# A STUDY OF THE RELATIONSHIP BETWEEN THE CHROMATOGRAPHIC MOBILITIES OF PEPTIDES AND THEIR CONSTITUENT AMINO ACIDS ON PAPER AND THIN LAYERS OF CELLULOSE

C. HAWORTH AND R. W. A. OLIVER

*Department of Chemistry and Applied Chemistry, University of Salford, Salford, Lancs. (Great Britain)*

## SUMMARY

The thin-layer chromatographic properties of twenty-seven commercially available di- and tripeptides and fourteen constituent amino acids in five different solvent systems on thin-layers of cellulose are reported. Theoretical $R_F$ values determined using the equation of PARDEE are compared with the observed $R_F$ values and found in many cases to give wide differences.

New empirical relationships for calculating the $R_F$ values of small peptides from the $R_F$ values of the constituent amino acids are proposed. The relationships are based on the additivity of $R_F$ values and are valid for both thin-layer and paper chromatography.

It is shown that $R_F$ values determined using these new equations agree more closely with the observed and reported $R_F$ values than does the PARDEE equation.

## INTRODUCTION

The techniques of paper (PC) and thin-layer chromatography (TLC) are widely used for the separation of amino acids and related compounds found in biological fluids and protein hydrolysates[1-3]. The identification of a particular spot on the chromatogram of such fluids is normally achieved by comparison of its $R_F$ value with those of known amino acids or other reference compounds[1]. In certain difficult cases it may be necessary to augment the co-chromatographic conclusions by observing the colours given by the spot containing the unknown compound with a variety of chromogenic reagents[4].

Because of the very large number of peptides which may be formed by the commonly occurring amino acids, any attempt to apply these procedures to the problem of identification of peptides is clearly impracticable. This is also the case even if restricted to those peptides containing eight or less amino acid residues which are commonly found in biological fluids[5]. However, the analysis time required for peptide characterisation may be greatly reduced if some information as to its identity

could be obtained from PC or TLC. Even using advanced modern techniques artefacts arising from the hydrolysis and work up procedures may still lead to misidentification of certain labile amino acid residues such as asparaginyl and glutaminyl. This may also be the case when the peptides are combined with acetyl and sulphate groups or sugar moieties. One possible solution is to correlate the chromatographic properties of the intact peptide with the chromatographic properties of the individual constituent amino acids. In this way it should be possible to obtain valuable information on the structure of the peptide. Only two papers[6,7] have been published which report a relationship between the amino acid composition of a peptide and its mobility in PC systems. In each case the equation devised by PARDEE[7] was used; which is in turn based on the MARTIN[8] equation.

According to PARDEE[7]

$$RT \ln \left( \frac{1}{R_F} - 1 \right)_{\text{peptide}} = (n - 1) A + B + \Sigma RT \ln \left( \frac{1}{R_F} - 1 \right)_{\text{amino acid}} \quad (1)$$

where $R$ is the molar gas constant, $T$ is the absolute temperature, $A$ and $B$ are experimentally determined constants for a given solvent system and $n$ is the number of amino acid residues in the peptide. A review of the current chromatographic literature revealed that no similar reports had been published for TLC systems. Because of their speed and sensitivity TLC systems are becoming more extensively used for the identification of mixture components. Consequently, in view of the importance of the determination of the structure of peptides we decided to investigate the applicability of the above equation to TLC systems. Cellulose was chosen as the support since such systems are reported to correspond well with PC[9] and also it is widely used for the separation of amino acids[1-3].

The present paper gives the $R_F$ data of twenty-seven commercially available synthetic di- and tripeptides and their constituent amino acids in five different solvent systems. From the results obtained a new empirical relationship is derived which may be used to confirm the amino acid analyses for small peptides. Reported PC data[10] is used in an attempt to extend the equation to higher peptides. It is shown that the empirical equation is valid for use in both TLC and PC systems.

MATERIALS AND EQUIPMENT

*Chromatographic equipment*
    Shandon* equipment was used throughout this work for the preparation and development of the cellulose thin-layers.

*Cellulose powder*
    The cellulose powder used in this investigation was MN-300 (without binder)**; it was washed by the technique described previously[2].

*Amino acids and peptides*
    With the exception of DL-alanyl-DL-alanine, DL-leucyl-DL-valine, and DL-

alanyl–DL-serine the amino acids and peptides used in this work were the L-enantio-
morphs and obtained from a variety of commercial suppliers.

