Journal of Steroid Biochemistry, Vol. 11, pp. 1629 to 1634
Pergamon Press Ltd 1979. Printed in Great Britain

ON THE ASSESSMENT OF VALIDITY OF STEROID RADIOIMMUNOASSAYS

S. Z. CEKAN
Swedish Medical Research Council, Reproductive Endocrinology Research Unit, Karolinska Hospital, 104 01 Stockholm, Sweden

(*Received* 8 *December* 1978)

SUMMARY

According to the generally accepted definition an assay is valid if the recognized criteria of validity are fulfilled, such as specificity as assessed by cross-reaction studies, parallelism of the dose response relationship for the standard an unknown, and agreement of results with the physiological state of the subject. However, these criteria are not stringent enough for the radioimmunoassay of steroids.

According to another definition an assay is valid only if it is intrinsically completely pure, i.e. if the substance measured is structurally unique, homogeneous and identical with the standard. However, this definition of validity has a limited practical meaning since probably no steroid radioimmunoassay of samples of biological origin can fulfil this condition.

As a practical solution to this problem it is proposed to couple the term validity to the outcome of the assays and to define a valid assay as the one which yields results identical with the true value within the limits of experimental error.

It is suggested that in practice, when various antisera and purification techniques are compared, the assay procedure yielding the lowest estimates may usually be assumed to be closest to validity. However, the experiments of this type cannot result in a proof of validity. Such a proof may be obtained by a validity test based on a comparison of results yielded by the radioimmunoassay tested for validity on the one hand, with those obtained by a reference radioimmunoassay performed on the sample purified to radiochemical purity on the other hand. Since the radiochemically pure sample can be assumed to represent the true value, the reference assay is valid. The assay tested is to be considered valid as well, if the results yielded by this assay cannot be distinguished statistically from those obtained by the reference assay. The advantage of this test of validity is that it can be performed in every radioimmunoassay laboratory.

The above concepts are illustrated by references to previous studies on progesterone, testosterone and norethisterone assays. An example of the validity test for a progesterone radioimmunoassay is given.

The term "validity" is often used for the assessment of assay systems or for the evaluation of individual assays. On these occasions the following questions are asked: Is this assay method valid or not, *i.e.* does it yield valid or invalid results, or, are these particular results valid or invalid? The answers depend on the definition of validity used.

A generally accepted definition of validity says that an assay is valid if the recognized validity (or reliability) criteria are fulfilled, such as high specificity as assessed by cross-reaction studies, parallelism of the dose response relationship for the standard and unknown, and agreement of results with the physiological state of the subject. However, it will be shown later that these criteria are not sufficiently stringent since they do not guarantee that the assay measures the real contents of the steroid in the biological material investigated.

Another definition of validity was presented recently by Ekins[1] who postulated that an assay is valid if the substance measured is structurally unique, homogeneous and identical with the standard. This definition is undoubtedly correct in theory since an ideal radioimmunoassay containing the unknown in a chemically pure form should yield valid results. However, this definition seems to be excessively rigorous for the daily routine, since no ordinary radioimmunoassay of samples of biological origin fulfills and probably ever will fulfil this condition.

Instead, a third definition of validity is proposed in this review. According to this definition an assay is valid, if the results express the true concentration of the steroid in the unknown sample within the limits of experimental error*. In contradistinction to the definition put forward by Ekins[1], the definition proposed takes into consideration the specificity of the antiserum used and it permits the presence in the assay of those compounds which do not appreciably react with the antibody.

In this review the validity criteria so far used will be discussed, but the main attention will be devoted to ways of approaching or achieving validity in the sense proposed, and of testing for it.

When testing a new progesterone antiserum some

* According to the terminology used previously [2, 3], a valid assay is an assay of an absolute accuracy and of an accepted degree of precision.

Table 1. Cross-reactions (%)* with two progesterone antisera

Steroid	Antiserum A†	Antiserum B‡
5α-Pregnane-3,20-dione	57	20
5β-Pregnane-3,20-dione	96	9
Deoxycorticosterone	17	0.5
20α-Dihydroprogesterone	3	0.1
17-Hydroxyprogesterone	3	0.1

* Estimated at 50% binding [4].

† Raised against bovine serum albumin conjugated with 11α-hydroxy-4-pregnene-3,20-dione hemisuccinate; the cross-reactions were published earlier [4].

