing plants and in food processing pilot plants. Spray wands operating at pressures up to 3.5 MPa (500 psig) are used to ensure removal of adherent material.

Clean-in-place (CIP) systems that accomplish thorough cleaning without disassembling piping and permit some reuse of cleaning agents have been developed for many plants. Because cleaning agents and rinse water run through pipelines and narrow-bore equipment in CIP systems, they must be laid out to permit hang-up-free drainage. Product contact surfaces in tanks and large-bore equipment are cleaned with the aid of sprays issuing from permanently installed spraying devices. Globular spray heads or reciprocating sprays are used to provide complete washing and rinsing of internal surfaces. Removal of cleaning agents during rinsing has been successfully modeled so that rinsing can be accomplished without excessive use of water.

Soil deposition and fouling interfere with process heat transfer and greatly reduce permeation rates in membrane-based processes. Therefore fouling and soil deposition mechanisms and kinetics for such processes have been extensively studied. The results of these studies have been used to optimize frequencies of cleaning and to select operating conditions and surface treatments that retard fouling.

**HOMOGENIZATION AND EMULSIFICATION**

Fat in freshly drawn cow's milk is dispersed as butterfat globules that individually contain a fatty core coated with a lipoprotein layer roughly 10 nm thick. Depending on the type of cow involved and how long the animal has been lactating, mean volume/surface diameters of butterfat globules range between 2.5 and 4.5 µm (43). Unless treated, the globules rise, producing a layer of cream on top of milk. To prevent this, and to minimize butterfat clumping, the globules are broken into droplets whose mean volume/surface diameter ranges between 0.22 and 0.5 µm. This is usually done by passing milk through spring-loaded, narrow-clearance valves in high-pressure piston pumps, called homogenizers (26,43). Cavitation produced by rapid increases in velocity in the pump valve and flow-induced shear break up the globules. The mean size of the resulting droplets is inversely proportional to the homogenization pressure used raised to the 0.6 power. Brownian motion would very rapidly cause collision, clumping, and coalescence of incompletely coated droplets. Therefore the lipoprotein coating must redistribute itself rapidly enough to completely coat freshly created droplets before they collide with each other and stick together.

Dispersions of small drops of oil in a continuous aqueous phase are often created by intensive agitation of a mixture of an oil and an aqueous solution. The agitation produces turbulent velocity fluctuations, which in turn cause local shearing intense enough to overcome the interfacial tension forces that tend to hold drops together. Mean drop sizes produced by such agitation are frequently proportional to the interfacial tension raised to the 0.6 power and inversely proportional to agitation power expenditure per unit volume raised to the 0.4 power. Simple dispersions rapidly coalesce and separate when mixing stops. To avoid this, surfactants or agents that coat drops and prevent or greatly retard drop coalescence are added. Dispersions stabilized in this way are called emulsions. Stiff emulsions that act like soft solids (e.g., mayonnaise) contain so much dispersed material that drops butt against each other. A great deal of work has been done to develop gelled low-fat food systems with textural properties resembling those of stiff, oil-rich emulsions.

**SOLIDIFICATION**

Gelling and Coagulation

Fluid foods often are converted into fluid-rich solids with desirable textures by means of gelling and coagulation. Acid produced by bacterial action and bond site production due to the action of the enzyme rennet serve to coagulate milk casein proteins and produce cheese (44–46). Acidification...
cation is used to precipitate and gel food proteins with low isoelectric pH values (e.g., casein, milk whey proteins, and soy proteins solubilized by base addition). Calcium salts precipitate soy proteins and cause them to gel during the production of tofu (47). Heat-induced gelation occurs during the cooking of many protein-rich solutions and slurries. Other protein solutions (e.g., gelatin) gel when cooled and melt when heated. High-methoxy pectin causes acidic sugar-rich fruit juices and slurries to gel when jams and jellies are produced. Fully hydrated (gelatinized) starches are used to produce cold-set soft puddings.

Liquid foods gel because reactions or physical chemical forces bond together sites on adjacent, long-chain food molecules or chains of globular proteins (48). Low extents of bonding produce increases in viscosity. Gelling occurs when bond junction densities become large enough to form coherent three-dimensional polymeric networks. Gelling occurs only if the gelling agent concentration C exceeds a critical value $C_0$. Gelling times are often inversely proportional to $(C - C_0)^n$, where $n$ is a constant. Simple kinetic considerations indicate that $n$ should equal 1, but it is frequently larger than 1. Gelling and coagulation rates may be controlled by the rate of bond site production or by rates of bond forming reactions or events. Cross-linking continues after gel forms, and the gels progressively stiffen. This reduces the mobility of network sections, slowing and eventually stopping the bonding process. At gelling agent concentrations substantially greater than $C_0$, gel strength tends to increase roughly linearly as $(C/C_0)^2$ increases.

**Cooling-Induced Solidification**

Hard sugar candies are made by evaporating water from a concentrated solution of mixed sugars until the solution is essentially converted into a hot melt containing less than 1% moisture. Then the melt is cooled. So viscous that crystal growth does not occur, the melt ultimately becomes a vitreous glass. Flavoring and coloring agents are blended into the melt. When the hardening melt becomes stiff enough, it is processed in devices that convert it into pieces of appropriate size and shape. Chewy candies are made by cooling, flavoring, coloring, and shaping evaporatively concentrated mixtures of sugars, corn syrup, fat, and milk solids containing 12–15% moisture. In both cases, solidification is effectively caused by large cooling-induced increases in viscosity and flow yield strength. Cooling is controlled to prevent excess stress from developing and to permit shaping (e.g., molding, rolling, and extrusion) before the completion of hardening.

Chocolate is made by molding and cooling finely milled, thoroughly worked mixtures of molten cocoa butter, cocoa, sugar, and in some cases milk solids. Cocoa butter can exist in four polymorphic forms that solidify or change form at different temperatures. Only one, the $\beta$-form, is stable at room temperature. Therefore carefully controlled cooling and seeding with fine $\beta$-form crystals is used in the solidification of chocolate.

**SHAPING AND TEXTURIZING OPERATIONS**

**Rolling and Flaking**

Food processing intermediates used to produce bread, noodles, cookies, and chewing gum are often rolled into sheets before being further shaped by cutting and other forming operations. Cooked cereal grits are rolled to form flakes that are dried and toasted to make ready-to-eat breakfast cereals. Oilseed grits are rolled into flakes to prepare them for extraction. Ropes of partially hardened candy melts are pulled through sequences of pairs of rollers turning at progressively faster speeds to produce narrow-diameter ropes, which are then cut into pieces that are pressed into shape by other devices. Engineering aspects of these operations can be dealt with by methods developed to analyze the calendering of plastics (48). These methods involve the use of partial differential equations to describe flow fields and pressure gradients that develop in plastic materials passing between rolls. These equations can be solved to determine the pressing force and torque acting on the rolls and the power required to drive them. Dimensional analysis has been applied to circumvent complications due to the complex rheological characteristics of food doughs (49).

**Extrusion**

Foods are frequently shaped, mixed, cooked, and texturized in screw-driven extruders (50). Flow driven by shear between the rotating screw and the stationary barrel conveys, mixes, and kneads the material being processed and forces it through dies that shape the product. Rotating or oscillating knives cut the extruded material into pieces of suitable size. Action of the screws on the processed material can generate a great deal of frictional heat, particularly when screws with shallow flights are used and the feed's moisture content is low. Supplemental heating is often provided by heated jackets. Cooling jackets are sometimes used. Initially, single-screw extruders were used; now, extruders containing corotating twin screws that contain changeable sequences of conveying, kneading, and pressurizing elements (Fig. 12) are frequently used. Flow, mixing, and power expenditure in extruders of both types have been treated by methods developed for extrusion processing of polymers (49,51). Foods processed in extruders usually contain 15–31% water so that they can be made plastic when worked on by the screw.

Extruders operating close to room temperature and containing screws with deep flights are used to form pasta by forcing semolina-based dough through large arrays of parallel holes. Die hole inserts that provide nonuniform clearance are used in the production of elbow macaroni. High-temperature extrusion, based on use of screws with shallow flights or elements that provide intense kneading, is used to produce confectionery doughs and puffed and textured foods. Temperatures between 130 and 180°C are frequently used in making these products. Extrusion is used to convert soy protein into products that resemble meat in structure and texture. Such texturization appears to involve the formation of links between $\epsilon$-amino groups and $\epsilon$-amino acids on neighboring protein chains. Kinetic models for the reaction indicate that its rate depends on shear rates in the extruder and die as well as the concentrations of the reacting species.

Concentrated purees or slurries containing components capable of entering into setting reactions are pumped or
extruded through dies into baths or vapor-rich atmospheres containing setting agents. Examples include the use of extruded, slurried, partially dissolved collagen or viscose solutions to form sausage casings and use of pumped slurries containing alginates and pureed fruits or vegetables to form artificially shaped fruit and vegetable pieces. Pump- or piston-based extrusion is used to form products that are subsequently heat set (e.g., doughnuts, masa-based puffed products, and certain types of cookie dough). Extrusion is also used to stuff meat emulsions into sausage casings. Coextrusion is used to form fruit-filled bars.

Moist feeds used to produce shaped cereal pieces and pelletized animal feeds and pet food are pushed by rollers through die holes in thick, rotating, perforated rings. Extrudate emerging from the holes is cut into short pieces by knives.

**OTHER OPERATIONS**

Because of space limitations, many food processing operations have not been covered in this article. These include mixing; grinding, and milling; cutting, slicing, and dicing; screening; pneumatic separation and classification; pumping; mechanical, pneumatic, and hydraulic conveying; slaughtering and carcass disassembly; casting and molding; distillation and deodorization; liquid-liquid extraction; gas absorption; crystallization; adsorption and ion exchange; chemical modification of food ingredients (e.g., hydrogenation of vegetable oils); various types of cooking and baking; treatment of foods with ionizing radiation; pore generation and puffing; coating and enrobing; sorting and inspection; packaging and filling; fermentation, including brewing, wine making, and the production of pickled vegetables and fermented dairy products; aquaculture; shellfish depuration; automated egg production, and automated growing of plants and plant tissue in controlled environments.

**BIBLIOGRAPHY**


See also Enzymes, Baking, Bread Making; Enzymes, Fruit Juice Processing; Freeze Drying, Pharmaceuticals; Production of L-Amino Acids by Aminocylase; Solid Substrate Fermentations, Enzyme Production, Food Enrichment; Yeast, Baker’s.
INTRODUCTION

For the last twenty years, many biotechnology-derived pharmaceuticals have been manufactured for the treatment of diseases such as diabetes, cardiovascular diseases, cancers, dwarfism, anemia, AIDS, cystic fibrosis, chronic granulomatous disease, and kidney diseases. Some of these recombinant protein drugs have already been approved by the U.S. Food and Drug Administration (FDA) for marketing, others are still being tested in clinics or are under review for approval. The success of these protein pharmaceutical drugs greatly depends on the delivery of their biologically active forms to the target site. Therefore, the development of a stable formulation, in which the protein can maintain its native conformation and bioactivity, is one of the important steps in the production of protein pharmaceuticals. In fact, the FDA usually requires the pharmaceutical manufacturers to provide real-time stability data to demonstrate that a pharmaceutical drug contains at least 90% potency throughout the entire shelf life period (usually ≥ 2 years) before it can be approved to be a marketed product. In addition, the manufacturers are required to demonstrate that the degradation products that may cause a loss in the potency of the drug do not have any adverse effects on the safety and efficacy of the drug.

In developing a protein drug formulation and designing a delivery system, formulation scientists must consider the physicochemical properties, clinical indication, site of action, pharmacokinetics, and toxicity of the drug. The physicochemical properties of proteins such as the amino acid composition, molecular weight, isoelectric point (pI), glycosylation, and conformation can affect the pharmacokinetics and toxicity as well as the clinical indication. The potential clinical application of a protein drug depends upon the biological function and potency of the product, and the physical and chemical properties of a protein determine its biological function and potency.

The unique physical and chemical properties of each protein also determine its in vitro and in vivo stability. To obtain in vivo information on the pharmacokinetics and toxicity of a drug, the drug must be administered with a stable formulation. In considering the best formulation for the intended applications of a protein drug, the formulation scientist must also consider the route of administration. For initial animal testing, protein drugs are usually administered systemically via an intravenous (i.v.) injection. However, some indications may require a high local drug dose that cannot be achieved by i.v. administration due to toxicity issues. In this case, an alternative route of delivery is necessary. For example, Pulmozyme® (rhDNase) is delivered to patients with cystic fibrosis via the pulmonary route of administration (1). The drug is best administered directly into the lungs of the patients as an aerosol using nebulizers in order to achieve an efficacious dose at the target site (lungs). Thus, the development of both a stable formulation and suitable delivery route or system is critical in determining the success of a final product.

During the past 10 years, several review articles and texts related to protein formulation and stability have been published (2–7). The purpose of this article is to provide a comprehensive overview of formulation and delivery aspects of protein drugs. Protein degradation pathways, analytical characterization of proteins, the process of formulation development, and drug delivery systems are reviewed.
idation, and oxidation are the three most common ones, others include isomerization, cleavage, thiol disulfide exchange, and β-elimination. To develop a stable protein formulation, protein degradation in the formulation must be inhibited or minimized throughout the shelf life of the product. Fortunately, with the advances in analytical techniques for protein analysis, most of these expected degradation routes can be identified, and the degradation products can also be characterized by formulation scientists. In addition, with the continued success of biotechnology, more recombinant proteins have become available for studying degradation mechanisms, thus increasing the general knowledge in formulating new proteins.