### Standard solutions

Stock solutions (0.025 $M$) of amino acids and peptides were made up in 2-
propanol in water (10%).

### Detection reagent

Ninhydrin–cadmium acetate chromogenic reagent was used throughout this
work for the detection of amino acids and peptides. It was prepared as described
previously[3].

## METHODS

### Preparation of thin layers

Thin layers of cellulose initially 400 $\mu$ thick were spread on 20 $\times$ 20 cm glass
plates as described by HAWORTH AND HEATHCOTE[2].

### Application of solutions to thin layers

First, a line was marked 1.5 cm from the edge of the layer using a soft lead
pencil. The starting points were then positioned 1.5 cm from each other along this
line, leaving a margin (2 cm) at each edge in order to reduce the error due to edge
effect to a minimum. The limit of the solvent front (13 cm from the origin) was marked
in a similar manner. The standard solutions (1 $\mu$l) were then applied to the marked
origins and the chromatograms developed in the chosen solvent system (100 ml).

### Solvents for chromatographic development

During the course of this work a systematic investigation of the mobility of
amino acids and peptides in fifty different solvent systems on thin layers of cellulose
was made of which 1–5 are examples.

(1) 2-propanol–butanone–1 $N$ hydrochloric acid (60:15:25).
(2) 2-methylbutan-2-ol–butanone–propanone–methanol–water–(0.88)ammonia
(50:20:10:5:15:5).
(3) 2-propanol–water (75:25).
(4) 2-propanol–water–acetic acid (75:20:5).
(5) 2-propanol–water–(0.88)ammonia (75:20:5).

### Development

When the solvent front had reached the required distance from the origin
(13 cm) the plates were removed from the tank and dried in a stream of cold air for
15 min followed by heating in a convection oven at 60° for 15 min to remove final
traces of solvent.

### Detection of amino acids and peptides

The amino acids and peptides were localized using the ninhydrin–cadmium
acetate chromogenic reagent[3]. After spraying with the reagent the plates were heated
at 60° for 15 min in a convection oven to develop the colours. The positions were then

TABLE I

$R_F$ × 100 VALUES OF AMINO ACIDS AND PEPTIDES AFTER ONE DIMENSIONAL CHROMATOGRAPHY ON THIN-LAYERS OF CELLULOSE

Solvent systems (1) 2-propanol–butanone–1 N hydrochloric acid (60:15:25); (2) 2-methylbutan-2-ol–butanone–propanone–methanol–water–(0.88) ammonia (50:20:10:5:15:5); (3) 2-propanol–water (75:25); (4) 2-propanol–water–acetic acid (75:20:5); (5) 2-propanol–water–(0.88) ammonia (75:20:5).

| Peptide; amino acid | Solvent systems | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Ala–Ala | 65 | 26 | 55 | 68 | 58 |
| Ala–Asp | 56 | 1 | 45 | 44 | 19 |
| Ala–Glu | 64 | 5 | 58 | 56 | 29 |
| Ala–Gly | 50 | 17 | 36 | 46 | 46 |
| Ala–Phe | 94 | 52 | 86 | 84 | 85 |
| Ala–Ser | 52 | 17 | 36 | 41 | 49 |
| Gly–Ala | 52 | 18 | 37 | 43 | 46 |
| Gly–Asp | 43 | 0 | 30 | 29 | 13 |
| Gly–Gly | 34 | 13 | 22 | 29 | 34 |
| Gly–His | 7 | 16 | 5 | 16 | 32 |
| Gly–Ile | 81 | 49 | 65 | 80 | 75 |
| Gly–Leu | 82 | 51 | 67 | 87 | 80 |
| Gly–Lys | 14 | 11 | 2 | 21 | 27 |
| Gly–Phe | 75 | 50 | 65 | 76 | 67 |
| Gly–Pro | 47 | 17 | 39 | 45 | 44 |
| Gly–Ser | 32 | 13 | 22 | 28 | 36 |
| Gly–Tyr | 68 | 28 | 45 | 64 | 51 |
| Gly–Val | 72 | 36 | 56 | 73 | 62 |
| Leu–Ala | 97 | 56 | 88 | 90 | 89 |
| Leu–Gly | 82 | 52 | 67 | 84 | 83 |
| Leu–Val | 100 | 77 | 98 | 96 | 95 |
| Val–Gly | 67 | 38 | 56 | 70 | 69 |
| Val–Leu | 100 | 80 | 98 | 100 | 95 |
| Ala–Gly–Gly | 47 | 13 | 34 | 39 | 43 |
| Glu–Cys–Gly | 16 | 0 | 7 | 7 | 0 |
| Gly–Gly–Gly | 32 | 8 | 20 | 31 | 30 |
| Val–Gly–Gly | 65 | 28 | 52 | 65 | 61 |
| Ala | 57 | 18 | 46 | 51 | 55 |
| Asp | 48 | 2 | 33 | 34 | 27 |
| Cys | 12 | 5 | 6 | 5 | 32 |
| Gly | 37 | 14 | 34 | 33 | 50 |
| Glu | 56 | 2 | 47 | 44 | 33 |
| His | 11 | 21 | 10 | 21 | 46 |
| Ile | 90 | 51 | 77 | 79 | 76 |
| Leu | 90 | 55 | 83 | 83 | 78 |
| Lys | 16 | 13 | 10 | 23 | 36 |
| Phe | 82 | 55 | 75 | 72 | 76 |
| Pro | 58 | 24 | 54 | 54 | 61 |
| Ser | 39 | 23 | 34 | 31 | 50 |
| Tyr | 72 | 27 | 58 | 60 | 57 |
| Val | 79 | 35 | 68 | 73 | 71 |