‡ Raised against bovine serum albumin conjugated with progesterone 3-(O-carboxymethyl)oxime; the cross-reactions have not been published.

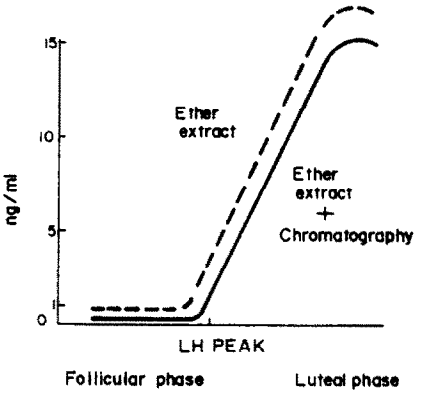

Fig. 1. Schematic representation of progesterone measurements in the human plasma following two different procedures for the purification of this compound prior to radioimmunoassay proper.

time ago, we expected a substantial increase in accuracy of the assay since the new antiserum (antiserum B in Table 1) exhibited a considerably lower degree of cross-reactions with some steroids (Table 1). However, when a direct radioimmunoassay (*i.e.* a radioimmunoassay of plasma ether extracts) was performed in samples of follicular phase plasma, overestimations were found which were much higher than those seen in a previous study [5]. Using the new antiserum, the results were approximately three times higher (Fig. 1), than those obtained by a radioimmunoassay in which the ether extraction was followed by a celite chromatography [4]. The latter radioimmunoassay yielded results which corresponded to those usually found during the follicular phase (approx. 300 pg/ml [6]). In luteal phase samples the direct assay also gave markedly higher results (Fig. 1). However, in spite of these overestimations, an agreement of the results was achieved with the well known physiological increase of circulating progesterone in the menstrual cycle. This example well illustrates the fact that the agreement of results with the physiological state of the subject is a weak criterion of the assay validity.

The inadequacy of parallelism as a criterion of validity was demonstrated previously [2, 3], but the progesterone assay mentioned may be used as an additional example. When progesterone was assayed using the direct radioimmunoassay and the logit responses to increasing doses were compared with the logit responses to increasing volumes of follicular phase plasma, non-parallelism was found* (Fig. 2).

On the other hand, a parallel relationship was obtained with the luteal phase plasma (Fig. 2). Thus the parallelism test did not detect the invalidity of the luteal phase plasma measurements. This happened in spite of the fact that the overestimation was of approximately the same (or even larger) magnitude in absolute terms than that seen in the follicular phase plasma†.

The above example demonstrated the unreliability of the validity criteria used so far. On the other hand, a fruitful working hypothesis could be derived from this example. According to this hypothesis, (1) the character of biological material investigated exerts a considerable influence on assay accuracy and validity, and (2) comparisons of various antisera and purification procedures are better suited to assess validity than the validity criteria used so far; these compari-

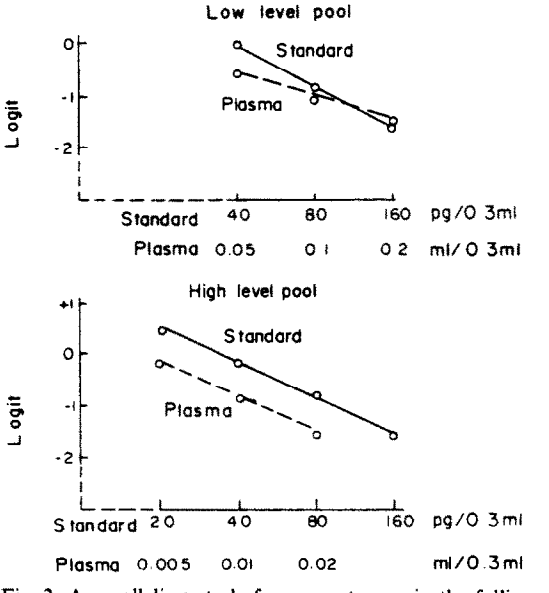

Fig. 2. A parallelism study for progesterone in the follicular phase plasma pool (= low-level pool) and in the luteal phase plasma pool (= high-level pool).

* Because of this non-parallelism, the measurements in the follicular phase samples mentioned above are to be considered as semi-quantitative estimations only.

† The lower sensitivity of the parallelism test to detect interfering compounds in the luteal phase plasma seems to be due to the "dilution effect" which means that this plasma can be assayed in a higher dilution than is possible with the follicular phase plasma. Due to the higher dilution, the amount of interfering compounds is lower in the assay tube.

