

Testing for Purity in Proteins by Gel Electrophoresis

Robert F. Peterson

Gel electrophoresis owes its high resolving power to imposition of the constraint of molecular sieving on movement due to the electrical force field. Since sieving is independent of the net charge of the protein, the contribution of net charge and size factors may be examined separately by newly developed methods. Milk proteins provide a convenient series of genetic variants, by which replacement of a single charged group may be evaluated. Electro-

phoresis in acrylamide gels containing sodium dodecyl sulfate removes net charge as a factor, since the charges are equalized. Mobility becomes a function of molecular weight and possibly configuration. Electrophoresis in a pH gradient using Vesterberg's ampholyte buffers is more sensitive for detecting impurities in protein preparations than ordinary gel electrophoresis.

Progress in protein research in the past three decades has been paced by the development of suitable analytical techniques such as ultracentrifugation and boundary and zone electrophoresis. The latter has been done with supporting media such as paper and starch gels, as well as in free solution. This paper will discuss the evolution of gel electrophoresis methods, using milk proteins as specific examples, and attempt to evaluate critically some newly developed procedures.

The first detection of genetic variation in proteins was done by Pauling and his associates using free electrophoresis (Pauling *et al.*, 1949). This was the identification of sickle cell anemia as a molecular disease. Milk proteins approach blood proteins in their number, diversity, and interest to biochemists. Possibly there is more freedom for variation in a protein made to be digested than in a circulatory protein which serves a more specialized function in the blood. Among the most important milk proteins are β -lactoglobulin and α -lactalbumin, which have been the subject of many physicochemical investigations. The study of the caseins, which are strongly complexed, lagged behind because the first analytical procedures did not completely resolve these complexes.

Free electrophoresis, developed in the 1930's by Tiselius (1937), was used by Mellander in 1939 to resolve whole casein at alkaline pH into three fractions which he named α -, β -, and γ -casein. Our own laboratory acquired an apparatus for free electrophoresis in 1940, and initiated intensive studies of the milk proteins.

Zone electrophoresis on paper was developed by Durrum (1950) and by Kunkel and Tiselius (1951). Here was an inexpensive and reliable method, easy to interpret, which avoided the confusion induced by monomer-polymer equi-

libria. Aschaffenburg and Drewry (1955) used paper electrophoresis at pH 8.6 in barbital buffer to show that β -lactoglobulin was genetically controlled, without dominance, and distinguished two types, A and B. Parenthetically, it is noteworthy that the electrophoresis had to be performed at room temperature. No separation occurs in the cold room. Later, using paper electrophoresis, he discovered genetic variation in β -casein. This involved addition of urea to a pH 7.15 phosphate buffer to dissociate the casein complexes (Aschaffenburg, 1961).

The next advance in resolution of proteins also involved zone electrophoresis; it was the introduction of starch gel electrophoresis (Smithies, 1955). The change from paper to starch gel increased the resolution tremendously. Wake and Baldwin (1961) extended the method to the caseins by addition of 7 M urea to the buffers. Thompson discovered genetic variation in α_{s1} -casein, the calcium precipitable fraction of the α -casein complex (Thompson *et al.*, 1962). Starch gels, despite their greatly improved resolving powers over paper, did have some disadvantages. To run a single gel required casting the gel, the actual electrophoresis run, a delicate slicing operation, staining the gel, and finally destaining. The total operation could take up to 3 days.

Acrylamide gel electrophoresis, developed independently by Raymond and Weintraub (1959) and by Ornstein (1964), provided yet another advance in separation techniques as well as convenience. The 3-day operation was condensed into 1 working day, and the slicing operation was eliminated. The new medium proved to be considerably more flexible to manipulation than starch gel. The porosity of the gel could be controlled by varying the concentration of acrylamide, and the extent of crosslinking could be varied independently. The commercial Cyanogum-41, containing 5% methylene-bisacrylamide, polymerizes to rather brittle gels above the 10% level. Reduction of the amount of bis improves the flexibility of the gels.