noted and the $R_F$ values determined. The $R_F$ data given in this work are the mean values of at least three determinations on separate plates.

## RESULTS AND DISCUSSION

The $R_F \times 100$ values of the peptides and constituent amino acids studied are given for solvents 1–5 in Table I. In Table II the observed $R_F \times 100$ values in the five solvent systems are listed with the theoretical values (eqn. 1) and the differences between the two ($\Delta$) for each of the di- and tripeptides studied. A study of the data in Table II shows that generally the results obtained by calculation in all five solvent systems gave good agreement with the observed values. However, the $R_F \times 100$ values given by calculation differed by as much as $\pm 10$ $R_F \times 100$ units in 6 cases. The $R_F \times 100$ data obtained for TLC during the course of this work generally lay within the range $\pm 2$ $R_F \times 100$ units; thus a $\Delta$ value of greater than 4 in Table II would imply that the equation was incorrect. Using this criterion the values obtained using the PARDEE[7] equation were incorrect in 44 out of 115 cases for the dipeptides and 3 out of 20 cases for the tripeptides. In view of this it was decided to re-investigate the relationship between the $R_F$ of a peptide and the $R_F$ values of the constituent amino acids.

Since we possessed several (23) dipeptides these were made the subject of the initial study. For each solvent system examined the following plots were made.

*Plot 1. $R_F \times 100$ value of the dipeptide versus the $R_F \times 100$ value of the N-terminal amino acid.*