Table 2. Means (n = 9) and standard deviations (in parentheses) of the levels of testosterone* in two plasma pools†, as measured using three antisera and two purification procedures

Antiserum	X		Y		Z	
Purification‡	E	C	E	C	E	C
Female pool (pg/ml)	1110	313	700	283	337	311
	(101)	(10.0)	(89.1)	(15.8)	(31.6)	(33.4)
	h§	1	h	L	1	1
Male pool (ng/ml)	5.09	4.36	4.23	4.29	4.37	4.40
	(0.19)	(0.21)	(0.20)	(0.09)	(0.27)	(0.30)
	h	1	L	1	1	1

* Parallelism was obtained in all assays.

† The female pool was collected from eight normally menstruating women (age 22–36 years), the male pool from six men (age 32–45 years).

‡ E: extraction with ether, C: chromatography on celite [4] of the ether extract.

§ A one-way analysis of variance and appropriate contrasts were computed for each pool; the significance of the differences between individual means is expressed as follows: L represents the lowest mean in a pool, 1 indicates a value which is statistically non-distinguishable from L, h denotes a value which is statistically higher than L at 95% or higher confidence level.

sons can detect which antiserum and which purification procedure gives the lowest results which in turn may usually* be considered as being closest to validity.

The effect of the type of biological material investigated on assay accuracy could be seen in a very distinct manner in a study (M. S. de Gomez, S. Z. Cekan, P. Vitins and A. Frei; to be published) in which three different antisera were compared in a testosterone radioimmunoassay and, simultaneously, two plasma pools (a female and a male pool) were investigated. For each antiserum and each pool the assay was performed both on ether extracts of plasma (direct assay) and ether extracts further purified by celite chromatography (chromatographic assay). In order to avoid the influence of the day-to-day assay variation on the comparisons, all assays involving one pool were performed on one day. It can be seen from the results (Table 2) that the chromatographic method used in the assay of the female pool yielded statistically undistinguishable results for all three antisera. It may be assumed that these results are closest to validity. When the direct method was used, various degrees of accuracy were observed. The result closest to validity was obtained with antiserum Z, an overestimate by approximately 100% was found with antiserum Y, and with antiserum X a 3–4 fold overestimate was observed. In contradistinction to the female pool the differences seen in the male pool were much smaller. Again, the chromatographic procedures gave a result which seemed to be close to validity, but also two direct procedures yielded results which could not be distinguished from those obtained by the chromatographic method.

This example may be used to demonstrate still another aspect of the effect of biological material on accuracy. It may be assumed that the direct assay using antiserum Y (Table 2) is valid for the male plasma pool. It is apparent from the data shown in Table 2 that it would be extremely hazardous to extrapolate the assumption of validity to the female pool, if the same antiserum Y and the same direct method were used.

The next investigation seems to be well suited to demonstrate the usefulness of the comparisons of various antisera and purification procedures for the detection of a procedure (or procedures) which are closest to validity. In this investigation [7] 4 norethisterone antisera were compared. Both the direct and chromatographic assays were performed using each antiserum. In this case 35 plasma samples obtained from seven women receiving norethisterone contraceptive pills were assayed. In order to avoid the influence of the day-to-day variation, the samples from one patient were always assayed on one day using all four antisera, and for each antiserum using both the direct and chromatographic procedures. It may be seen (Table 3) that antiserum C gave the lowest results when both the direct and chromatographic method were used. Antiserum D yielded results of the same magnitude, but only after chromatography. Again, the lowest results can be considered as being closest to validity. It is interesting to note that the chromatographic procedure gave significantly higher results with antisera A and B. This signifies that a single chromatography did not succeed in separating all interfering compounds, and that those interfering compounds were not detected by antisera C and D, but were registered by antisera A and B. Another interesting fact is that the best antiserum (C) was not the best one as far as the apparent cross-reactions were concerned. This antiserum which was raised against an antigen coupled to BSA in position 3 exhibited a considerable degree of cross-reaction with norethisterone metabolites, in contrast to antiserum

* The lowest result will not be closest to validity (closest to the true value) in cases when a negative blank is present.