Aggregation

Protein degradation via aggregation pathways is considered a physical reaction that does not involve the breakage or formation of covalent bonds. For a protein to retain its biological function and stability, it must maintain its conformational or tertiary structure. Protein conformation is stabilized mainly by hydrophobic interactions (8–11). Thus, monomeric globular proteins often exist in native or folded conformations such that the hydrophobic groups are not exposed on the surface (12). Loss in the tertiary globular structure or unfolding of proteins due to denaturation results in exposure of hydrophobic residues to the environment. Protein aggregation or denaturation in a formulation (liquid or lyophilized) can be caused by a change in temperature, extreme pH, ionic strength, pressure, or presence of denaturants (11,13). Usually, the native, N, and unfolded, U, states of a protein are in equilibrium during denaturation and folding in a two-state process, as shown in equation 1. The reversibility of the unfolding process depends upon the formulation conditions.

\[ N \rightleftharpoons U \]  
\[ N \rightleftharpoons I \rightleftharpoons U \]  
\[ A \]

Sometimes, a protein denatures via a different pathway that involves the formation of at least one stable, partially unfolded intermediates. As shown in equation 2, the native protein unfolds to form an intermediate, I, that has the internal hydrophobic residues exposed to the environment. This hydrophobic intermediate on the refolding pathway may then form aggregates, A, due to intermolecular hydrophobic interactions. According to this scheme, the protein must be denatured or unfolded before nonnative aggregates can be formed. The formation of a stable intermediate during denaturation and folding has been observed for many proteins (14–18). For example, the dimeric native form of interferon-γ (IFN-γ), unfolds to form a partially denatured monomeric intermediate, below pH 4.5 and in the absence of NaCl (19,20). Upon dialysis, both native IFN-γ and aggregates were obtained. The aggregates lost most of the native tertiary structure as well as bioactivity.

Deamidation

Deamidation of protein is the acid- and base-catalyzed hydrolysis of the side-chain amide on glutamine (Gln) and asparagine (Asn) residues to form a carboxylic acid (Scheme 1). Studies of deamidation in peptides by Robinson and his coworkers suggested that Asn is more susceptible to deamidation than Gln and degrades more readily in the presence of an adjacent glycine (Gly) residue on the C-terminal side of the sequence (21,22). It was also discovered that the deamidation of Asn-Gly was accelerated at pH ≥ 7.0. In addition, the studies found that the amino acid sequence plays a major role in determining the rate of deamidation. For example, the presence of polar amino acid residues adjacent to an Asn or Gln enhances the rate of deamidation, whereas adjacent bulky hydrophobic residues decrease the deamidation rate. Serine (Ser) or threonine (Thr) next to an Asn or Gln can act as a general acid by providing a proton in the acid-catalyzed reaction, thereby enhancing the rate of deamidation. The major factors influencing the rate of protein deamidation include pH, temperature, ionic strength, and buffer species. Degradation of proteins via deamidation has been shown to exist in recombinant protein pharmaceuticals such as insulin (23,24) and human growth hormone (25,26) in liquid formulations.

Isomerization

Deamidation of an asparagine (Asn) residue to yield aspartate (Asp) can also lead to the formation of an isoaspartate (iso-Asp) via intermediate succinimide formation (27). Nucleophilic attack of the main-chain amide nitrogen on the carbonyl carbon of the Asn side-chain amide group results in a five-membered succinimide ring. This cyclic imide can be hydrolyzed to form either aspartate or isoaspartate (Scheme 2). An aspartate residue can also undergo deamidation via intermediate succinimide formation to yield isoaspartate. In this case, nucleophilic attack of the main-chain amide nitrogen on the carbonyl carbon of the Asp side-chain carboxylic group results in the formation of a cyclic imide (Scheme 3). The additional CH₂ group in the glutamine (Gln) side chain makes the formation of a cyclic imide intermediate by nucleophilic attack less favorable than for an Asn or Asp residue. Therefore, the deamidation rate of Gln via isomerization is very slow. When a Gly or Ser residue is next to an Asn or Asp residue, the isomerization via cyclic imide formation is accelerated in solution (28). An increase in temperature or pH also increases the rate of isomerization for peptides containing an Asn-Gly or Asp-Gly sequence (29). Cyclic imide formation was demonstrated during the storage of lyophilized recombinant methionyl human growth hormone (met-rhGH) at 45 °C for 4 months (30), and the Asp-130 residue in the protein underwent isomerization (31).

Cleavage

Cleavage of peptides at an asparagine residue can result from the nucleophilic attack of the side-chain amide nitrogen on the main-chain peptide carbonyl to form a C-terminal succinimide (Scheme 4). The reaction is spontaneous and dependent on protein sequence. Asparagine
having an adjacent Gly or Ser residue on the C-terminal side is more labile to spontaneous cleavage. Cleavage at an Asn-Ser linkage has been demonstrated in proteins such as bovine and porcine somatotropins (32).

### Oxidation

Oxidation is one of the major degradation pathways for protein pharmaceuticals. Amino acids that can undergo
oxides (42).

Oxidation of methionine has been demonstrated in many recombinant proteins such as interleukin 2 (33), relaxin (34), parathyroid hormone (35), and human growth hormone (36). At low pH, the thioether group of methionine is not protonated and is susceptible to oxidation, resulting in methione sulfides. Under extremely oxidative conditions, the sulfoxides can be further oxidized to sulfones. Methionine can react with a variety of oxygen-reactive species such as hydrogen peroxide, alkylhydroperoxides, molecular oxygen, and singlet oxygen generated by heat and light. Methionine oxidation can be catalyzed by the presence of transition metal ions such as Cu2⁺, Fe2⁺ or Fe3⁺. Although there are many oxidative pathways for methionine reported in literature (3,7,37,38), oxidation via a free radical, singlet oxygen, or nucleophilic substitution are the three common mechanisms. In contrast to methionine, oxidation of cysteine usually occurs at higher pH where the thiol is deprotonated (39). Cysteine can react with molecular oxygen to form cystine disulfide. The reaction can also be catalyzed by transition metal ions, especially Cu²⁺. In some proteins cysteine can be oxidized to give sulfenic, sulfenic, or sulfenic acid (40). Histidine, tryptophan, and tyrosine are susceptible to photooxidation via the singlet oxygen pathway. The rate of photooxidation is pH dependent. At neutral pH, histidine and tryptophan photooxidize faster than tyrosine (41). At low pHs, tryptophan and methionine are the most photoreactive amino acids (42).

Oxidation of methionine:

\[ RSCH_3 \rightarrow RS(O)CH_3 \rightarrow RS(OO)CH_3 \]

Oxidation of cysteine by oxygen:

\[ RS^- + O_2 \rightarrow RS^- + O_2^- \]
\[ RSH + O_2^- \rightarrow RS^- + HOO^- \]
\[ RS^- + RS^- \rightarrow RS-SR \]

Thiol Disulfide Exchange

Thiol disulfide exchange occurs when a disulfide bond is reduced to two cysteines, and one of them reacts with another cysteine to form a new disulfide (5). The incorrect linkage of two cysteines in a disulfide bond can induce loss of biological activity of the protein. For example, interleukin 2 contains three cysteines at positions 58, 105, and 125. The native protein form has a disulfide linkage between the two cysteines at 58 and 105. In the presence of copper ions, two less active isomers with disulfide linkages between the cysteines at 58 and 125 and the cysteines at 105 and 125 are formed (5). The reaction is also base catalyzed and concentration dependent. The reaction rate can be influenced by pH, temperature, buffer composition, and the presence of other metal ions.

\[ \beta\text{-Elimination} \]

Inactivation of proteins in an alkaline solution can result from \( \beta \)-elimination of the cysteine residue, with the heterolytic deavage of the disulfide and the formation of dehydroalanine and persulfide (equation 3). The persulfide can further react with hydroxide to form hydrosulfide (HS), as shown in equation 4 (5). The rate of \( \beta \)-elimination increases with increase in temperature and pH. The presence of metal ions also enhances the reaction rate.

\[ R'CH_2SSCH_2R^- + OH^- \rightarrow R'CH_2SS^- \]
\[ + CH_2 = CR^+ + H_2O \]
\[ R'CH_2SS^- + OH^- \rightarrow R'CH_2SO^- + HS^- \]

PROTEIN FORMULATION DEVELOPMENT

After gathering all the background information (physicochemical properties, application, site of action, etc.) on the protein to be formulated, the first step that a formulation scientist usually takes is to characterize the protein in the initial formulation by analytical methods. During this step, analytical methods are also developed or optimized for use in formulation and stability studies. Because protein degradation is often dependent on pH and temperature, the second step of formulation development is to determine the relationship between the major degradation pathways of the protein and these parameters. Formulation scientists usually perform short-term studies of pH and elevated temperature (e.g., 25-40 °C) on initial liquid formulations. After determining the major degradation pathways in these formulations, the formulation scientist will select the most stable formulation that can achieve an optimum balance between the different degradation pathways. Finally, long-term stability testing is performed on the final chosen formulation to obtain real-time stability data. This section reviews the process of developing a stable formulation for protein pharmaceuticals. A simplified process diagram for protein formulation development is also illustrated in Figure 1.

### Analytical Methods for Assessment of Protein Formulations

The selection of formulations for protein pharmaceuticals greatly depends on the stability results obtained in the formulation screening. Therefore, it is important to have suitable and sensitive analytical methods to study the degradation of proteins in formulations. Review articles by Jones provide detailed principles and applications of many analytical methods that can be used for protein characterization (43,44). In the following section, the commonly used analytical methods for determining protein formulation stability during formulation development are discussed. A summary of the analytical methods used for identifying protein degradation is shown in Table 1.

#### High Performance Liquid Chromatography

High performance liquid chromatography (HPLC) methods are widely used in the pharmaceutical industry for the analysis of peptides and proteins. The common HPLC methods for use in formulation development to assess protein stability are size exclusion chromatography (SEC), ion exchange chromatography (IEC), reversed-phase chromatography (RPHPLC), and multidimensional chromatography (MDC) methods.

### Table 1. Analytical Methods for Protein Formulations

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>IEC</td>
<td>Ion exchange chromatography</td>
</tr>
<tr>
<td>RPHPLC</td>
<td>Reversed-phase high performance liquid chromatography</td>
</tr>
<tr>
<td>MDC</td>
<td>Multidimensional chromatography</td>
</tr>
</tbody>
</table>

The selection of analytical methods depends on the protein to be formulated, the type of formulation, and the specific questions being addressed. For example, SEC analysis is often used to assess the size and molecular weight distribution of recombinant proteins, while IEC analysis is commonly used to determine the charge and isoelectric point of proteins. RPHPLC methods are often used to characterize the purity and homogeneity of proteins, while MDC methods can provide additional information on protein structure and interactions.
**Figure 1.** Simplified process diagram for protein formulation development.

Protein drug candidate

Background information
- Physicochemical properties
  - pI, molecular weight, and composition, etc.
- In vivo information:
  - application, site of action, toxicity, and pharmacokinetics, etc.

Protein characterization by analytical methods

Analytical method development

Liquid formulation screening
- pH stability studies
- Elevated temperature studies

Determinant of degradation pathways

Formulation selection

Stress stability testing
- Shaking
- Freeze-thawing
- Light exposure

Stable in solution
- Shelf life > 2 years

Unstable in solution
- Shelf life < 2 years

Lyophilized formulation screening

Lyoprotectants/cryoprotectants selection

Lyophilization cycle development

Accelerated stability studies

Liquid formulation

---

RPC, and hydrophobic interaction chromatography (HIC). Degradation of proteins via aggregation and cleavage results in the formation of soluble and insoluble aggregates (oligomers) and fragments, respectively. These soluble aggregates and fragments have different apparent molecular weights and can be separated from the intact protein on a size exclusion column that contains porous particles with different pore diameters and has been calibrated with molecular weight standards, usually globular proteins of known molecular weight.

IEC is often employed during formulation and stability studies for the separation of degraded proteins with charge heterogeneity resulting from deamidation or isomerization. Charged protein tends to bind to ion-exchanged resin (e.g., sulfopropyl, DEAE) of opposite charge by ionic interaction. Bound proteins can be eluted at their own critical salt concentration, allowing separation of the various charge forms (44). The amount of salt required to elute a charged species depends on its affinity for the resin (e.g., number of binding sites), the net charge, and the ability of the salt to displace the bound protein. By this method, deamidated forms of a protein that exhibit different net charge from each other and the native protein can be separated and quantified.

RPC methods are commonly used for the separation of small molecules and peptides, especially for peptides that result from proteolytic digestion of proteins in peptide mapping (45). A hydrophobic surface on the protein binds to the hydrophobic site of the resin (e.g., alkyl groups such as butyl [C4], hexyl [C6], or octadecyl [C18] or aromatics such as phenyl) in a reversed-phase column by hydrophobic interaction. Such hydrophobic interactions can be weakened by increasing the content of organic modifier in the mobile phase (46). Thus, proteins can be eluted in the order of their hydrophobic interaction strengths by an organic modifier such as acetonitrile. Although the application of RPC for analysis of recombinant proteins is less common than for peptides due to difficulties in resolution of large molecules, it has been used for characterization or separation, as in the case of met-rhGH and rhGH, which differ by the only one amino acid (47). This method is also useful in detecting the formation of more polar degraded proteins resulting from oxidation.

Similar to RPC, HIC is another useful technique for the separation of proteins based on their differences in surface hydrophobicity. Most proteins have some hydrophobic groups exposed to the surface. When the hydrophobic surface of a protein is in contact with the hydrophobic ligands on the beads of a column, water is released from the hydrophobic regions, and the entropy of the system is increased. The release of water increases the entropy of the
Electrophoresis. Electrophoresis is a common method for the separation of proteins based on their net charge and size. Upon applying an electric field to a protein, it migrates toward the anode or cathode, depending on its net charge. There are two basic forms of electrophoresis: sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and isoelectric focusing (IEF). In SDS-PAGE, proteins are complexed with the denaturing agent SDS to give similar mass-to-charge ratios so that they can have free electrophoretic mobility. When SDS-complexed proteins are electrophoresed in a polyacrylamide gel with the appropriate pore size, their migration rate will be based on the size of the proteins. Thus, a mixture of proteins with different molecular weights can be separated using this method.