Fig. 1. Plot of $R_F \times 100$ value of dipeptide vs. $R_F \times 100$ value of N-Terminal amino acid for the solvent system 2-propanol–butanone–1 $N$ hydrochloric acid (60:15:25) on thin layers of cellulose. O = Glycyl; △ = alanyl; □ = phenylalanyl; ▽ = leucyl.
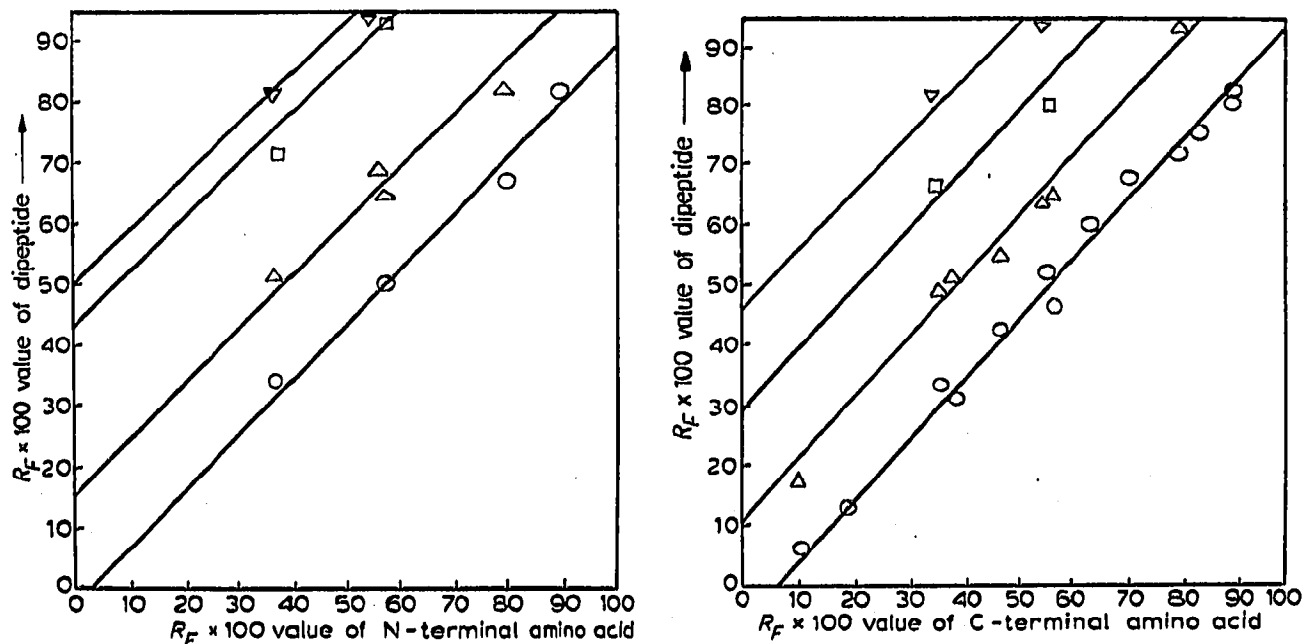
Fig. 2. Plot of $R_F \times 100$ value of dipeptide vs. $R_F \times 100$ value of C-terminal amino acid for the solvent system 2-propanol–butanone–1 $N$ hydrochloric acid (60:15:25) on thin layers of cellulose. O = Glycyl; △ = alanyl; □ = valyl; ▽ = leucyl.

## TABLE II

COMPARISON OF OBSERVED $R_F \times 100$ VALUES AND $R_F \times 100$ DATA CALCULATED USING THE EQUATION OF PARDEE FOR SOME SYNTHETIC DI- AND TRI-PEPTIDES AFTER CHROMATOGRAPHY ON THIN-LAYERS OF CELLULOSE USING THE SOLVENT SYSTEMS LISTED IN THE TEXT