Table 3. Results of norethisterone (NET) radioimmunoassays* performed on female plasma samples† when four antisera and two purification procedures were used

Antiserum	A		B		C		D	
Position of coupling of NET to BSA	3		3		3		11	
Cross-reactions (%)‡								
5β-Dihydro-NET	28		8		23		6	
3β,5α-Tetrahydro-NET	10		19		19		2	
3β,5β-Tetrahydro-NET	7		3		13		0.4	
Purification§	E	C	E	C	E	C	E	C
Geometric means	504	356	410	353	317	314	532	295
95% Confidence limits	427–596	281–450	336–500	283–440	249–404	242–408	458–619	230–379
Significance‖	H	H	H	H	L	L	H	L

* In all assays parallelism was obtained. For further details cf [7].

† NET levels were measured in 35 plasma samples collected from seven subjects (five samples per subject); the samples from one subject were assayed using both purification procedures and all four antisera always on the same day.

‡ The cross-reactions were estimated at 50% binding.

§ E: extraction of plasma with ether, C: chromatography on celite of the ether extract.

‖ The assessment of significance is based on an analysis of variance and on a computation of appropriate contrasts. The results denoted by H are significantly higher than those denoted by L (at at least 95% confidence level).

D which was raised against the 11-hydroxy conjugate with BSA and gave a relatively low degree of cross-reaction.

The study of norethisterone radioimmunoassays demonstrated that in fact the degree of accuracy may be independent of the formal cross-reactions (cf. also the progesterone study mentioned above), that the antigen structure may have a low predictive value from the point of view of validity, and that the only way of arriving at a procedure which most approaches validity is a comparison of purification procedures and antisera.

However, the results of such comparisons cannot be considered as a proof of validity. Such a proof may be obtained *e.g.* by a test of validity which is based on a comparison of results yielded by the radioimmunoassay tested for validity on the one hand, with those obtained by a reference radioimmunoassay performed on the sample purified to radiochemical purity* on the other hand (Fig. 3). Since it may be assumed that the contents of the radiochemically pure compound represent the true concentration of the compound in the sample, the reference assay may be assumed to be valid. If the results of this assay are statistically undistinguishable from those yielded by the assay under test, also the latter assay is to be considered valid.

As an example a validity test for a progesterone radioimmunoassay [8] may be given. A sample of a plasma pool from normal women in the follicular phase of the menstrual cycle was extracted with ether, chromatographed on a celite–propylene glycol (1:1, w/v) column in isooctane [4] and individual chromatographic fractions (1 ml) were collected (Fig. 4). The radioactivity of freshly purified [^{3}H]-progesterone—added to the plasma prior to the extraction and chromatography—was measured in each fraction. Furthermore, the contents of progesterone were determined by means of radioimmunoassay† in the same fractions. It may be seen in Fig. 4a that a contamination would be present in the progesterone fraction, if fractions 4–7‡ were pooled as is the case in a current radioimmunoassay [4]. When fractions 0–3 were pooled and re-chromatographed in the same system (Fig. 4b), a peak of compounds was obtained which behaved as progesterone in the radioimmunoassay, but was completely separated from the radioactive standard. When fractions 4–6 were re-chromatographed, an agreement of the mass and radioactivity

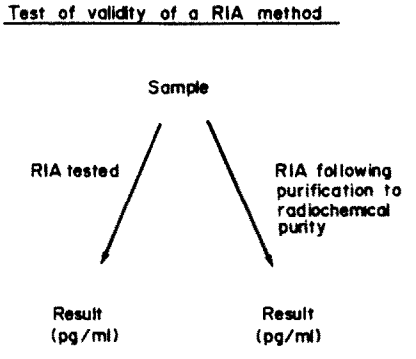

Fig. 3. Schematic representation of a validity test.

* Radiochemical purity in the present sense means that a sample exhibits a constant ratio of radioactivity and mass (constant specific activity) in individual fractions of a chromatographic zone containing the compound assayed.

† Since there was a varying amount of radioactive progesterone in individual fractions, it was important to correct the radioimmunoassay results for the mass of the radioactive compound in each fraction. The computation of this correction was described earlier [3, 8].

‡ Neither radioactivity nor mass were found in fraction 7 of the chromatography shown in Fig. 4a.