Separation of proteins by IEF is based on their isoelectric points (pIs), where the mobility of each protein in an electric field is zero and the net charge of the each protein is also zero. Because the mobility of non-denatured protein is dependent on its net charge, and this charge is determined by the pH of the solution, an electric field in the presence of a pH gradient causes the protein to migrate until the pH is the same as the pI (49). This technique is usually employed by preparing a polyacrylamide gel containing ampholytes of appropriate pH and buffer capacity to produce a pH gradient that is stable in the applied electric field. This method has been used to assess protein deamidation during stability studies of many pharmaceutical products such as recombinant human growth hormone (50).

**Capillary Electrophoresis.** Due to recent advances in technology, capillary electrophoresis (CE) is being more widely used in the biotechnology industry for protein analysis. The sensitivity and speed of CE make it a useful technique in early protein formulation development when only a small amount of material is available. CE provides a wide range of separation modes that can be performed with a single instrument. The basic separation principles are the same as with conventional electrophoresis techniques.

Capillary zone electrophoresis (CZE) is the most common CE separation method, and it employs a single buffer system in free solution. Separation of proteins using CE is based on their differences in mass-to-charge ratio. Separation takes place in free solution inside a capillary (usually made of fused silica) filled with electrolyte. Proteins with different mass-to-charge ratios are separated into individual zones as electrical force drives them at different rates through the capillary. As the mass-to-charge ratio increases, the electrophoretic mobility decreases, resulting in a longer migration time. Coated capillaries and buffer additives are often used to prevent protein adsorption and electroendosmotic flow. CZE has been used for the detection of the heterogeneity of recombinant proteins (51) as well as the separation of the deamidation products of growth-hormone-releasing factor (52).

Capillary isoelectric focusing (cIEF) is another separation mode in CE for the separation of proteins based on their isoelectric points. There are three basic steps in cIEF: sample injection, focusing, and mobilization. Proteins are premixed with ampholytes and injected into the capillary by pressure. High voltage is applied to the capillary during focusing so that sample ions can migrate to their neutral charge point in the pH gradient formed by the ampholytes along the capillary. After focusing, the focused protein bands are moved to the detection point to generate a signal. The on-line UV detection of cIEF provides a more accurate quantitation than the slab gel technique. A coated capillary is recommended for cIEF because the proteins stay in the capillary for focusing and mobilization. cIEF has been applied to the characterization of charge heterogeneity of recombinant monoclonal antibodies as a result of deamidation (53,54). The high resolving power of the method also makes it useful for the characterization of the

---

**Table 1. Analytical Methods Used for Analysis of Protein Degradation During Formulation Development**

<table>
<thead>
<tr>
<th>Analytical method</th>
<th>Protein degradation/alteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>High performance liquid chromatography</td>
<td>Protein degradation/alteration</td>
</tr>
<tr>
<td>Size exclusion</td>
<td>Aggregation, polypeptide cleavage</td>
</tr>
<tr>
<td>Ion exchange</td>
<td>Charge alteration (e.g., deamidation)</td>
</tr>
<tr>
<td>Reversed phase</td>
<td>Oxidation, disulfide alterations, charge alteration (using peptide map)</td>
</tr>
<tr>
<td>Hydrophobic interaction</td>
<td>Neutral alteration (e.g., Met oxidation)</td>
</tr>
</tbody>
</table>

**Spectroscopy**

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UV absorption</td>
<td>Tertiary structure alteration</td>
</tr>
<tr>
<td>Circular dichroism</td>
<td>Secondary structure alteration (far-UV CD)</td>
</tr>
<tr>
<td>Light scattering</td>
<td>Aggregation</td>
</tr>
</tbody>
</table>

**Capillary electrophoresis**

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capillary zone electrophoresis</td>
<td>Charge alteration (e.g., deamidation)</td>
</tr>
<tr>
<td>Capillary isoelectric focusing</td>
<td>Charge alteration (e.g., deamidation)</td>
</tr>
<tr>
<td>SDS-dynamic sieving</td>
<td>Aggregation, polypeptide cleavage, disulfide alterations</td>
</tr>
<tr>
<td>Capillary electrophoresis</td>
<td>Neutral alteration (e.g., Met oxidation)</td>
</tr>
</tbody>
</table>

**Capillary zone electrophoresis** (CZE) is the most common CE separation method, and it employs a single buffer system in free solution. Separation of proteins using CE is based on their differences in mass-to-charge ratio. Separation takes place in free solution inside a capillary (usually made of fused silica) filled with electrolyte. Proteins with different mass-to-charge ratios are separated into individual zones as electrical force drives them at different rates through the capillary. As the mass-to-charge ratio increases, the electrophoretic mobility decreases, resulting in a longer migration time. Coated capillaries and buffer additives are often used to prevent protein adsorption and electroendosmotic flow. CZE has been used for the detection of the heterogeneity of recombinant proteins (51) as well as the separation of the deamidation products of growth-hormone-releasing factor (52).
heterogeneity of the glycoforms of glycoproteins such as recombinant tissue plasminogen activator (55).

SDS-dynamic sieving capillary electrophoresis (SDS-DSCe) is a useful separation mode in CE for the molecular weight determination of proteins. Because all proteins have the same mass-to-charge ratio after complexation with SDS, the migration rate of proteins depends only on their molecular size. Although SDS-DSCe is equivalent to the conventional SDS-PAGE, it provides a rapid and automated system for quantitative analysis of proteins based on their molecular weight. Monoclonal antibodies degraded by the formation of aggregates and fragments can easily be separated by this method (54,56,57).

**Spectroscopy.** Ultraviolet (UV) absorption spectroscopic measurement is the most convenient and accurate method for the determination of protein concentration. Proteins with aromatic amino acids have a maximum in UV absorption between 275 and 282 nm at pH values below 8 (58). Thus, the protein concentration in a sample can be determined by measuring its absorbance at the wavelength maximum near 280 nm, and this absorbance is then compared with the absorbance of a solution with a known protein concentration or, if known, the molar extinction coefficient or absorptivity of the protein may be used. The absorptivity (often incorrectly described as the extinction coefficient) of a protein is defined as the absorbance of a 1 mg/mL protein solution through a 1-cm path at the wavelength maximum (e.g., near 280 nm). The molar extinction coefficient is the absorptivity of a 1.0 M solution of the protein through a 1.0-cm path. It is important to measure the protein concentration accurately for extinction coefficient determination. Quantitative amino acid analysis, nitrogen determination, or dry weight measurement are often used for this purpose.

Circular dichroism (CD) spectroscopy is a useful technique to evaluate the secondary and tertiary structures of protein (59,60). The secondary structure of a protein such as a-helices, b-sheets, and random coils can give rise to CD signals in the far-UV region (170–250 nm). CD spectroscopy is commonly employed for the study of denaturation or unfolding of proteins. The tertiary structure of a protein affects the local environment of the aromatic amino acids such as tryptophan, tyrosine, and phenylalanine, as well as the disulfide bonds, and these effects contribute to CD signals in the near-UV region (240–320 nm). Because CD signals can be either positive or negative, it is not easy to obtain detail from CD spectra. However, a change in the far- or near-UV CD spectrum can represent a change in the secondary or tertiary structure of the protein, respectively. Therefore, CD is a useful tool to measure protein denaturation as well as aggregation, both of which result in structural or conformational changes.

Light scattering spectroscopy can be used to detect protein aggregation due to the fact that larger particles scatter more light per weight unit than small ones. Aggregated proteins can be either soluble or insoluble. Insoluble aggregates usually cause opalescence (turbidity) in a protein liquid formulation, yielding light scattering at wavelengths even in the visible region of the spectrum. Turbidity can also be conveniently measured in the near UV (340–360 nm) if no other chromophores are present in the same region. Thus, an increase in absorbance in this region often indicates the formation of aggregates. Correction for light scattering errors is necessary for the measurement of protein concentration made by the absorbance at the peak wavelength near 280 nm (61). Pharmaceutical final product solutions are required by the FDA to be inspected for clarity and degree of opalescence. Therefore, light-scattering determination may be substituted for the visual inspection method. Because native protein in solution also scatters light, light-scattering spectroscopy can be used to estimate the molecular weight of a protein. Dynamic laser light scattering, also known as photon correlation spectroscopy (PCS) or quasi-elastic light scattering (QELS), can be used to calculate the average hydrodynamic diameter and determine the diffusion coefficient of a protein by autocorrelation methods (62,63).

**Formulation Screening Methods**

The methods commonly used by formulation scientists for protein formulation screening to select the best candidate formulation include the short-term accelerated pH and temperature stability studies. Accelerated conditions are defined as the extremes of the parameter range (e.g., 40 °C and high (≥9.0) or low (≤3.0) pH).

Accelerated pH Stability Studies. Solution pH often plays an important role in the stability of the protein product because most of the major protein degradation pathways are pH dependent. Thus, the stability of a protein in different formulations can be compared by performing stability studies on the protein formulated at a range of pH using a variety of buffer solutions. Examples of buffers that are used to maintain the pH of formulations in commercially available recombinant protein drug products are listed in Table 2. Acetate, succinate, citrate, and phosphate are the most common buffers used in recombinant protein formulations. For rapid formulation screening, a pH stability study is usually performed in a short period of time and at an elevated temperature of 40 °C because the rate

**Table 2. Examples of Buffers Commonly Used in Recombinant Protein Formulations for pH Maintenance**

<table>
<thead>
<tr>
<th>Buffers</th>
<th>pH</th>
<th>Commercial products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium acetate</td>
<td>4.0</td>
<td>Neupogen® (Amgen)</td>
</tr>
<tr>
<td>Sodium succinate</td>
<td>5.0</td>
<td>Actimmune® (Genentech)</td>
</tr>
<tr>
<td>Sodium citrate</td>
<td>6.0</td>
<td>Nutropin® AQ (Genentech)</td>
</tr>
<tr>
<td>Sodium phosphate</td>
<td>7.3</td>
<td>Intron-A® (Schering-Plough)</td>
</tr>
<tr>
<td></td>
<td>7.4</td>
<td>Protropin® (Genentech)</td>
</tr>
<tr>
<td></td>
<td>7.5</td>
<td>Humatrope® (Lilly)</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>Orthodone OKT® 3 (Ortho)</td>
</tr>
<tr>
<td></td>
<td>7.5</td>
<td>Prolekine® (Chiran)</td>
</tr>
<tr>
<td></td>
<td>7.4</td>
<td>Nutropin® (Genentech)</td>
</tr>
<tr>
<td>Phosphoric acid</td>
<td>7.3</td>
<td>Activase® (Genentech)</td>
</tr>
</tbody>
</table>

*Recombinant biological products approved for marketing in the United States. Source Ref. 64.*
of degradation is slow at low temperature. Often, a pH stability profile for each formulation can be developed by plotting the rate constants of all degradation reactions versus pH. The most stable protein formulation selected should be the one with a pH at which the overall degradation reactions are minimal.

Accelerated Temperature Stability Studies. Because temperature also has effects on protein degradation reactions, stability testing at accelerated temperatures is often carried out by formulation scientists for formulation screening, and the degradation rate profiles in various formulations are compared using Arrhenius plots. An Arrhenius plot allows one to predict the degradation rate constants at storage temperatures other than those actually studied, assuming that the protein does not undergo physical changes at a given temperature. Formulation scientists often use the Arrhenius plot to extrapolate results obtained at higher temperatures to 5 °C for product shelf life prediction. However, at least three temperatures are required to obtain an Arrhenius plot, and extrapolation of these data should be done with understanding of the assumptions made regarding the linearity of the response and the accuracy of the rate constants. Small-molecule drugs that often degrade via a single degradation pathway usually follow Arrhenius behavior. However, due to complexity in protein structure, more than one degradation reaction can occur in a protein product. This may result in nonlinear Arrhenius behavior. Sometimes, the degradation pathways for proteins at higher temperatures are not the same as those at lower temperatures. For example, interleukin 1/α undergoes deamidation at or below 30 °C and aggregation at or above 39 °C (65). In this case, the formulation selected based on accelerated stability results may not be stable enough to protect the protein from degradation at lower temperatures. However, accelerated stability testing can still be a useful tool for protein formulation screening if the protein primarily degrades via only one degradation pathway within a selected temperature range, or if the degradation pathways are not influenced by one another (e.g., deamidation rate not affected by oxidation).

Stress Testing

After a stable protein formulation is selected from the formulation screening studies, stress tests are usually performed to determine whether additives are required to stabilize the protein against stresses such as agitation, freeze-thawing, and light exposure.

Agitation Studies. Proteins are susceptible to denaturation by mechanical force such as shaking or agitation. Shaking of liquid protein products can occur during filling, shipping, and handling. The mechanism of denaturation may involve an increase in area of the air–liquid interface during shaking. Proteins have a tendency to unfold and expose their hydrophobic groups at an air–liquid interface, resulting in denaturation and the formation of aggregates. Proteins also tend to concentrate at the air–liquid interface due to the presence of both polar and nonpolar side chains. Irreversible unfolding or aggregation of the protein occurs when the interface surface area increases (66). Henson explained that the turbidity observed in some protein solutions during vigorous shaking could be due to the formation of aggregates when the air–liquid interface increases (67). Formulation scientists often perform agitation studies on liquid protein formulations to determine if a surfactant is required to protect the protein from denaturation when agitation is unavoidable during manufacturing, shipping, and handling. The addition of surfactants could either reduce the interfacial tension or solubilize the concentrated protein at the interface (2). Surfactants commonly used as stabilizers in the pharmaceutical industry include polysorbate 20, polysorbate 80, pluronic 68, and brij, which are polymers composed of various combinations of ethylene oxide, propyl oxide, sorbitan, and fatty alcohols.