We experimented with water-cooled cells for acrylamide, so that higher voltages could be applied. About this time Raymond (1962) designed a vertical slab cell for electrophoresis which was produced commercially. We adopted it for its convenience in casting gels and because of its compactness. It also was water-cooled, and higher voltages could be applied with improved resolution. Our first studies with the milk proteins (Peterson, 1963) showed that Thompson's genetic variants of α_{s1} -casein and the Aschaffenburg variants of β -casein could be conveniently resolved at pH 9 in a gel containing 4.5 M urea. β -Lactoglobulin variants and α -lactalbumin could also be resolved if Bell's boric acid buffer (Bell, 1962) was used in 8% acrylamide gels. Addition of urea to these gels produced additional zones, probably due to disulfide interchange. Some protein denaturation is produced by the pH 9 buffers.

Structural studies of β -casein were started using trypsin hydrolyzates of β -casein A (Peterson *et al.*, 1958). High voltage electrophoresis of trypsin hydrolyzates of individual β -caseins and analyses of these peptides showed amino acid variation involving histidine replacement by proline. Considering that histidine would be charged at acid pH and proline would not be, we reasoned that electrophoresis at acid pH should theoretically resolve these caseins. We tried several buffer systems to achieve separation. One of these, formic and acetic acid, did resolve the histidine variants. A third histidine variant was discovered when studies were started on the frequency of distribution of the variants. β -Casein A had multiplied into A¹, A², and A³ (Peterson and Kopfler, 1966; Peterson *et al.*, 1966).

By using mercaptoethanol to reduce the disulfide linkage in κ -casein (the calcium-insensitive portion of the α -casein complex), the aggregate size of the κ -casein components was reduced to the point where they would migrate in acrylamide gels. Woychik (1964) used this technique to discover genetic variation in κ -casein, where two polymorphs A and B existed.

Groves used the acrylamide gel electrophoresis procedure in Raymond cells and showed genetic variation of the red protein, lactoferrin, in milk (Groves *et al.*, 1965). He then used the disc electrophoresis method of Ornstein to elucidate a complex picture of genetic variation in many of the minor proteins (Groves, 1969). Genetic variation in the γ -caseins seems to be linked to the occurrence of β -casein polymorphs (Groves and Gordon, 1969; Groves and Kiddy, 1970).

Why did gel electrophoresis have such increased resolving power? Paper is a web of coarse impure cellulose fibers. Adsorption of proteins on paper produces tailing of zones. Starch is purer and nearly devoid of electroosmotic effects, while acrylamide produces no electroosmotic effect. Inspection of the starting slots in starch or acrylamide gels shows that some proteins, which may stabilize complexes, are removed by molecular sieving. κ -Casein is removed in acrylamide gels, unless it is reduced with mercaptoethanol.

Migration in gels is a combination of the mobility in buffer (as in free electrophoresis) and the constraint imposed by gel filtration. Mathematical relations obtained between rate of flow of proteins on beads of acrylamide gel and pore size of the gels have application to acrylamide gel electrophoresis (Fantes and Furminger, 1967). The basic equation for electrophoresis in gels as stated by Tombs and Ackroyd (1967) is:

$$M = qE/f$$

where M = electrophoretic velocity, cm/sec, q is the net charge on the protein or aggregate, E is the field strength in V/cm, and f is the frictional constant. f is related to the diffusion

constant, D , in the medium by the relation $f = kT/D$, where k is Boltzmann's constant and T is the absolute temperature. A more detailed explanation of the relation between gel composition, pore size, and mobility is contained in a paper by Morris (1967).

We can utilize the sieving function of acrylamide gels and separate proteins on the basis of their ability to diffuse through a gel by conducting electrophoresis in a medium combining 0.1% sodium dodecyl sulfate, SDS, with phosphate buffer, 0.025 M, at pH 7. This method, described by Shapiro *et al.* (1967) for disc electrophoresis, also involves the addition of 2-mercaptoethanol to the protein solutions to reduce the disulfide linkages. The migrating units are then monomer chains. Reynolds and Tanford (1970) have shown that a wide variety of proteins bind identical amounts of sodium dodecyl sulfate at monomer equilibrium concentrations of SDS above 0.5 mM. When a large number of charged SDS molecules are bound to a protein, small charge differences are indistinguishable. Lysozyme, which still has a positive net charge at pH 9, in the presence of SDS has the charge reversed and migrates toward the anode. If the distance moved in a gel is plotted against the log of the molecular weight of the monomeric proteins, a straight line is obtained. Blattler and Reithel (1970) have published very useful graphs relating acrylamide concentration to exclusion limits for various proteins. They concluded that shape differences must be rather extreme to affect the results.