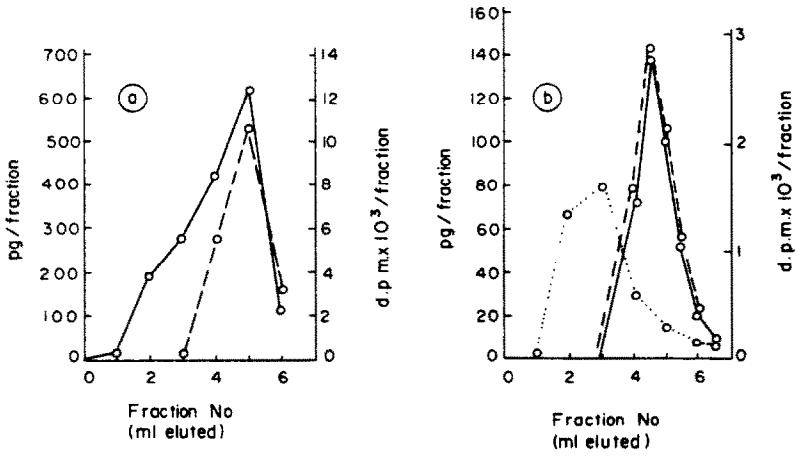

Fig. 4. Chromatographic isolation of progesterone from plasma. (a) Chromatography of plasma ether extract on a celite–propylene glycol (1:1, w/v). The solid lines indicate the contents of progesterone-like compounds in individual fractions (1 ml) of the eluate, as established by radioimmunoassay. The broken lines demonstrate the radioactivity present in the same fractions. (b) Re-chromatography on celite of the fractions obtained from the above chromatography. The dotted lines demonstrate the mass (pg) of progesterone-like compound(s) obtained from the pooled fractions 1–3 of the above chromatography, the solid lines that of the fractions 4–6. The broken lines indicate the radioactivity in the latter fractions.

was seen. Thus the small amount of impurity present in fractions 4–6 of the 1st chromatography was apparently completely separated in the 2nd chromatography.

In order to test the agreement of radioactivity and mass statistically, specific activity can be calculated for each fraction. These values can be then subjected to a regression analysis [9]. Radiochemical purity can be assumed, if the slope of the values over the fractions measured cannot be distinguished from zero. In order to further increase the weight of evidence yielded by such an experiment, the radiochemical purity may be investigated in replicated experiments so that even the linearity of the specific activities can be tested statistically by a slightly more complex regression analysis [10]. Our experiment (Fig. 5a) performed in three replicates showed that the celite chromatography did not yield uniformly pure progesterone in the fraction between the 4th and 7th ml of the eluate, usually collected for radioimmunoassay*. However, a repeated chromatography of this fraction in the same chromatographic system resulted in radiochemically pure progesterone (Fig. 5b). This conclusion follows from the constancy of the specific activity in the progesterone fraction, as demonstrated by the fact that the slope of the specific activities (0.15; standard error of slope 0.21) was not distinguishable from zero and that there was not found any evidence against linearity (F-value found = 1.47; $F_{0.95(2,8)} = 4.46$).

The test of validity proper consisted then of the comparison of radioimmunoassay measurements performed on progesterone purified by a single chromatography (as in a current radioimmunoassay) on the one hand, and on progesterone which was re-chromatographed on the other hand. The same pool of follicular phase plasma was used as in the experiments mentioned above. The results yielded by the radioimmunoassay of the re-chromatographed progesterone (representing the true value as shown above) were significantly lower than those obtained without re-chromatography (Table 4). This means that the radioimmunoassay currently used cannot be considered valid in terms of the definition proposed, and that two chromatographies in the same system would be needed to achieve validity.

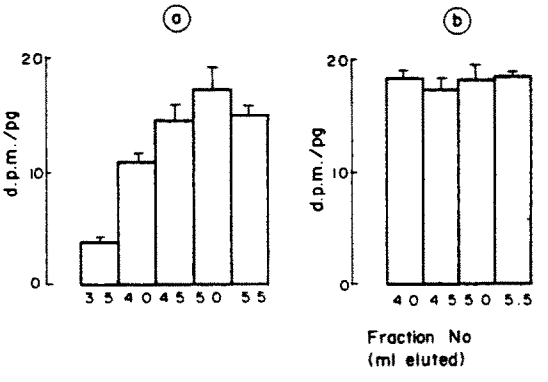

Fig. 5. Specific activity of progesterone in individual chromatographic fractions of 0.5 ml. Means ($n = 3$) are depicted as columns and standard deviations as vertical bars. (a) Chromatography of a plasma ether extract on a celite column. (b) Re-chromatography on a celite column of fractions 4–6 obtained from a previous chromatography on celite.