Freeze-Thaw Studies. Under many circumstances, such as processing, storage, shipping, and stability testing, freezing and freeze-thawing stresses may be applied to proteins. For instance, it is very common for a liquid protein bulk to be manufactured and stored as frozen bulk for long-term storage and then thawed before filling into containers for packaging as a liquid protein product or for freeze-drying as a lyophilized product. Sometimes, a protein formulation is frozen as a concentrated bulk for storage and then thawed and diluted with placebo before use. The risks of accidental freezing and thawing of proteins in liquid formulations may also exist during shipping and handling. Formulation scientists often perform short-term stability testing such that samples at each timepoint are kept frozen and then thawed at the end of the study for analysis. In lyophilized formulations, proteins are subjected to freezing during the lyophilization process. Therefore, freeze-thawing stability studies (at least three cycles of freezing and thawing) are usually performed on potential formulation candidates during formulation development to ensure the final product is stable upon freeze-thawing.

Protein denaturation during freezing is mainly caused by the phase separation due to the formation of ice. When ice is formed in a liquid protein formulation, the local concentration of protein as well as all excipients increases dramatically. The concentrated excipients in the non-ice phase may become destabilizing agents and cause the protein to denature (68). Some buffer species such as sodium phosphate and sodium succinate exhibit a decrease in pH during freezing. Freezing-induced pH changes may cause instability of recombinant proteins, as observed in IFN-γ formulated in succinate buffer (69).

If a protein formulation is evaluated to be denatured during freeze-thawing studies, formulation scientists may consider adding cryoprotectants such as sugars, polyols, amino acids, and inorganic and organic salts to prevent denaturation. The preferential exclusion mechanism of protection of proteins by these cryoprotectants during freeze-thawing was proposed by Timashelf and his coworkers (70). They suggested that the cryoprotectants, acting as cosolutes in the solution, are preferentially excluded from the protein. Consequently, the protein itself is preferentially hydrated by water. Hydration of a protein plays an important role in maintaining the conformational struc-
Lyophilization, also known as freeze-drying, is a process usually employed in the protein pharmaceutical industry to convert protein liquid formulations into solids of sufficient stability for storage. The process involves three major steps: freezing, primary drying, and secondary drying. During the initial freezing step, solvent (usually water) from an aqueous solution of protein drug that has been filled into glass vials is frozen and crystallized into solid ice at a low temperature such as −40 °C in a freeze dryer. The solutes including the protein and excipients are usually converted into an amorphous solid phase, which consists of uncrystallized solutes and uncrystallized water below their glass transition temperature, although some excipients such as buffer salts and mannitol may crystallize (75). The second step of the freeze-drying process, in which ice is removed from the frozen product by sublimation and condensation, is called primary drying. After the initial protein solution has been solidified, vacuum is applied to the drying chamber, and the shelf temperature is increased to supply heat to sublime ice from the frozen solid. Water vapor formed as a result of sublimation is condensed on cold surfaces in the condenser chamber. Primary drying is the longest part of the freeze-drying process and may require several days. After the bulk water (ice) has been removed from the partially dried product, secondary drying begins. During the stage of secondary drying, the final step of the freeze-drying process, the shelf temperature is increased to provide an elevated product temperature to remove the unfrozen water remaining in the amorphous solid phase. Secondary drying is usually performed over several hours at a shelf temperature between 25 and 50 °C, and the product temperature is maintained at about 25–35 °C.

Photostability testing is often performed during protein formulation development. The aim of a photostability study is to accelerate the formation of photodegradation products, if any, in a particular formulation upon exposure to light. FDA guidelines call for light stability testing of finished drug formulations but do not provide specifics of how to perform the test. A recent review article by Nema et al. provides an extensive overview of the protocols currently being used to perform photostability studies in pharmaceutical industry (73). Fluorescent lamps are usually used by pharmaceutical formulation scientists to simulate sunlight in photostability studies. Finished drug products are stored in a light cabinet with light intensity of 2–180 klx. The total exposure is between 8 and 4,500 klx days. Temperature inside the light cabinet is maintained at 25 ± 2 °C. Control samples, wrapped in aluminum foil, are exposed concurrently with the test samples.

Photolytic degradation of protein can be prevented by adding antioxidants to the formulation. Antioxidants can be classified into four main categories: chelating agents, reducing agents, oxygen scavengers, and chain terminators (74). If the addition of antioxidants in a protein formulation does not inhibit oxidation, the use of light-resistant containers is an alternative way of protecting light-sensitive protein drugs.

Lyophilized Formulation

Protein products that are unstable in solution or whose most-stable liquid formulations have insufficient long-term storage stability for marketing require lyophilization. However, the lyophilized products are usually limited to reconstitution and use in a short period of time. Lyophilized products also require longer time and more labor for formulation development than the ready-to-use liquid products. Lyophilization is also expensive due to the added costs of drying. Therefore, lyophilization is usually considered as an alternative method in protein formulation development when success with a liquid formulation cannot be achieved.

Light Stability Studies.

Some amino acid residues in proteins can absorb energy from UV light and become modified to form photooxidized degradation products, especially in the presence of photosensitisers such as molecular oxygen, riboflavin, dyes, and polymers containing other linkages (e.g., polysorbates). For example, the methionine residue in recombinant human relaxin can undergo photooxidation to form methionine sulfoxide (71), and the tryptophan residue in monoclonal IgG is converted to kynurenine and N-formylkynureneine upon exposure to UV light (72). The degree of photolytic degradation in the protein formulation can be influenced by many factors, including the buffer species and its concentration, excipients, and the formulation pH. Light intensity, duration of exposure, and storage temperature can also affect the rate of photolytic degradation.

Light stability testing is often performed during protein formulation development. The aim of a photostability study is to accelerate the formation of photodegradation products, if any, in a particular formulation upon exposure to light. FDA guidelines call for light stability testing of finished drug formulations but do not provide specifics of how to perform the test. A recent review article by Nema et al. provides an extensive overview of the protocols currently being used to perform photostability studies in pharmaceutical industry (73). Fluorescent lamps are usually used by pharmaceutical formulation scientists to simulate sunlight in photostability studies. Finished drug products are stored in a light cabinet with light intensity of 2–180 klx. The total exposure is between 8 and 4,500 klx days. Temperature inside the light cabinet is maintained at 25 ± 2 °C. Control samples, wrapped in aluminum foil, are exposed concurrently with the test samples.

Photolytic degradation of protein can be prevented by adding antioxidants to the formulation. Antioxidants can be classified into four main categories: chelating agents, reducing agents, oxygen scavengers, and chain terminators (74). If the addition of antioxidants in a protein formulation does not inhibit oxidation, the use of light-resistant containers is an alternative way of protecting light-sensitive protein drugs.
Supercooling is undesirable. The ideal freezing method is slower primary drying. Therefore, a very high degree of supercooling will result in very small pores, which means and appearance of the dried cake. A very high degree of tals and therefore determines the drying characteristics of supercooling determines the number and size of ice crys-

rapid ice growth as a result of low freezing temperature transition temperature (79). Typically, the final product reduced to well below the lowest eutectic temperature to If the solute crystallizes, the shelf temperature must be controlled by dosing the drug i.v., intramuscularly (i.m.),

Figure 2. Schematic representation of the conformation of a protein during freezing, drying, and rehydration. N (native), U (unfolding), $K_1$ (conformational equilibrium formed upon freezing), $k_1$ (rate constant for refolding), $k_2$ (rate constant for formation of irreversibly denatured forms). Source: Reprinted with permission from Ref. 78.

mannitol fail to protect labile proteins during freeze-drying because the sugar forms a separate crystalline phase during freezing (77). It has been hypothesized that lyoprotectants stabilize proteins by acting as a water replacement through hydrogen bonding to the dried protein, preventing conformational changes during drying (78).

In developing a freeze-drying cycle for lyophilized protein product, temperature control is important in each step of the process. During freezing, the protein solution is frozen so that most of the water is converted to ice and all solutes are converted to either crystalline solids or amphi-

N_{Dry} \quad N_{Frozen} \quad U_{Dry}

K_1

U_{Frozen} \quad U_{Frozen, Irreversible}

mannitol fail to protect labile proteins during freeze-drying because the sugar forms a separate crystalline phase during freezing (77). It has been hypothesized that lyoprotectants stabilize proteins by acting as a water replacement through hydrogen bonding to the dried protein, preventing conformational changes during drying (78).

In developing a freeze-drying cycle for lyophilized protein product, temperature control is important in each step of the process. During freezing, the protein solution is frozen so that most of the water is converted to ice and all solutes are converted to either crystalline solids or amphi-

N_{Dry} \quad N_{Frozen} \quad U_{Dry}

K_1

U_{Frozen} \quad U_{Frozen, Irreversible}

Long-term Stability Studies.

Long-term premarket stability studies of protein formulations are designed to establish the specific storage conditions and expiration dating for the drug product. It is usually required by the FDA that a pharmaceutical product contain not less than 90% of the label claim of active ingredient at the expiration date. Although real-time stability data are required by the FDA to demonstrate that the pharmaceutical product can meet this requirement prior to approval, an extrapolated expiration date based on the stability results obtained from clinical batches is usually allowed. Thus, stability-indicating assays are important for determining the stability of drug products. Analytical methods such as HPLC, electrophoresis, and spectroscopy have been widely used for this purpose.

In designing a long-term stability protocol for drug potency testing, one should consider the storage conditions and storage intervals. For refrigeration (5°C) and room temperature (15–30°C, by U.S.P. definition) storage, samples should be stored at 5, 15, and 25–30°C and monitored at 0, 1, 2, 3, 6, 9, 12, 18, and 24 months, and once yearly after. For lyophilized products, samples can be assayed at longer storage intervals and at higher storage temperatures such as 40°C. Pharmaceutical drugs are often monitored at high temperatures for the estimation of the degradation constants at low temperatures by Arrhenius plots. For high-temperature storage, samples are usually stored at 37 or 40°C and assayed at 1, 3, and 6 months. Samples may also be stored at 45 or 50°C and assayed at 2 weeks and 1 and 2 months. However, these elevated temperatures may approach the denaturation temperature of the protein, resulting in physical denaturation of the protein that would not be encountered in real situations, or extrapolate to lower-temperature behavior. In addition to determining the amount of active ingredient in a drug product, evaluation of the product should also include the following: pH, ionic strength, appearance, color, clarity, particulate matter, sterility, and pyrogenicity.

PROTEIN DRUG DELIVERY

Typically, proteins and many peptides must be administered by injection because of their poor bioavailability (fraction of administered drug in the circulation) by other routes. Initial clinical studies of these molecules is often conducted by dosing the drug i.v., intramuscularly (i.m.),
or subcutaneously (s.c.). These clinical studies assess the safety and tolerability of systemic exposure to the new drug. The results from early clinical trials combined with preclinical studies in animal models may indicate the need for a drug delivery system. Drug delivery systems are often required when the efficacious dose is not well tolerated, indicating a requirement to localize the drug at the site of action or maintain a low systemic steady-state level (e.g., provide same total drug dose at a lower maximum serum concentration). Localization of the drug to the site of action may also be necessary due to biological barriers preventing sufficient delivery of the drug via the systemic circulation (e.g., blood–brain barrier, blood–ocular barrier). In some situations, the disease state may be more readily accessible to an alternative delivery method than traditional parenteral administration. In the case of cystic fibrosis, a disease of the lungs, aerosol delivery of rhDNase has proven to be an effective therapy (81). Another rationale for using a drug delivery system is patient compliance or ease of use. For example, if a patient is required to have multiple injections of a drug per day or per week, it will be difficult to maintain proper compliance unless the drug effects are immediately felt by the patient and/or the disease is debilitating. These issues along with the in vitro (physicochemical properties) and in vivo (pharmacokinetics, pharmacodynamics, and toxicology) characteristics of the drug must be well understood prior to embarking on the development of a drug delivery system.

Selection of a Delivery System

Several potential options are available for drug delivery, as shown in Table 3 (82). These options include both invasive methods involving injections or noninvasive approaches that may or may not provide sufficient bioavailability. In each case, a stable formulation must first be developed, as described in the previous sections. For invasive methods, drug delivery may be applied with either devices or biodegradable depots. Delivery devices such as syringe pens and needleless injectors (injection by high-pressure gas) are gaining acceptance for many frequently administered drugs (83,84). In addition, implantable pumps are currently used with protein drugs such as insulin (85). Several injectable depot formulations made from the biodegradable polymer, poly(lactic-co-glycolic acid) (PLGA), are either approved (86,87) or in clinical trials (88). These depot systems offer the advantage of eliminating multiple injections and localizing drug to the site of action, if necessary. However, a major challenge in the development of these systems is the production of sterile injectable depots, usually microspheres on the order of 20 μm or greater in diameter. Approaches to developing microsphere depot systems for clinical and commercial use have recently been reviewed (89,90).

For noninvasive administration, aerosol delivery of proteins and peptides is the most developed and accepted delivery option. Aerosol delivery offers the unique option of delivering the drug through a single inhalation instead of a single injection. The lung provides a large surface area for absorption of the drug, but the aerodynamic properties of the drug (liquid or solid) play a key role in the deposition and, ultimately, the amount of drug delivered systemically. Several key issues such as drug stability and device selection have been reviewed recently (91). This work demonstrates that if the key parameters are analyzed and the appropriate drug formulation and device are selected, then aerosol delivery of proteins may be an effective method of administration. In contrast, the bioavailability of the protein or peptide when delivered by the other noninvasive routes is usually quite low (<1%), indicating that they are not an effective method for achieving significant systemic levels of the drug (for a review see Ref. 92). However, they may be very effective for local delivery of protein drugs if a low dose is efficacious.

Protein Stability in a Delivery System

After selecting a delivery system, the formulation scientist is required to develop a stable formulation that is compatible with the system and its intended use. These studies usually involve the formulation development described in the previous sections, as well as protein–delivery system interaction studies. The type of studies undertaken to assure delivery of biologically active and nonimmunogenic protein depend upon the system characteristics.