We have applied this method to the principal milk proteins using the published method of Koenig *et al.* (1970), which uses the Raymond vertical cell. As shown in Figure 1, the distance migrated is proportional to the logarithm of the molecular weight. This is in good agreement with the molecular weight found by amino acid analysis and also by sedimentation coefficients combined with diffusion constants. As shown, the method is sufficiently sensitive in 5% acrylamide gels to detect the deletion of eight amino acid residues out of 236 in the genetic variants of α_{s1} -casein A compared to the larger α_{s1} -casein C. It does not show any variation between the β -casein genetic variants, which have very similar molecular weights.

The inversion of the normal order in SDS-acrylamide gel electrophoresis is that the components of the casein complex, instead of migrating as α_{s1} followed by β , then κ , are ordered as β , α_{s1} , then κ , with little resolution. We conclude that the differences seen in ordinary acrylamide gels are due mainly to charge differences to which phosphoric acid groups are a major contributor.

We shall now consider the effect of net charge alone, since the β -casein genetic variants are inseparable by molecular sieving. Comparison of the amino acid analyses (Groves and Gordon, 1969; Peterson *et al.*, 1966; Ribadeau-Dumas *et al.*, 1970) of the β -caseins shows the following replacements, with the corresponding coding triplets.

A ¹ -A ²	His-Pro	CAC-CCC
A ² -A ³	His-Gln	CAC-CAG
A ¹ -C	Glu-Lys	GAA-AAA
A ² -B	Pro-His	CCC-CAC
	Ser-Arg	UCU-CGU
A ² -C	Pro-His	CCC-CAC
	Glu-Lys	GAA-AAA
B-C	Arg-Ser	CGU-UCU
	Glu-Lys	GAA-AAA

All of the variants can be accounted for by single base changes in the genetic code. If variation occurs at the same

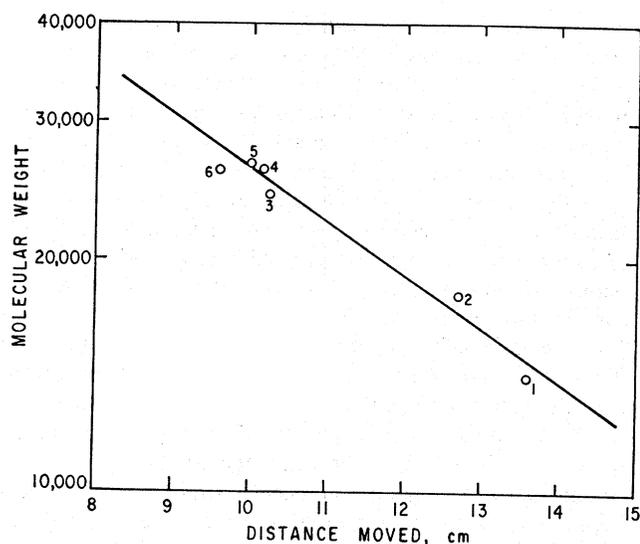


Figure 1. Molecular sieve electrophoresis of milk proteins. Electrophoresis in 5% acrylamide gel containing 0.025 M phosphate buffer, pH 7.0 and 0.1% sodium dodecyl sulfate. α -Lactalbumin, 1; β -lactoglobulin A, 2; β -casein A, 3; α_{s1} -casein A, 4; α_{s1} -casein C, 5; and κ -casein, 6

rate in nonpolar amino acids, probably three times as many β -casein variants await discovery.