* In order to obtain as many values of specific activity as possible, 0.5 ml fractions were collected between the 1st and 7th ml of the eluate. Unmeasurable values of specific activity were found in fractions 1.0–3.0 and in fractions 6.0, 6.5 and 7.0.

Table 4. Progesterone levels (pg/ml) in a plasma pool*, as measured following a single chromatography† and re-chromatography‡ in three replicates

Parameters	Chromatography	Re-chromatography
Mean	422	344
Standard deviation	15.7	12.9
t-test	6.66	
Significance	$P < 0.01$	

* A follicular phase plasma pool was collected from 14 women (age 22–35 years).

† Chromatography was performed as described previously [4] on four celite columns; fractions 4–6 were pooled and one half of this pool was assayed.

‡ The other half of the pool described under "†" was chromatographed on two fresh celite columns in the same chromatographic system; fractions 4–6 were pooled and assayed.

One may speculate that if the antisera used were even less specific than the present one, and if the concentration of the steroid assayed were even lower in relation to the interfering compounds, it might happen that several chromatographies in different chromatographic systems might be needed to achieve validity. If, on the other hand, both the specificity of the antiserum and the ratio steroid:impurities were much higher than in the present case, it would be conceivable that one single chromatography, or even a non-chromatographic method might yield valid results.

The above test seems to provide a satisfactory evidence of validity and it may be generally used in radioimmunoassay laboratories since it does not require any other additional equipment than that for chromatography.

In case of any doubts that an interfering compound would exhibit the same chromatographic behaviour in the system(s) used and, at the same time, would cross-react significantly with the antiserum, a validity test should be performed which would be based on a comparison with an independent method such as gas liquid chromatography-mass spectrometry (*e.g.* [11]).

Finally, it should not be forgotten that the examples shown above clearly indicate that the test of validity described is applicable only to samples of the same character, *i.e.* of the same species, sex and source, and of a similar physiological and pathological state of the subject.

Acknowledgements—The expenses of these investigations were defrayed by research grants from the Swedish Medical Research Council, the Ford Foundation and the World Health Organization.

REFERENCES

1. Ekins R. P.: The future development of immunoassay. Int. Symposium on Radioimmunoassay and Related Procedures in Medicine, Berlin (West), 1977, Int. Atomic Energy Agency and WHO (1977) 1–39.
2. Cekan Z.: Assessment of reliability of steroid radioimmunoassays. *J. steroid Biochem.* **6** (1975) 271–275.
3. Cekan S. Z.: Reliability of steroid radioimmunoassays. *Acta Univ. upsalien.* No. **14** (1976) 1–48.
4. Brenner P. F., Guerrero R., Cekan Z. and Diczfalusy E.: Radioimmunoassay method for six steroids in human plasma. *Steroids* **22** (1973) 775–794.
5. Aso T., Guerrero R., Cekan Z. and Diczfalusy E.: A rapid 5 hour radioimmunoassay of progesterone and oestradiol in human plasma. *Clin. endocr.* **4** (1975) 173–182.
6. Guerrero R., Aso T., Brenner P. F., Cekan Z., Landgren B.-M., Hagenfeldt K. and Diczfalusy E.: Studies on the pattern of circulating steroids in the normal menstrual cycle. 1. Simultaneous assay of progesterone, pregnenolone, dehydroepiandrosterone, testosterone, dihydrotestosterone, androstenedione, oestradiol and oestrone. *Acta endocr., Copenh.* **81** (1976) 133–149.
7. Bedolla-Tovar N., Rahman S. A., Cekan S. Z. and Diczfalusy E.: Assessment of the specificity of norethisterone radioimmunoassays. *J. steroid Biochem.* **9** (1978) 561–567.
8. Cekan S. Z. and de Gomez S. M.: A method to assess radiochemical purity of compounds measured by radioimmunoassay. *Analyt. Lett.* **12(B6)** (1979) 589–602.
9. Baggett B. and Engel L. L.: The use of countercurrent distribution for the study of radiochemical purity. *J. biol. Chem.* **229** (1957) 443–450.
10. Scientific Tables (Edited by K. Diem and C. Lentner), 7th edition. Ciba-Geigy Ltd., Basel (1970).
11. Björkhem I., Blomstrand R. and Lantto O.: Validation of routine methods for serum progesterone determination using mass fragmentography. *Clin. chim. Acta* **65** (1975) 343–350.