For injection devices, it is necessary to assess the protein stability in the container used to store the protein (e.g., cartridges) and the effect of the injection itself on the protein stability. These studies often involve storage stability studies of the protein in the device container as already described for protein filled into vials. The protein delivered from the device is also assessed by injecting into an empty sample tube or a buffer solution. Both recombinant human growth hormone (rGH) (93) and insulin (94) are currently available in injection devices. Typically, high-pressure devices such as the needleless injectors may cause denaturation of the protein due to the air–water interface and the high shear force applied to the liquid. Surface-induced denaturation may also occur at the device surface or at an air–water interface in the device. This degradation can be reduced and potentially eliminated by the addition of surfactants (e.g., polysorbates). In addition, implantable pump devices require studies to demonstrate maintenance of protein stability during the protein’s residence time in the pump at physiological temperatures (~37 °C). Several studies have been performed with insulin in these pumps (85), indicating the difficulty in designing stable formulations for these pumps. These systems pose a formidable challenge due to the potential for accumulation of degraded protein over time with repeated use of a refillable pump.

Similar issues are also encountered in the development of the injectable depot formulations. In this case, the protein remains in a biodegradable depot at physiological conditions for prolonged periods. The protein must maintain its potency during its in vivo release from the depot. Recent studies with rGH revealed that while it degrades during incubation at physiological pH, temperatures, and ionic strength, it maintains its biological activity and is not immunogenic (95,96). This result, however, cannot be assumed for each protein. Therefore, stability studies at physiological conditions should be performed prior to the development of a depot formulation to define the feasibility of the system and the acceptable duration of release (i.e.,
what level of degradation can be tolerated in terms of both potency and immunogenicity). These systems also pose a challenge in protein stability during the manufacture of the depot. For example, PLGA formulations require the dissolution of the polymer in organic solvents or use of high temperatures to melt the polymer, followed by dispersion of the protein in the PLGA and extraction of the organic solvent or reduction of the temperature, respectively. This exposure to harsh conditions often causes denaturation of the protein. Screening studies are performed to select excipients that will stabilize the protein against denaturation under these conditions. An example of this type of study was recently described using the model proteins (97). These studies revealed that excipients (e.g., sugars), which cause preferential hydration of the protein (cause an increased water layer at the protein surface), also increase protein stability during exposure to organic solvents. If these stability issues are overcome, an injectable depot system such as PLGA microspheres may be developed and tested in clinical trials.

Unlike depot systems, aerosol formulations for proteins do not require long-term stability at physiological conditions or stability in organic solvents. However, the aerosolization of proteins in their liquid state often leads to some protein degradation due to the air–water interface as well as the shear stress generated during the process. Several studies have been performed with traditional nebulizer devices that generate aqueous droplets of the appropriate diameter for efficient delivery to the lung. For rhGH, a surfactant was required to stabilize the protein during the process of aerosol formation (98), whereas rhDNase was very stable in an unbuffered isotonic solution (1). In many cases, it may not be possible to develop an aqueous protein formulation that both prevents degradation and is not an irritant to the lung. In addition, devices required to generate liquid droplets are usually not easily portable, making them less attractive to patients that are not hospitalized or bed-ridden. However, dry powder protein formulations may be delivered in small hand-held devices. To produce a protein powder with the proper aerodynamic properties, spray-drying of the protein is usually performed (99). This process usually involves high temperatures and aerosolization of the proteins of which may cause denaturation, as observed for rhGH (100). Once again, a formulation must be developed to prevent this denaturation. In addition, the dry powder protein formulation may be required to have stability at controlled room temperature (15–30 °C) and variable levels of residual humidity. The formulation scientist then carries out formulation screening experiments at different temperatures and residual humidities. If a protein powder is feasible, it offers advantages in ease of use and, perhaps, greater long-term stability.

**Manufacturing Issues**

The development of drug delivery systems may also require elaborate manufacturing procedures due to the complexity of both the system and the protein. When considering the manufacture of injection devices such as pens and pumps, some of the key issues are the ability to per-
form a sterile filling operation, the stability in the device, and the interaction between the drug, excipients, and the device or manufacturing components. The level of complexity is further increased in the case of pulmonary delivery. A nebulized liquid formulation is usually filled into ampules or other containers that allow easy addition of the liquid to the nebulizer by the patient. The filling of plastic ampules also has unique challenges in the ability to maintain sterility of the final product, the permeability of the material to gases and volatile solvents, and the physical properties of the container. In contrast, if a dry powder is used as the aerosol, the manufacturing process requires large-scale spray-drying equipment and machines for filling the dry powder into blister packs that can be used for metered-dose inhalers (99).

Manufacturing complexity is further increased when considering depot formulations such as PLGA microspheres. These microspheres are typically 20–90 μm in diameter and are therefore not readily sterilized by the traditional filtration methods (e.g., 0.22-μm filters). Terminal sterilization of these microspheres causes degradation of the polymer backbone, altering the release properties and often causing significant degradation of the protein within the microspheres. Microspheres must then be produced aseptically to avoid terminal sterilization while still maintaining the delivery of a sterile product to the patient. An aseptic process may be achieved through the use of barrier technology (isolators) for raw material input into the process and final product removal from the process. The remainder of the process operations is then designed as steam-in-place systems to assure a sterile environment throughout the production process.

If the unique challenges of a drug delivery system are overcome for a given protein, protein delivery may be successfully achieved. To reach commercial use, the drug delivery system must be well matched with the clinical and marketing requirements as well as the physiological requirements of the protein.

**BIBLIOGRAPHY**


INTRODUCTION

The primary object of freeze-drying is to preserve biological materials and pharmaceuticals without injury by freezing the contained water and then removing the ice by sublimation. Thus, one can expect to combine the advantages of both freezing and low-temperature dessication to obtain a more propitious state of preservation. In freeze-drying, the water or other solvent is removed as a vapor by sublimation from the frozen material in a vacuum chamber, leaving a dry porous mass of approximately the same size and shape as the original frozen mass. The resulting product is in a stable form and can be redissolved rapidly in water; the material displays what is called lyophilic behavior, which gave rise to the term lyophilization to describe the process of freeze-drying.

Lyophilization is employed soon after production for injectable pharmaceuticals that exhibit poor stability in solution, preventing their deactivation over a period of time and preserving their bioactivity. For example, a compound that is heat sensitive and undergoes rapid decomposition in aqueous solution can be formulated in a stable, rapidly soluble form.

THE FREEZE-DRYING PROCESS

Many pharmaceuticals and biological materials, which may not be heated even to moderate temperatures in ordinary drying, can be freeze-dried (1-6). Furthermore, many pharmaceutical products in solution deactivate over a period of time; the bioactivity of such compounds can be preserved by freeze-dried soon after their production, which stabilizes their molecules (2,6). Freeze-drying can also be applied to systems in which interfacial forces during drying need to be eliminated and where homogeneous powder products (derived from multicomponent, multiphase precursors) are required (2,5,7).

As a rule, freeze-drying produces the highest-quality pharmaceutical product obtainable by any drying method. One important factor is the surface structural rigidity of the frozen substance, where sublimination occurs. This rigidity, to a large extent, prevents collapse of the solid matrix remaining after drying. The result is a porous, nonshrunken structure that facilitates rapid and almost complete rehydration when water is added at a later time. Freeze-drying of biological materials and pharmaceuticals also has the advantage of low processing temperatures, the relative absence of liquid water, and the rapid transition of any local region of the material being dried from a fully hydrated to a nearly completely dehydrated state, all of which minimize the degradative reactions that normally occur in ordinary drying processes, such as protein denaturation and enzymatic reactions. In any pharmaceutical or biological material, some nonfrozen water (bound, or sorbed, water) will almost unavoidably be present during freeze-drying, but there is very often a rather sharp transition temperature for the still-wet region during drying (5), below which the product quality improves markedly. This improvement shows that sufficient water is frozen to provide the beneficial characteristics of freeze-drying.
However, freeze-drying is an expensive form of dehydration for pharmaceuticals because of the slow drying rate and the use of vacuum. The cost of processing is offset to some extent by the absence of any need for refrigerated handling and storage.

Systematic freeze-drying is a procedure mainly applied to the following categories of material (1–18):

1. Nonliving matter, such as blood plasma, serum, hormone solutions, foodstuffs, pharmaceuticals (e.g., antibiotics), ceramics, superconducting materials, and materials of historical importance (e.g., archaeological wood)
2. Surgical transplants, which are made nonviable so that the host cells can grow on them and use them as a skeleton (e.g., artery, bone, and skin)
3. Certain living cells that must remain viable for long periods of time, such as bacteria, yeasts, and viruses

Freeze-drying requires very low pressures or high vacuum in order to produce a satisfactory drying rate. If the water were in a pure state, freeze drying at or near 0 °C at an absolute pressure of 4.58 mmHg could be performed. But because the water usually exists in a combined state or a solution, the material must be cooled below 0 °C to keep the water in the solid phase. Most pharmaceutical freeze-drying is done at −20 °C or lower at absolute pressures of about 0.3 mmHg or less.

In short, freeze drying is a multistep operation:

1. The material is frozen hard by low-temperature cooling.
2. It is dried by direct sublimation of the frozen solvent and by desorption of the sorbed or bound solvent (nonfrozen solvent), generally under reduced pressure.
3. It is then stored in the dry state under controlled conditions (free of oxygen and water vapor and usually in airtight, opaque containers, filled with inert dry gas).

If correctly processed, most products can be kept in such a state for an almost unlimited period of time while retaining all their initial physical, chemical, biological, and organoleptic properties, and remaining available at any time for immediate reconstitution. In most cases reconstitution is accomplished by addition of the exact amount of solvent that has been extracted, thus giving to the reconstituted product a structure and appearance as close as possible to the original material. However, in some instances, reconstitution can be monitored in order to yield more concentrated or diluted products by controlling the amount of solvent. Vaccines and pharmaceutical materials are very often reconstituted in physiological solutions that are quite different from the original but are best suited for intramuscular or intravenous injection.

These examples are not exhaustive; detailed presentations on the uses of the freeze-drying process and of freeze-dried products are given in Refs. 1–10, 12, and 17–22.

**FREEZE-DRYING STAGES**

The freeze-drying separation method (process) involves three stages: (1) the freezing stage, (2) the primary drying stage, and (3) the secondary drying stage.

In the freezing stage, the product or solution to be processed is cooled down to a temperature at which all the material is in a frozen state.

In the primary drying stage, the frozen solvent is removed by sublimation; this requires that the pressure of the system (freeze-dryer) where the product is being dried must be less than or near to the equilibrium vapor pressure of the frozen solvent. If, for instance, frozen pure water (ice) is processed, then sublimation of pure water at or near 0 °C and at an absolute pressure of 4.58 mmHg could occur. But because the water usually exists in a combined state (e.g., gel suspension or biological material) or a solution (e.g., pharmaceutical product), the material must be cooled below 0 °C to keep the water in the frozen state. For this reason, during the primary drying stage the temperature of the frozen layer (Fig. 1) is most often at −20 °C or lower and at absolute pressures of about 0.3 mmHg or less.

As the solvent (ice) sublimes, the sublimation interface (plane of sublimation), which started at the outside surface (Fig. 1), recedes, and a porous shell of dried material remains. The heat for the latent heat of sublimation (2,840 kJ/kg ice) can be conducted through the layered material and through the frozen layer, as shown in Fig. 1. The vaporized solvent (water vapor) is transported through the porous layer of dried material. During the primary drying stage, some of the sorbed water (nonfrozen water) in the dried layer may be desorbed. The desorption process in the dried layer could affect the amount of heat that arrives at the sublimation interface, and therefore, it could affect the velocity of the moving sublimation front (interface). The time at which there is no more frozen layer (that is, there is no more sublimation interface) is taken to represent the end of the primary drying stage.

The secondary drying stage involves the removal of solvent (water) that did not freeze (sorbed, or bound, water). The secondary drying stage starts at the end of the primary drying stage, and the desorbed water vapor is transported through the pores of the material being dried.

**Figure 1.** Diagram of a material on a tray during freeze-drying; the variable $X$ denotes the position of the sublimation interface (front) between the freeze-dried layer (layer I) and the frozen material (layer II).
The Freezing Stage

The freezing stage represents the first separation step in the freeze-drying process, and the performance of the overall freeze-drying process depends significantly on this stage. The objective in freezing is to convert most of the water from the system by formation of ice and convert all solutes into solids, either crystalline solids or a glass. The material system to be processed (e.g., gel suspension, pharmaceutical solution, or biological material) is cooled down to a temperature (which depends on the nature of the product) that is always below the solidification temperature of the material system. For instance, if the material to be freeze-dried is a solution whose equilibrium phase diagram presents a eutectic point (e.g., the solution of NaCl and one of two different types of freezing behavior: (1) the liquid phase suddenly solidifies (eutectic formation) at a temperature that depends on the nature of the solids in the sample, or (2) the liquid phase does not solidify (glass formation), but rather it just becomes more and more viscous until it finally takes the form of a very stiff, highly viscous liquid. In this latter case there is no eutectic temperature, but there is a minimum freezing temperature.

At the end of the freezing step, there already exists a separation between the water to be removed (frozen water in the form of ice crystals) and the solute. In many cases, at the end of the freezing stage about 65–90% of the initial (at the start of the freezing stage) water is in the frozen state, and the remaining 10–35% of the initial water is in the sorbed (nonfrozen) state. The shape of the pores, the pore-size distribution, and pore connectivity (1–3, 6, 9–11, 13, 15, 16, 21, 23, 24) of the porous network of the dried layer formed by the sublimation of the frozen water during the primary drying stage depend on the ice crystals that formed during the freezing stage; this dependence is of extreme importance because the parameters that characterize the mass and heat transfer rates in the dried layer are influenced significantly by the porous structure of the dried layer. If the ice crystals are small and discontinuous, then the mass transfer rate of the water vapor in the dried layer could be limited. On the other hand, if large dendritic ice crystals are formed and homogeneous dispersion of the preeutectic and posteutectic frozen solution can be realized, the mass transfer rate of the water vapor in the dried layer could be high, allowing the product to be dried more quickly. Thus, the method and rate of freezing, as well as the shape that contains the solution and the nature of the product, are critical to the course of lyophilization because they affect the drying rate and the quality of the product.