The mobility of the known β -casein variants in acrylamide gels at pH 3 containing 4.5 M urea decreases in the order C, B, A¹, A², A³. The calculation of exact charges on β -caseins, which are assumed to be monomeric in urea containing buffers, is not possible because of unknown factors such as the amount of buffer ions bound and the shifts in ionization of groups. Assuming one oxygen of the phosphoric acid groups is protonated, the charges would be approximately: C, 20+; B, 19+; A¹, 17+; A², 16+; and A³, 15+, which is in agreement with the order found. At alkaline pH the β -caseins are resolved into three groups, the A variants being unresolved from each other. At alkaline pH exact determination of amide nitrogen is important, since each amide nitrogen subtracts one charge from the net negative charge. The usual calculation from the ammonia peak in amino acid analyses has too many sources of error. A rough calculation shows the A variants have 19-, 20-, and 21- charges, B has 18- charges, and C has 10- charges. The figure for B would seem to be in error; possibly the environment of the charged groups is a factor.

The β -lactoglobulins separate well at pH 9. Basch and Timasheff (1967) and Brignon *et al.* (1969) have completed titration data on the four genetic variants of β -lactoglobulin, A, B, C, and D. We can interpolate charge values of 52-, 50-, 48-, and 46- for the dimers, which agree with the order in electrophoresis.

A radically different type of electrophoresis which may be capable of adding another order of magnitude of resolution to gel electrophoresis is electrophoresis in a pH gradient. When Kolin (1954) pioneered the method, the first application was to colored proteins which moved rapidly to their isoelectric point in a miniature cell. The pH gradient rapidly disappeared and the results were difficult to record for proteins with no chromophore.

Svensson (1961) derived the theoretical requirements for an ampholyte which would generate its own natural pH gradient when a direct current was passed through its solution. Histidine peptides derived by hydrolysis of proteins were used in the first preparative apparatus. When histidine and

histidyl-histidine were used in a vertical water-cooled cell, the pH gradient was stabilized by a density gradient of sucrose. Hemoglobins A, S, and C, which differ by 0.2 pH unit, were separated after moving to their isoelectric points (Svensson, 1962). Vesterberg (1969a) developed a series of synthetic ampholytes from acrylic acid reacted with diamines. These covered a wide pH range and had low spectral absorbance at 280 m μ .

We obtained the preparative apparatus, 440 ml capacity, and a supply of Ampholine from LKB for a preliminary study of the milk proteins. First experiments with the milk proteins gave an unexpected result. Because of the low electrolyte concentration at the completion of the formation of the pH gradient, β -lactoglobulin was quite insoluble and the quantity of protein applied had to be reduced. For a wide pH range ampholyte mixture, such as pH 3-10 used in initial tests, the loading was only 10-20 mg. However, when narrow range ampholytes were used, pH 4-6, the loading could be increased. β -Lactoglobulins A and B were easily separated with isoelectric points measured at 4.85 and 5.00 units, a fair agreement with published values (Peterson, 1969).

The applications to analytical gel electrophoresis were quickly realized. Awdeh *et al.* (1968) published the details of forming a 23 cm \times 15 cm \times 0.1 cm gel by photopolymerization of 5% acrylamide gel which contained 2% of ampholyte. The solution to be gelled is placed between glass plates to exclude air and is exposed to light from a fluorescent tube viewing plate to activate the riboflavin catalyst. The polymerized gels, with the cover removed, were inverted and laid on carbon rods. One rod, the anode, was saturated with 5% sulfuric acid and the other was saturated with 5% ethylene diamine. The gels require only small amounts of the expensive ampholytes. To study the caseins, we added sufficient urea to the gel solution to make it 4.5 M with respect to urea. The samples were applied by laying small 1-cm squares of filter paper saturated with solutions of the samples in aqueous urea near the anode on the surface of the gels. After 16 hr at 200-400 V, the pH gradients were achieved and the proteins had formed narrow bands at their respective isoelectric points. The gels were transferred to 10% trichloroacetic acid and the proteins became immediately visible as opaque zones. After complete removal of the ampholytes the gels could be dyed with Amidoschwartz 10B.