In industrial pharmaceutical freeze-dryers the freezing of the product is usually done in the same plant where the drying occurs. In certain nonsterile pharmaceutical freeze-dryers the freezing of the product is accomplished by spraying liquid nitrogen into the drying chamber containing the product. In sterile pharmaceutical freeze-dryers the freezing stage is realized by contact between cooled plates and product-supporting containers.

The Primary Drying Stage

After the freezing stage, the drying chamber where the product is placed is evacuated and the chamber pressure is reduced down to a value that allows the sublimation of solvent (water) to take place. When the water molecules sublime and enter the vapor phase, they also keep with them a significant amount of the latent heat of sublimation (2,840 kJ/kg ice), and thus, the temperature of the frozen product is again reduced. If there is no heat supplied to the product by a heat source, then the vapor pressure of the water at the temperature of the product reaches the same value as that of the partial pressure of the water vapor in the drying chamber; therefore, the system reaches equilibrium and no additional water sublimation from the product occurs. In order to have continuous sublimation of water from the product, the latent heat of sublimation must be provided by a heat source. The heat is usually supplied by conduction, convection, and/or radiation; conduction is realized by contact between heated plates and product-supporting containers.

The amount of heat that can be supplied to the product can not be increased freely because there are certain limiting conditions that have to be satisfied during the primary drying stage. One of the constraints has to do with the maximum temperature that the dried product (freeze-dried layer in Fig. 1) can tolerate without (1) loss of bioactivity, (2) color change, (3) the possibility of degradative chemical and biochemical reactions, and (4) structural deformation in the dried layer. The maximum temperature that the dried product can tolerate without suffering any of these deleterious effects is denoted, for a given product, by $T_{scor}$ ($T_{scor}$ is often called, by convention, the temperature of the scorch point of the dried product). Another constraint has to do with the maximum temperature the frozen layer can tolerate and still remain frozen. If the material has a eutectic form and if the temperature of the lowest eutectic point is exceeded during the primary drying stage, then melting in the frozen layer (Fig. 1) can occur. Melting at the sublimation interface or any melting that occurs in the frozen layer can give rise to gross material faults such as puffing, shrinking, and structural topologies filled with liquid solution. When melting has occurred at some point in the frozen layer, then the solvent at that point cannot be removed by sublimation, and therefore, there is process failure (Fig. 1). There has also been, at least, a loss in structural stability. If the material is in a glass form and if the minimum freezing temperature is exceeded during the primary drying stage, then the phenomenon of collapse can occur, with a loss of rigidity in the solid matrix. Again in this case, there is a process failure because the water can no longer be removed from the frozen layer only by sublimation, and there has also been, at least, a loss in structural stability. The structural stability of a material relates to its ability to go through the freeze-drying process without change in size, porous structure, and shape. The maximum allowable temperature in the frozen layer is determined by both structural stability and product stability (e.g., product bioactivity) factors; that is, the maximum value of the temperature in the frozen layer during the primary drying stage must be such that the
The drying process is conducted without loss of product properties (e.g., bioactivity) and structural stability. Sometimes the product stability factors are related to structural stability factors (as in melting). There are systems in which the product stability factors do not depend on structural stability factors, as is the case for many vaccines, viruses, and bacteria where the temperature of the frozen layer during the primary drying stage must be kept well below the melting temperature so that there is a good level of bioactivity and organism survival after drying. In general, product stability is related to the temperature of the frozen layer during the primary drying stage. The maximum allowable temperature that the frozen layer can tolerate without suffering melting, puffing, shrinking, collapse, and loss of product property or stability, is denoted, for a given product, by $T_m$ ($T_m$ is often called, by convention, the melting temperature of the sublimation interface or of the frozen layer).

The water vapor produced by the sublimation of the frozen water in the frozen layer and by the desorption of sorbed (nonfrozen) water in the dried layer during the primary drying stage travels by diffusion and convective flow through the porous structure of the dried layer and enters the drying chamber. (It should be noted that most of the water removed during the primary drying stage is produced by sublimation of the frozen water in the frozen layer.) The water vapor must be continuously removed from the drying chamber in order to maintain nonequilibrium conditions for the drying process in the system. This is usually accomplished by fitting a refrigerated trap (called an ice condenser) between the drying chamber and the vacuum pump; the water vapor is collected on the cooled surface of the condenser in the form of ice. The time at which there is no more frozen layer is taken to represent the end of the primary drying stage.

**Secondary Drying Stage**

The secondary drying stage involves the removal of water that did not freeze (sorbed or bound water). In an ideal freeze-drying process, the secondary drying stage starts at the end of the primary drying stage. The word ideal is used here to suggest that in an ideal freeze-drying process only frozen water should be removed during the primary drying stage, whereas the sorbed water should be removed during the secondary drying stage. But, as just discussed, in real freeze-drying systems a small amount of sorbed water could be removed by desorption from the dried layer of the product during the primary drying stage, and thus, there could be some secondary drying in the dried layer of the product during the primary drying stage. In real freeze-drying processes the secondary drying stage is considered to start when all the ice has been removed by sublimation (end of primary drying stage). It is then considered that, during the secondary drying stage, most of the water that did not freeze (bound water) is removed. The bound moisture is present due to mechanisms of (1) physical adsorption, (2) chemical adsorption, and (3) water of crystallization. Although the amount of bound water is about 10–35% of the total moisture content (65–90% of the total moisture could be free water that was frozen and then removed by sublimation during the primary drying stage), its effect on the drying rate and overall drying time is very significant, and the time that it takes to remove the sorbed water could be as long or longer than the time that is required for the removal of the free water.

The bound water is removed by heating the product under vacuum. But, as in the case of primary drying, the amount of heat that can be supplied to the product cannot be increased freely because there are certain constraints that have to be satisfied during the secondary drying stage. The constraints have to do with the moisture content and the temperature of the product; these two variables influence the structural stability, as well as the product stability, during and after drying. For structural stability, the same phenomena, as in the case of the primary drying stage, have to be considered: collapse, melting (if the temperature is increased at constant moisture), or dissolution (if moisture is increased at constant temperature) of the solid matrix can occur. Product stability (e.g., bioactivity) is a function of both moisture content and temperature in the sample, and during secondary drying the moisture concentration and temperature in the sample could vary widely with location and time. This implies that the potential for product alteration to occur in the sample will vary with time and location. The moisture concentration profile is related to the temperature profile in the dried layer, and thus, the moisture content in the sample cannot be controlled independently. Because many products are temperature sensitive, it is usual to control product stability by limiting the temperature during the secondary drying process, and the final moisture content is checked before the end of the cycle (6,10,25).

In the secondary drying stage, the bound water is removed by heating the product under vacuum, and the heat is usually supplied to the product by conduction, convection, and/or radiation. The following product temperatures are usually employed: (1) below 0°C for vaccines, (2) between 10°C and 35°C for heat-sensitive products, and (3) 50°C or more for less heat-sensitive products.

The residual moisture content in the dried material at the end of the secondary drying stage, as well as the temperature at which the dried material is kept in storage, are critical factors in determining product stability during storage life. Some vaccines remain stable for many years when they are stored at ~20°C, but a significant loss of titer can be found after one year if they are stored at 37°C (26). Furthermore, certain vaccines such as live rubella and measles can be damaged by overdrying (final moisture content of about 2% is required for best titer retention), whereas other materials such as chemotherapeutics and antibiotics must be dried to a residual moisture content as low as 0.1% for best results.

**FREEZE-DRYING EQUIPMENT**

Freeze-drying equipment for pharmaceuticals is able to perform the three stages of the freeze-drying process: (1) freezing of the pharmaceutical product (usually a pharmaceutical aqueous solution), (2) primary drying of the product by removing the frozen solvent by sublimation,
and (3) secondary drying of the product by removing the sorbed, or bound, water (solvent that did not freeze).

All three stages of the freeze-drying process are energy consuming. The freezing stage involves freezing of the wet product. Because this is normally considered one of the preparatory steps before the actual freeze-drying, we concentrate here on the other two steps, which take place in the freeze-drying chamber (2,5,12). The primary and secondary drying stages involve the controlled supply of heat to the product to cover requirements for the sublimation and desorption processes, respectively. The removal from the freeze-drying chamber of the vast volumes of water vapor released during the sublimation and desorption processes always consumes the largest amount of energy. The efficiency of water vapor removal (the vapor trap system) therefore has a decisive effect on the total energy consumption of the freeze-drying plant.

The vapor trap is placed in an ice condenser chamber communicating with the freeze-drying chamber. The water vapor condenses to ice on its refrigerated surfaces.

When in operation, the efficiency of the vapor trap is shown by a small total temperature difference, $\Delta T$, between the saturation temperature for water vapor at the pressure in the freeze-drying cabinet and the evaporation temperature of the refrigerant (Fig. 2). This total temperature difference, $\Delta T$, results mainly from the following three resistances:

1. Pressure difference, $\Delta P$; equivalent to the pressure drop caused by the resistances to the vapor flow from the freeze-drying chamber to the cold surfaces of the vapor trap placed in the ice condenser chamber.
2. The temperature difference $\Delta T_{\text{ice}}$ through the layer of ice on the cold surface.
3. The temperature difference $\Delta T_{\text{refr}}$ between the cold surface and the evaporating refrigerant.

For an efficient vapor trap it is necessary to have a combination of a large cross-sectional area for the vapor flow (low $\Delta P$), an efficient deicing system (low $\Delta T_{\text{ice}}$), and an efficient refrigerating system (low $\Delta T_{\text{refr}}$). A less efficient vapor trap means a higher $\Delta T$, thus demanding a lower evaporation temperature of the refrigerating plant to maintain the required vacuum in the freeze-drying cabinet. A lower evaporation temperature means higher operation costs. In this temperature range, an evaporation temperature 10°C lower means a 50% increased in energy consumption.

The principal problem in the freeze-drying of pharmaceutical solutions is to operate in sterile conditions. In modern plants, the internal sterilization of the equipment is usually made with pressurized steam at 121°C or more; in old plants, sterilization is realized with the use of certain proprietary sanitizing agents (27,28). A pharmaceutical freeze-dryer must be designed according to current good manufacturing practice (cGMP) in order to be validated before and during use.

When evaluating industrial freeze-drying plants, the following characteristics are thus of prime importance:

1. Operation reliability
2. Ease, safety, and quality of process control
3. Product losses
4. Vapor trap efficiency
5. Internal sanitization
6. cGMP requirements

In the following sections, some of the commonly used types of pilot and industrial freeze-drying equipment for pharmaceuticals are presented, and their most important technical features are discussed.

Pilot Freeze-dryers

The freeze-drying of pharmaceutical products can be performed in the freeze-dryer as: (1) bulk solution freeze-drying in trays, or (2) freeze-drying in vials.

Freeze-drying pilot units appropriate for use in the pharmaceutical industries and in the laboratory are in high demand because they are used to explore possibilities for the preservation of labile products, especially those of biological origin. These units are portable and of convenient size for developmental work on freeze-dried products in laboratories and factories around the world. A large number of designs incorporate self-contained facilities for refrigeration, heating, and vacuum pumping, and they can freeze-dry batches consisting of 1 to 10 kg of frozen product. Because of the large variety of pilot freeze-dryers that are employed in industries and laboratories, and because of the limitation space we have for description, a pilot freeze-dryer is presented here whose characteristics are very close to the characteristics of industrial large-scale lyophilizers.

A schematic diagram of the pilot unit (Criofarma model C8-2), is shown in Figure 3a. The unit consists of (1) a freezing fluid system (R404A) that can be routed to the heat exchanger section of the condenser or into the refrigeration chamber pressure

Ice condenser Chamber pressure

$\Delta P$

$\Delta T_{\text{ice}}$

$\Delta T_{\text{refr}}$

Refrigerant

Figure 2. Graphic presentation of the variables $\Delta P$, $\Delta T_{\text{ice}}$, $\Delta T_{\text{refr}}$, and $\Delta T$.
Figure 3. Pilot freeze-dryer (Criofarma model C8-2). (a) Diagram of a pilot freeze-dryer: 1, drying chamber; 2, ice condenser chamber; 3, refrigeration unit; 4, cooling/heating system for the plates; 5, vacuum unit; 6, isolation butterfly valve; 7, silicon oil pump; 8, cooling/heating plate; 9, refrigerated coil; 10, condenser vacuum valve. (b) Frontal view of a pilot freeze-dryer: 1, drying chamber; 2, ice condenser chamber; 3, cooling/heating plate; 4, inspection window; 5, computer system; 6, vacuum indicator and regulator; 7, temperature control panel; 8, printer. (c) Frontal view of a laboratory freeze-dryer (Criofarma model Criolab-8). Source: Courtesy of Criofarma.

The rectangular drying chamber shown in Figure 3b is mounted on top of the condenser; the dimensions are 0.45 m x 0.45 m, and 0.5 m deep. Viewing windows are incorporated in the drying and condensation sections. The refrigeration and vacuum systems are in the internal part of the apparatus, whose complete dimensions are 1.1 m x 0.8 m, and 2.0 m high.

For pilot use, the freeze-dryer offers full control of the process variables and is able to achieve conditions of pressure and temperature beyond the limits of production units. The shelf and ice condenser temperatures of the pilot unit can be –55 °C and –75 °C, respectively, and the pressure in the drying chamber can be as low as 1 Pa or less. The pilot freeze-dryer has a control panel fully accessorized with instruments that record and display (1) the temperature inside the product, (2) the temperature on the plates, (3) the temperature of the coils of the condenser, (4) the pressure in the drying chamber, (5) the pressure in the vacuum unit, and (6) the pressure in the condenser section. The use of a personal computer in this pilot unit, with programmable temperature during the freeze-drying cycle and programmable input–output logic in the different freeze-drying stages, offers a wide variety of drying cycles and the capability for extensive data acquisition, so
Industrial Freeze-Dryers

The objective in a freeze-drying process is to convert most of the material into ice in the freezing stage, remove the ice by direct sublimation in the primary drying stage, and finally remove most of the unfrozen water in the secondary drying stage by desorption. In the industrial freeze-dryer, it is possible to perform all stages of the freeze-drying process. Most of the industrial freeze-dryers in operation are of the vacuum-batch type with freeze-drying of the product in trays. There are two main subtypes, depending on the type of condenser used. In one, the condenser plates are alongside the tray-heater assembly and in the same chamber; in the other, the ice condenser is in a separate chamber joined to the first, generally, by a wide butterfly valve. This latter type of plant is always used in pharmaceutical industries. Because of the wide variety and complexity of the problems associated with the production of pharmaceuticals by freeze-drying, in the following paragraphs the principal features of an industrial tray freeze-dryer for pharmaceuticals are presented.

The principal problem in the freeze-drying of pharmaceutical solutions is to operate in sterile conditions. The pharmaceutical solution may be sterile filtered immediately before introduction into a vial or a tray. Also the location of the plant must be able to guarantee sterile conditions during the filling, charging before drying, and discharging after drying of the pharmaceutical product. This is accomplished by facing the drying chamber door in a wall that separates the sterile room from the machine, or non-sterile, room. In the plant, this separation is accomplished with an isolation valve that separates the ice condenser from the drying chamber; this valve is also able to permit (1) the pressure rise test at the end of the freezing-drying cycle, (2) the simultaneous discharging/loading of the product and condenser defrosting, and (3) the reduction of cross-contamination between batches to a minimum. All the internal parts of the freeze-dryer are of stainless steel type AISI 316L with a finished surface of 0.1–0.4 rugosity or less. In modern plants, in order to avoid particle generation, the use of hydrogen peroxide (27) or certain proprietary sanitizing agents (28).

The product containers (vials or bottles loaded on stainless steel trays) are usually sterilized in a separate unit before filling and charging in the freeze-dryer. These operations require the presence of people in the sterile room, with consequent handling of the containers and possible contamination of the batch. For this reason, the human presence in the sterile room is usually reduced to a minimum necessary. A new freeze-dryer plant concept has been developed in order to reduce the risk of product contamination. The plant, as shown in Figure 4 (Criofarma model C292-BS), has two full doors. In the first (usually closed) there is a small door for loading the product before drying; the second full door, located opposite to the small door, is for discharging the product after drying. The condenser could be placed on the same floor, or on a floor that is below the floor where the drying chamber is located. The shelves of the freeze-dryer are lowered to the bottom of the drying chamber and are then lifted one by one to a position in line with the loading machine. The charging of the product is made under a laminar flow of sterile air; the small door is opened only for each plate loading and is then immediately closed. If the product is unstable and must be frozen within a short time after it is filled into its container, then it is possible to load trays of product onto the precooled shelves, half plate at a time. When the product container is a bottle, as shown in Figure 4, it usually has a silicon plug, which is partially inserted into the bottle; the solvent vapor leaves the container from the free space between the inserted portion of the plug and the container. After drying and before product discharge, the bottles are stopped in the drying chamber with the plugs, which are now fully introduced into the bottles. The stoppering operation is done in vacuum conditions, or at atmospheric pressure by breaking the drying chamber vacuum with sterile nitrogen, which prevents successive oxidation of the product; the latter method is most often employed in practice. The silicon plug in the stopped bottles provides a protection from contamination, and it may be possible to discharge the product in a less-sterile environment from the full door of the freeze-dryer in only one operation. The entire process may be fully automated as the bottles are removed from the filling machine; the disadvantage of automation is that the loading time of the freeze-dryer may become as large as the time it usually takes to complete the filling step of the operation, and this could reduce the theoretical freeze-dryer production for a large installation.

Pharmaceutical freeze-dryers are very often used to produce raw materials such as ampicillin, doxycycline, and cefazolin (usually as sodium salt), or other specialty materials such as collagen. In these systems, the product is usually charged on stainless steel or polyethylene film trays, and the plant is usually a medium or large unit with a loading surface varying from 20 m² to 60 m². If the product to be freeze-dried is not particularly unstable (e.g., collagen) and can withstand a delay of some hours before being filled into its tray and being frozen, then one can usually accumulate the trays of product on a loading trolley. When the loading trolley is filled, it is placed in front of the freeze-dryer, and the trays are automatically pushed on the shelves without sliding contact (in order to avoid particle generation), in a single operation. This system is advantageous because it permits maximum utilization of the freeze-dryer; the trolley may be loaded ahead of time, probably giving the average temperature of the product without placing temperature sensors in the product and, therefore, without risk of sterility compromise.
so that loading into the freeze-dryer can be carried out in a few minutes. If the product is not stable in the liquid state (e.g., ampicillin sodium salt) and must be frozen within a short time after its preparation, it is common to charge the empty trays on the precooled shelves and then to fill the trays, so that the freezing step is very quick and can proceed during the whole loading operation. This approach is also advantageous because it reduces the freeze-drying cycle time; this happens because the cooling phase starts at the same time as the loading phase, with a consequent reduction in the total time of these two steps.

The freeze-drying system used in the pharmaceutical industry basically consist of the following six units:

1. Drying chamber. Holds the product to be dried on shelves and provides an evacuated space for freeze-drying.
2. Ice condenser chamber. Collects and disposes of the water vapor produced during drying stages.
3. Shelf heating and cooling unit. Permits the heat transfer required for the freeze-drying cycle.
4. Vacuum pumping unit. Evacuates both the drying chamber and the ice condenser chamber to operating pressure and removes noncondensables such as air.
5. Refrigeration unit. Provides refrigeration for the chamber shelves during the freezing stage, and ice condenser plates for the primary and secondary drying stages.
6. Instruments and control panel. Indicates, records, and controls the shelf, product, and condenser temperature, as well as drying chamber, ice condenser chamber, and vacuum unit pressure.

A typical freeze-drying equipment layout (Criofarma model C292-8S) is shown in Figure 5.

The user of a pharmaceutical freeze-dryer must observe current cGMP for processes and equipment to be validated before and during use. As is the case in any validation effort, the protocol includes sections for installation, operation, and process. Each section of the protocol needs to be comprehensive and sufficiently detailed. The objectives of the validation program are to adequately demonstrate the equipment’s ability to support the process and to develop documentation demonstrating that control of the unit’s operation can be maintained. To develop and implement a protocol for lyophilization, careful consideration should be given to the installation qualification (IQ) and operational qualification (OQ) sections. The IQ section is the foundation of the protocol. Documentation of a general description, specifications, the support utility services, the equipment subsystems, and the installation of the lyophilization system should be addressed within the IQ section. This portion of the protocol is similar to that of any operation being validated. As part of the installation, an initial calibration should be performed to ensure that the instrumentation is in working condition after shipment and to ensure the accuracy of any data collected during start-up of the system. Calibration procedures and results also should be documented.

The OQ section can begin after successful completion of the installation qualification portion of the validation protocol. The objective of the operational qualification is to assure that the system performance is adequate to support the process for which the system is intended. The performance of the system needs to be compared either to the original equipment manufacturer’s specification or to the process parameter requirements as noted in the acceptance criteria. Tests should be performed on shelf temperature control, condenser cool-down, vacuum pumping rates, vacuum integrity, sublimation and condensation rates, vacuum level control, steam-in-place (SIP) system,
clean-in-place (CIP) system (cleaning the inside part of the plant with flushing steam or sterile water), and control hardware and software. The tests for validation of control hardware and software basically require suitable computer hardware and computer programs that perform consistently within preestablished operational limits so that analysis of the effects of possible failures can be carried out. If the tests are conducted in a manner that is in compliance with the procedures and that satisfies the acceptance criteria, the system will perform adequately at levels required to support the process.

The process validation section can be done as concurrent validation or as retrospective validation. Concurrent validation is applicable in situations in which the process data are collected and analyzed for a product that is to be marketed. Retrospective validation, which is particularly applicable to validation of lyophilization, is achieved through analysis of historical batch records. In this method, the data are gathered from records and then evaluated and analyzed. It is important to note that the validation of lyophilization processes is a sophisticated area that requires an understanding of the equipment, the process, and the product if a validation program for the process is to be implemented successfully.

Useful information for cGMP compliance, process, and computer system validation, can be found in Refs. 30–37.

FREEZE-DRYING OF PHARMACEUTICALS

Freeze-drying of pharmaceuticals can be divided according to the kind of product containers used to perform the process. It is possible to have freeze-drying of pharmaceutical products on trays, and freeze-drying of pharmaceutical products in vials. The different amount and kind of product processed (usually raw materials are freeze-dried on trays, whereas vaccines or drugs are freeze-dried in vials) and the different geometry of the container are very important relative to the process parameters and drying rates.

In Figure 1, a material being freeze-dried in a tray is shown. The thickness of the sides and bottom of the tray, as well as the material from which the tray is made, are most often in practice such that the resistance of the tray to heat transfer can be considered to be negligible (1–2,5–6, 8–10, 38–39). Heat $q_I$ can be supplied to the surface of the dried layer by conduction, convection, and/or radiation from the gas phase; this heat is then transferred by conduction to the frozen layer. Heat $q_I$ is supplied by a heating plate and is conducted through the bottom of the tray and through the frozen material to reach the sublimation interface or plane. The magnitude of the amount of heat $q_{III}$ in the vertical sides of the tray is much smaller (1–2,5–6, 8–10, 38–39) than that of $q_I$ or $q_{II}$; $q_{III}$ represents the amount of heat transferred between the environment in
the drying chamber and the vertical sides of the tray. Because the contribution of \( q_{\text{rad}} \) is rather negligible when compared with the contributions of \( q_{\text{vial}} \) and \( q_{\text{gas}} \), the contribution of \( q_{\text{cont}} \) to the drying rate will not be further considered (1–2,8–9). The terms \( N_W \) and \( N_{\text{in}} \) in Figure 1 represent the mass flux of water vapor and the total mass flux, respectively, in the dried layer. The total mass flux is equal to the sum of the mass fluxes of water vapor and inert gas, \( N_{\text{in}} = N_W + N_{\text{in}} \), where \( N_{\text{in}} \) denotes the mass flux of the inert gas.

In Figure 6, a typical lyophilizing vial (40) is shown during freeze-drying. The energy of sublimation of ice and desorption of sorbed water is supplied to the vial through radiation from the two shelf surfaces above and below the vial, conduction from vial–vial and vial–shelf contacts, and gaseous transfer along the bottom and side. The thickness of the pharmaceutical product charged into the vial usually is equivalent to the diameter of the vial. In pharmaceutical freeze-drying in vials, therefore, the heat transferred between the environment in the drying chamber and the surface of the vertical sides of the vial cannot be considered negligible (as in the case of pharmaceutical freeze-drying on trays) when compared with the contribution of the heat transferred between the dryer shelf and the surface of the bottom of the vial. For this reason the sublimation interface at time \( t = 0 \) is planar (as in the case of pharmaceutical freeze-drying on trays), but after the lyophilization process progresses, the moving interface acquires a parabolic shape because of energy entering along the sides of the vial (3,6,11). The traditionally accepted course of the moving sublimation interface is approximately 0-1-2-3-4. Experimental results (41–42) suggest that the dried-product mass transfer resistance is the dominant resistance, accounting for 82–92% of the overall mass transfer resistance, even when small (13-mm) closures are used. Closure and drying chamber mass transfer resistances are estimated (42) to be of the same order of magnitude, and in a well-designed freeze-dryer the chamber resistance (1–3,6) should be less than the resistance of the closure. A simple but satisfactory expression for estimating the resistance of the closure can be found in Ref. 42.

The goal of the pharmaceutical process designer and of the processor is to provide an economical drying process that gives reliably uniform and high product quality (1–4,18–20,43–44). A knowledge of the basic phenomena and mechanisms involved in freeze-drying is essential. A qualitative description and a mathematical model of the freeze-drying process on trays can be found in Refs. 1–3, 6, 9–10, 13, 15–16, and 39; the model can be used to analyze rates of freeze-drying of pharmaceuticals on trays (1–6,8–11,13–16,19,45–48). Qualitative description and a mathematical model of the freeze drying process in vials can be found in Refs. 3, 6, 9, 11, and 49.

### Temperature and Pressure Constraints in Freeze-Drying

A point that should be stressed is that, in any freeze-drying process, it will be desirable to fix the design and operating conditions such that the process is not rate limited by external resistances to either heat or mass transfer. The internal heat and mass transfer resistances are characteristic of the material being dried, but the external resistances are characteristic of the freeze-dryer. The design conditions of the equipment refer to having appropriate capacities for the vacuum pump, water vapor condenser, and heaters, and that the spacings between trays are such that the external heat and mass transfer resistances are not significant.

It is important to note at this point that, because a batch of a pharmaceutical product in an industrial freeze-dryer can easily be worth significant amounts of money, it is of paramount importance that the units and the control systems always operate under conditions at which there is insignificant loss in the quality of the product being freeze-dried. For this purpose, the freeze-dryer usually has one additional refrigeration and/or vacuum unit in a standby condition, and the control instrumentation is designed in such a way that the control policies can be implemented either automatically or by manual override.

In a well-designed freeze-dryer, the external resistances should not be the controlling factor in determining the drying rate (1–4,8–16). The internal heat and mass transfer resistances control the drying rate, and therefore, the freeze-drying process should be conducted within certain temperature and pressure limits in order to avoid product damage during the process.