As shown in Figure 2, the complexity of the patterns is much greater than that for acrylamide gel electrophoresis of whole caseins. Many bands appear in regions where we know of no casein fraction having such an isoelectric point. A clue to some of this behavior was noted when a preparation of casein was run in the preparative apparatus in 6 M urea and pH 3-10 ampholytes. An opalescent band, which formed in the alkaline region, broke up after 24 hr; a portion of this band migrated into the pH 4-5 region. However, if the preparative column was discharged into a fraction collector when the protein bands were first formed but sharply focused, as shown by incipient precipitation, the apparent isoelectric point was 7.0. This differs from the isoelectric point of 3.7 reported by Zittle and Custer (1963). This fraction behaved like κ -casein after ampholyte removal by dialysis. It was not precipitated by calcium at pH 7, and was coagulated by rennin.

At the present time the question of how many of the unknown bands are artifacts cannot be determined. β -Caseins A¹, A², A³, B, and C are resolved in pH 4-6 ampholytes. The resolution of the histidine variants is explained by the positive charges on the imidazole side-chain of the histidines in the region of the β -casein isoelectric points 4.6-4.9.

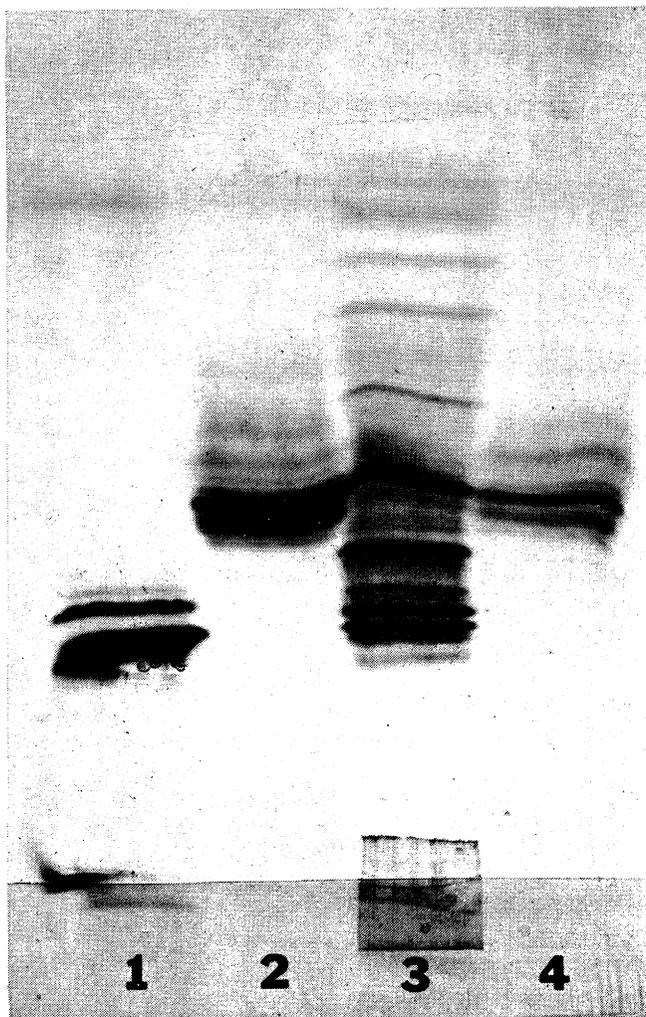


Figure 2. Electrophoresis in a pH gradient. pH 3–10 ampholyte, 5% acrylamide, 24 hr at 100–400 V. (1) α_{s1} -Casein C; (2) β -casein B; (3) whole acid casein; and (4) β -casein B

Preparative pH gradient electrophoresis of myoglobins in our apparatus has given three fractions. Preliminary histidine analyses of these fractions have given higher values than other published results. Vesterberg (1969b) has labeled ampholytes with C-14 in the carboxylic acid residues. The protein fractions, isolated with the labeled ampholytes, were brought to 0.5 M in NaCl and 0.1 M phosphate buffer at pH 7, and separated on Sephadex G-50 columns. The proteins retain only 0.01% of their original labeling C-14. At present we plan to isolate additional protein fractions by preparative pH gradient electrophoresis and analyze the isolated proteins in analytical pH gradient gels to see if they form single bands.

The last discussion points up the basic question we ask in gel electrophoresis: Is it pure if it migrates as a single band? Smithies (1955) found that when his bands were eluted from starch gel, they migrated exactly as before. Acrylamide gels, because of the very active catalyst systems employed, have a history of producing artifacts (Brewer, 1967; Fantes and Furminger, 1967; Schyns, 1968). Persulfate was first regarded as the culprit. However, it was demonstrated that a short pre-run with pH 9 gels removes the persulfate as detected by toluidine-KI reagent from the sample slots (Peterson and Bradshaw, 1967). Preparative acrylamide gel electrophoresis permits recovery of the sample from a scaled-up analytical apparatus. Despite a pre-run in alkaline gels,

β -lactoglobulin and β -casein lost most of their tryptophan content (Peterson and Peterson, 1969). If it is not persulfate, it could be long lived free radicals trapped in the gels. This hypothesis was tested by electron-spin spectrometry. Weak resonances were present in persulfate catalyzed gels; these were not removed by electrophoresis. Pre-electrophoresis with hydroquinone, cysteine, and thioglycollate is effective in removing the reactive agents. An easy proof for the live polymer theory exists: simply polymerize 10% acrylamide without crosslinking agent. A day later a portion will still catalyze the polymerization of deaerated acrylamide solutions. Possibly all proteins in acrylamide gels at the end of a run are different from the starting protein. Only if the electrophoresis is sufficiently prolonged (Peterson, 1963) will the artifactual bands be noted. If a fragment differs greatly in mobility, it may detach. It must be remembered that our detection system picks up only large molecules which will fix the dyes. Peptides are washed out from the gels if sufficiently soluble.

Another consequence of the persulfate catalyst is that the conductance of the gel at the beginning of the run is about three times its conductance at the end of the run. Hoagland (1968) has developed a programable power supply for electrophoresis which eliminated the constant attention required at the beginning of a run. Because of the poorer heat transfer, 10% acrylamide gels cannot be run at as high a current as 5% gels.

A last word of caution should be given to those who incorporate urea in their gels. Cyanate in the urea will carbamylate proteins in alkaline solutions. A protein sample, dissolved in urea, should never be used on a successive day. Cole and Mecham (1966) showed that the appearance of extra bands is minimal if the gel solutions are made up freshly from analytical grade urea. The heating necessary to form starch gels forms large amounts of cyanate. However, Melamed (1967) has found a way around the difficulty. By adding ZeoKarb 225 resin in the hydrogen form to the starch gel components before they are heated, the resin removes the cyanate, as shown by conductivity measurements. After the gel solution has cooled, the coarse resin is removed with a filter.

In conclusion, the most reliable way to check protein purity is to run the protein at pH 3 and pH 9 and then test for molecular heterogeneity by SDS electrophoresis. If very subtle differences exist, the new method of pH gradient gel electrophoresis may be able to select for smaller charge differences than any other method available, and possibly even select for changes in the environment of charged groups.

LITERATURE CITED

- Aschaffenburg, R., *Nature (London)* **192**, 431 (1961).
 Aschaffenburg, R., Drewry, J., *Nature (London)* **176**, 218 (1955).
 Awdeh, Z. L., Williamson, A. R., Askonas, B. A., *Nature (London)* **219**, 66 (1968).
 Basch, J. J., Timasheff, S. N., *Arch. Biochem. Biophys.* **118**, 37 (1967).
 Bell, K., *Nature (London)* **195**, 705 (1962).
 Blattler, D. P., Reithel, F. J., *J. Chromatogr.* **46**, 286 (1970).
 Brewer, J. M., *Science* **156**, 256 (1967).
 Brignon, G., Ribadeau-Dumas, B., Garnier, J., Pantaloni, D., Guinand, S., Basch, J. J., Timasheff, S. N., *Arch. Biochem. Biophys.* **129**, 720 (1969).
 Cole, E. G., Mecham, D. G., *Anal. Biochem.* **14**, 215 (1966).
 Durrum, E. L., *J. Amer. Chem. Soc.*, **72**, 2943 (1950).
 Fantes, K. H., Furminger, I. G. S., *Nature (London)* **215**, 750 (1967).
 Groves, M. L., *J. Dairy Sci.* **52**, 1155 (1969).
 Groves, M. L., Gordon, W. G., *Biochim. Biophys. Acta* **194**, 421 (1969).
 Groves, M. L., Kiddy, C. A., *J. Dairy Sci.* **53**, 931 (1970).
 Groves, M. L., Peterson, R. F., Kiddy, C. A., *Nature (London)* **207**, 1007 (1965).