Two temperature constraints (limits) may possibly be reached during the primary drying stage. First, the surface temperature of the dried layer must not become too high because of the risk of thermal damage; in fact the surface temperature of the dried layer must be kept below the scorch temperature, \( T_{\text{scor}} \), of the material being dried. Second, the temperature of the interface, \( T_I \), must be kept...
well below the melting temperature, \( T_m \), if the outer surface temperature limit \( T_{scor} \) is encountered first as the surface temperature of the dried layer is raised, the process is considered to be heat transfer controlled; to further increase the drying rate, the thermal conductivity, \( k_{ie} \), of the dried layer must be raised \((50,51)\). Many commercial freeze-drying processes are heat transfer controlled \((1–6)\). If the melting point temperature, \( T_m \), is encountered first, then the process is considered to be mass transfer limited, and in order to increase the drying rate, the effective diffusivity of water vapor in the dried layer, \( D_{win,e} \), and the total mass flux, \( N_t \), must be raised (an increase in \( N_t \) implies that the convective velocity of the vapor in the pores of the dried layer is increased); the values of \( D_{win,e} \) and \( N_t \) could be raised by decreasing the pressure in the drying chamber \((1,2,43)\). Because of the risks of lost bioactivity or structural and chemical damage of the product, the frozen layer temperature must be maintained below the melting point, which may in some cases be \(10^\circ C\) or more below the melting point of ice \((1–6, 8–17)\). Typical ice temperatures existing in the freeze-drying of pharmaceuticals under conditions in which the total pressure was primarily due to water vapor and the heat transfer took place via the dried \((I)\) and frozen \((II)\) layers are shown in Table 1.

The end of the primary drying stage occurs \((1–3,5,6,10)\) when the position of the moving sublimation interface is at \(X = L\); this condition implies that at the end of the primary drying stage there is no frozen \((II)\) layer, and therefore, there is no (moving) sublimation interface.

During the secondary drying stage, the temperature everywhere in the dried \((I)\) layer must be kept below the value of the scorch temperature, \( T_{scor} \).

The pressure constraints \((limits)\) are more complex to define because of the pressure effect on both the heat transfer \((pressure has effect on the thermal conductivity, \( k_{ie} \), of the dried layer) and the mass transfer \((pressure has effect on the effective diffusivity of water vapor in the dried layer, \( D_{win,e} \) and the total mass flux, \( N_t \)).

The effective thermal conductivity, \( k_{ie} \), in the dried material \((51)\) has been found to vary significantly with the total pressure and with the type of gas present. At very low pressures the thermal conductivity reaches a lower asymptotic value independent of the surrounding gas. This asymptotic conductivity reflects the geometric structure of the solid matrix itself, with no contribution from the gas in the voids of the material because the gas pressure is so low. At high pressures the thermal conductivity levels out again at a higher asymptotic value. This higher asymptote is characteristic of the heterogeneous matrix composed of solid material and the gas in the voids. Consequently, the high-pressure thermal conductivity is dependent upon the nature of the gas present and increases as the thermal conductivity of the gas increases and, hence, as the molecular weight of the gas decreases. When the thermal conductivity attains the high-pressure asymptotic value, the mean free path of the gas molecules rivals the void space dimensions in magnitude, but once the mean free path is reduced to the point at which the gas phase within the solid matrix obeys simple kinetic theory, the thermal conductivity stops rising. This reflects the fact that the thermal conductivity of a gas obeying simple kinetic theory is independent of the pressure.

The transition in thermal conductivity between asymptotes usually occurs between 0.1 and 100 mmHg, which includes the pressures characteristic of freeze-drying processes. The pressure range over which the transition in thermal conductivity between asymptotes occurs is characteristic of the pore-size distribution of the void spaces within the freeze-dried material \((52)\). A smaller pore dimension means that the gas must achieve a higher pressure in order for the mean free path of the gas to become comparable to the pore spacing and, hence, means that the transition between asymptotes will occur at higher pressures. Because fast freezing before freeze-drying leads to smaller pore spacing after freeze-drying \((53,54)\), it follows that faster freezing should lead to lower thermal conductivities at a given pressure. If a freeze-drying process is rate limited by internal heat transfer, the rate of freeze-drying for fast-frozen material should then be less than that of a slowly frozen material \((45,46)\). Thermal conductivity data for freeze-dried substances and pharmaceuticals are reported in Ref. 2.

The pressure effect on the mass transfer is explained by the fact that the effective diffusivity \((2,6)\) of water vapor in the dried layer, \( D_{win,e} \), and the total mass flux, \( N_t \), are a function of the total pressure, \( P \), in the drying chamber \((43)\). Elaborate expressions of \( D_{win,e} \) and \( N_t \) can be found in Refs. 2, 6, 9, 11, 15, and 16. In all cases the magnitude of \( D_{win,e} \) and \( N_t \) decrease if the total pressure, \( P \), is increased. If the value of \( D_{win,e} \) is decreased because of increased total pressure, then \((2,6)\) the value of the effective pore diffusivity, \( D_{win,e} \), decreases as the total pressure is

### Table 1. Frozen Layer and Maximum Dry Surface Temperatures in a Typical Freeze-Drying Operation Conducted with Heat Input through the Dry and Frozen Layers

<table>
<thead>
<tr>
<th>Pharmaceutical material</th>
<th>Chamber pressure (mm Hg)</th>
<th>Maximum surface temperature (°C)</th>
<th>Frozen-layer temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin sodium salt</td>
<td>0.15</td>
<td>40</td>
<td>-24</td>
</tr>
<tr>
<td>Cephalazin sodium salt</td>
<td>0.15</td>
<td>40</td>
<td>-25</td>
</tr>
<tr>
<td>Cloxacillin sodium salt</td>
<td>0.20</td>
<td>40</td>
<td>-20</td>
</tr>
<tr>
<td>Collagen</td>
<td>0.30</td>
<td>70</td>
<td>-20</td>
</tr>
<tr>
<td>Fructose-1,6-diphosphate</td>
<td>0.05</td>
<td>30</td>
<td>-45</td>
</tr>
<tr>
<td>Lactobacillus</td>
<td>0.10</td>
<td>20</td>
<td>-25</td>
</tr>
<tr>
<td>Rubella vaccine</td>
<td>0.05</td>
<td>10</td>
<td>-25</td>
</tr>
</tbody>
</table>
increased. Therefore, if the total pressure in the drying chamber is increased, then the effective diffusivity, $D_{inw,e}$ decreases, and thus, the diffusional mass flux of water vapor in the dried layer decreases. Furthermore, when the total pressure in the drying chamber is increased, the gradient of the total pressure, $\nabla P$, in the dried layer is reduced, and this decreases the convective velocity, $V_c$, and the total mass flux, $N_t$. Because the freeze-drying process will become internal mass transfer controlled above a certain pressure ($D_{inw,e}$ and $N_t$ decrease with increasing pressure, and $k_{ie}$ increases with pressure), the highest rate under mass transfer control will occur at the pressure of transition from heat transfer control to mass transfer control, and the attainable drying rate will decrease at higher pressures.

In general, operating conditions in freeze-drying pharmaceuticals include maximum surface temperatures of 20–40 °C and chamber pressures of 0.1–1 mmHg. Freeze drying of biological specimens, vaccines, and microorganisms is usually conducted with maximum surface temperatures of −20 to +20 °C and chamber pressures below 0.1 mmHg.

There has been considerable interest in investigating the factors affecting the batch time of freeze-dryers because this variable is most amenable to control, and efforts have been made to minimize the batch time (2,5,10,13,15,16,45–47).

The heat variables $q_i$ and $q_l$ from the energy sources, and the drying chamber pressure $P_{ch}$, $P_{ch} = P^0 = P_w^0 + P_{in}^0$ ($P_{ch} = P^0$ when the external mass transfer resistance is insignificant, as it would be the case with a well-designed freeze-dryer) are natural control variables. It should be emphasized at this point that (as the equations of the mathematical models in Ref. 2 indicate) the effects on the heat and mass transfer rates resulting from changes in the values of $q_i$, $q_l$, and $P_{ch}$ are coupled. The variable $P_w^0$ is taken to represent the water vapor pressure in the drying chamber (the external mass transfer resistance is taken to be insignificant), and its value is determined by the design and the operational temperature of the ice condenser. Thus, $P_{ch}$ may be changed by changes in $P_w^0$ ($P_w^0$ could be changed by changes in the temperature of the ice condenser), and by increasing or decreasing $P_{in}^0$. Therefore, changes in the temperature of the ice condenser affect the pressure, $P_{ch}$, (through $P_w^0$) in the drying chamber, and thus, the mass transfer rate in the dried layer.

Lipas and Litchfield (13) performed a quasi-steady state analysis for a system where $q_i \neq 0$ and $q_l = 0$ and obtained general guidelines about the optimal control policy at the beginning of the drying process (when neither of the state constraints is active), as well as during operation, when the process may be heat and/or mass transfer limited (10,13).

Millman et al. (10) studied the freeze-drying of skim milk under various operational policies that included the case where $q_i \neq 0$ and $q_l \neq 0$. They found that the control policy that produced the shortest primary drying stage was also the policy that provided the shortest overall drying time. Their results show that at least 80% of the heat used during the primary drying stage was transferred through the frozen layer of the sample. They also showed that the type of criterion used in terminating the secondary drying stage is of extreme importance, especially for samples of large thicknesses because it may lead to an undesirable sorbed (bound) water profile that may deteriorate the quality of the dried product.

### CONCLUSION

The evolution of freeze-drying in the past 40 years indicates that this separation process (unit operation) is a convenient method for drying those decomposable products (mostly pharmaceuticals) that cannot be stabilized in any other way (e.g., plasma, vaccines, antibiotics, sera, and growth hormones).

The economics of the process indicated that freeze-drying can be suitable for high-value products with specific biological and/or physicochemical properties. The highest cost advantages would be obtained from the processing of concentrated solutions of expensive materials; in this respect, freeze-drying could represent a viable alternative to filtration and crystallization. It is certain that freeze-drying has a future, and it is likely that this will essentially be in the fields of biological sciences and biotechnology. Its potential evolution is still great, and, of course, depend upon progress in basic research (4) and upon the level of creativity in the design and operating conditions of plants and instruments.

### NOMENCLATURE

- $D_{inw,e}$: Effective pore diffusivity of a binary mixture of inert gas and water vapor in the dried layer, $m^2/sec$
- $D_{win,e}$: Effective pore diffusivity of a binary mixture of water vapor and inert gas in the dried layer, $m^2/sec$
- $k_{ie}$: Effective thermal conductivity in the dried layer, kW/(m K)
- $k_{il}$: Thermal conductivity in the frozen layer, kW/(m K)
- $L$: Sample thickness, m
- $M_{in}$: Molecular weight of inert gas, kg/kg mol
- $M_w$: Molecular weight of water vapor, kg/kg mol
- $N_{in}$: Mass flux of inert gas in the dried layer, kg/(m² s)
- $N_t$: Total mass flux in the dried layer ($N_t = N_{in} + N_w$), kg/(m² s)
- $N_w$: Mass flux of water vapor in the dried layer, kg/(m² s)
- $p_{in}$: Partial pressure of inert gas at x = 0, N/m²
- $p_w$: Partial pressure of water vapor at x = 0, N/m²
- $p_{wX}$: Partial pressure of water vapor in equilibrium with the sublimation front ($p_{wX} = g(T_X)$), N/m²
- $P$: Total pressure ($P = p_{in} + p_w$) in the dried layer, N/m²
\[ P^0 \quad \text{Total pressure at } x = 0, \text{ N/m}^2 \]
\[ P_{ch} \quad \text{Total pressure in the drying chamber, N/m}^2 \]
\[ P_{chamber} \quad \text{Total pressure in the drying chamber, N/m}^2 \]
\[ P_{vial} \quad \text{Vial pressure, N/m}^2 \]
\[ q_i \quad \text{Heat flux at } x = 0, \text{ kW/m}^2 \]
\[ q_{li} \quad \text{Heat flux at the bottom of the tray, kW/m}^2 \]
\[ q_{lii} \quad \text{Heat flux at the sides of the tray, kW/m}^2 \]
\[ q_{contact} \quad \text{Conduction heat flux, kW/m}^2 \]
\[ q_{gas} \quad \text{Heat flux due to gaseous transfer, kW/m}^2 \]
\[ q_{rad} \quad \text{Radiation heat flux, kW/m}^2 \]
\[ t \quad \text{Time, s} \]
\[ T_i \quad \text{Temperature in the dried layer, K} \]
\[ T_{lii} \quad \text{Temperature in the frozen layer, K} \]
\[ T_m \quad \text{Melting temperature, K} \]
\[ T_{scor} \quad \text{Scorch temperature, K} \]
\[ T_X \quad \text{Temperature of the sublimation front, K} \]
\[ v_p \quad \text{Convective velocity of the binary mixture of water vapor and inert gas in the porous dried layer, m/sec} \]
\[ X \quad \text{Position of sublimation front (interface), m} \]

Greek Symbols

\[ \Delta P \quad \text{Pressure drop, N/m}^2 \]
\[ \Delta T \quad \text{Total temperature difference in Figure 2, K} \]
\[ \Delta T_{ice} \quad \text{Temperature difference through the layer of ice on the cold surface, K} \]
\[ \Delta T_{refr} \quad \text{Temperature difference between the cold surface and the evaporating refrigerant, K} \]

Subscripts

I \quad \text{Dried layer} \]
II \quad \text{Frozen layer} \]

BIBLIOGRAPHY

40. M.D. Cise, AIChE Summer National Meeting, Cleveland, Ohio, 1982, paper 22d.
43. C.J. Geankoplis, Transport Processes and Unit Operations, Allyn and Bacon, Boston, 1983.

See also Culture collections; Good manufacturing practice (GMP) and good industrial large scale practice (GLSP).