- Hoagland, P. D., *Anal. Biochem.* **25**, 194 (1968).
Koenig, R., Stegemann, H., Francksen, H., Paul, H. L., *Biochim. Biophys. Acta* **207**, 184 (1970).
Kolin, A., *J. Chem. Phys.* **22**, 1628 (1954).
Kunkel, H. G., Tiselius, A., *J. Gen. Physiol.* **35**, 89 (1951).
Melamed, M. D., *Anal. Biochem.* **19**, 187 (1967).
Mellander, O., *Biochem. Z.* **300**, 240 (1939).
Morris, C. J. O. R., *Protides Biol. Fluids, Proc. 14th Colloq. Bruges, 1966*, p 543, Elsevier Publ. Co., Amsterdam, 1967.
Ornstein, L., *Ann. N. Y. Acad. Sci.* **121**, 321 (1964).
Pauling, L., Itano, H. A., Singer, S. J., Wells, I. C., *Science* **110**, 543 (1949).
Peterson, R. F., *J. Dairy Sci.* **46**, 1136 (1963).
Peterson, R. F., *J. Dairy Sci.* **52**, 903 (1969).
Peterson, R. F., Bradshaw, E., Abstracts, 154th Meeting ACS, C-135, Chicago, Ill., September 1967.
Peterson, R. F., Kopfler, F. C., *Biochem. Biophys. Res. Commun.* **22**, 388 (1966).
Peterson, R. F., Nauman, L. W., Hamilton, D. F., *J. Dairy Res.* **49**, 601 (1966).
Peterson, R. F., Nauman, L. W., McMeekin, T. L., *J. Amer. Chem. Soc.* **89**, 95 (1958).
Peterson, R. F., Sr., Peterson, R. F., Jr., Abstracts, 158th Meeting ACS, Biol. 151, New York, September 1969.
Raymond, S., *Clin. Chem.* **8**, 455 (1962).
Raymond, S., Weintraub, L., *Science* **130**, 711 (1959).
Reynolds, J. A., Tanford, C., *Proc. Nat. Acad. Sci.* **66**, 1002 (1970).
Ribadeau-Dumas, B., Grosclaude, F., Mercier, J. C., *C. R. Acad. Sci.* **270**, 2369 (1970).
Schyns, R., *J. Chromatogr.* **36**, 549 (1968).
Shapiro, A. L., Vinuela, E., Maizel, J. V., Jr., *Biochem. Biophys. Res. Commun.* **28**, 815 (1967).
Smithies, O., *Biochem. J.* **16**, 629 (1955).
Svensson, H., *Acta Chem. Scand.* **15**, 325 (1961).
Svensson, H., *Arch. Biochem. Biophys. Suppl.* **1**, 132 (1962).
Thompson, M. P., Kiddy, C. A., Pepper, L., Zittle, C. A., *Nature (London)* **195**, 1001 (1962).
Tiselius, A., *Trans. Faraday Soc.* **33**, 524 (1937).
Tombs, M. P., Ackroyd, P., Shandon Instrument Application, No.18, Shandon Scientific Co., Inc., Sewickley, Pa., January 1967.
Vesterberg, O., *Acta Chem. Scand.* **23**, 2653 (1969a).
Vesterberg, O., *Sci. Tools* **16**, 24-27, LKB Producter, Bromma 1, Sweden, 1969b.
Wake, R. G., Baldwin, R. L., *Biochim. Biophys. Acta* **47**, 225 (1961).
Woychik, J. H., *Biochem. Biophys. Res. Commun.* **16**, 267 (1964).
Zittle, C. A., Custer, J. H., *J. Dairy Sci.* **46**, 1183 (1963).