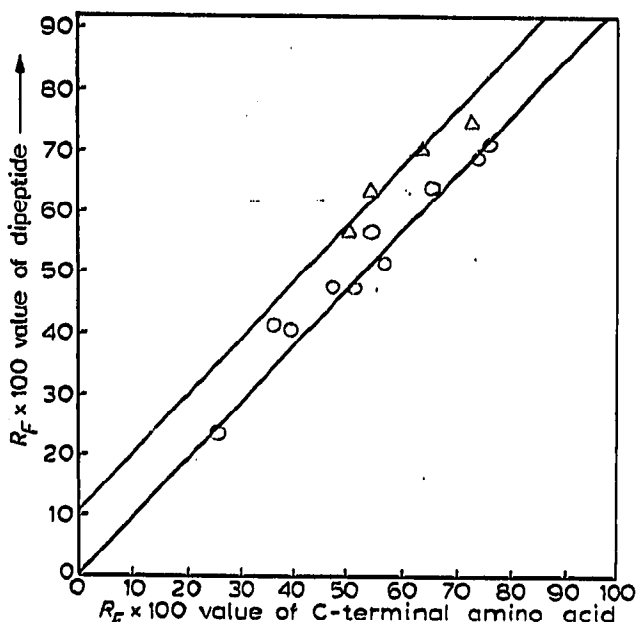| Peptide | Solvent systems | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | 1 | | | 2 | | | 3 | | | 4 | | | 5 | | |
| | Obs. | Calc. | Δ | Obs. | Calc. | Δ | Obs. | Calc. | Δ | Obs. | Calc. | Δ | Obs. | Calc. | Δ |
| Ala–Ala | 65 | 69 | +4 | 26 | 19 | −7 | 55 | 50 | −5 | 68 | 66 | −2 | 58 | 54 | −4 |
| Ala–Asp | 56 | 61 | +5 | 1 | 2 | +1 | 45 | 37 | −8 | 44 | 49 | +5 | 19 | 26 | +7 |
| Ala–Glu | 64 | 68 | +4 | 5 | 2 | −3 | 58 | 51 | −7 | 56 | 60 | +4 | 29 | 32 | +3 |
| Ala–Gly | 50 | 49 | −1 | 17 | 15 | −2 | 36 | 38 | +2 | 46 | 48 | +2 | 46 | 49 | +3 |
| Ala–Phe | 94 | 88 | −6 | 52 | 57 | +5 | 86 | 78 | −8 | 84 | 83 | −1 | 85 | 75 | −10 |
| Ala–Ser | 52 | 52 | 0 | 17 | 24 | +7 | 36 | 38 | +2 | 41 | 46 | +5 | 49 | 49 | 0 |
| Gly–Ala | 52 | 49 | −3 | 18 | 15 | −3 | 37 | 38 | +1 | 43 | 48 | +5 | 46 | 49 | +3 |
| Gly–Asp | 43 | 41 | −2 | 0 | 2 | +2 | 30 | 26 | −4 | 29 | 32 | +3 | 13 | 23 | +10 |
| Gly–Gly | 34 | 30 | −4 | 13 | 11 | −2 | 22 | 27 | +5 | 29 | 31 | +2 | 34 | 44 | +10 |
| Gly–His | 7 | 8 | +1 | 16 | 17 | +1 | 5 | 7 | +2 | 16 | 19 | +3 | 32 | 40 | +8 |
| Gly–Ile | 81 | 87 | +6 | 49 | 45 | −4 | 65 | 70 | +5 | 80 | 77 | −3 | 75 | 72 | −3 |
| Gly–Leu | 82 | 87 | +5 | 51 | 49 | −2 | 67 | 78 | +11 | 87 | 81 | −6 | 80 | 74 | −6 |
| Gly–Lys | 14 | 12 | −2 | 11 | 11 | 0 | 2 | 7 | +5 | 21 | 21 | 0 | 27 | 31 | +4 |
| Gly–Phe | 75 | 77 | +2 | 50 | 49 | −1 | 65 | 68 | +3 | 76 | 70 | −6 | 67 | 72 | +5 |
| Gly–Pro | 47 | 51 | +4 | 17 | 20 | +3 | 39 | 45 | +6 | 45 | 51 | +6 | 44 | 53 | +9 |
| Gly–Ser | 32 | 32 | 0 | 13 | 19 | +6 | 22 | 27 | +5 | 28 | 29 | +1 | 36 | 44 | +8 |
| Gly–Tyr | 68 | 66 | −2 | 28 | 23 | −5 | 45 | 50 | +5 | 64 | 57 | −7 | 51 | 51 | 0 |
| Gly–Val | 72 | 74 | +2 | 36 | 30 | −6 | 56 | 60 | +4 | 73 | 71 | −2 | 62 | 66 | +4 |
| Leu–Ala | 97 | 94 | −3 | 56 | 57 | +1 | 88 | 85 | −3 | 90 | 90 | 0 | 89 | 77 | −12 |
| Leu–Gly | 82 | 87 | +5 | 52 | 49 | −3 | 67 | 78 | +11 | 84 | 81 | −3 | 83 | 74 | −9 |
| Leu–Val | 100 | 100 | — | 77 | 76 | −1 | 98 | 93 | −5 | 96 | 96 | 0 | 95 | 90 | −5 |
| Val–Gly | 67 | 74 | +7 | 38 | 30 | −8 | 56 | 60 | +4 | 70 | 71 | +1 | 69 | 66 | −3 |
| Val–Leu | 100 | 100 | — | 80 | 76 | −4 | 98 | 93 | −5 | 100 | 96 | −4 | 95 | 87 | −8 |
| Ala–Gly–Gly | 47 | 48 | +1 | 13 | 13 | 0 | 34 | 34 | 0 | 39 | 38 | −1 | 43 | 40 | −3 |
| Gly–Cys–Gly | 16 | 14 | −2 | 0 | 1 | +1 | 7 | 12 | +5 | 7 | 4 | −3 | 0 | 6 | +6 |
| Gly–Gly–Gly | 32 | 29 | −3 | 8 | 7 | −1 | 20 | 23 | +3 | 31 | 27 | −4 | 30 | 35 | +5 |
| Val–Gly–Gly | 65 | 63 | −2 | 28 | 29 | +1 | 52 | 56 | +4 | 65 | 69 | +4 | 61 | 57 | −4 |

Fig. 3. Plot of $R_F \times$ 100 value of dipeptide *vs.* $R_F \times$ 100 value of C-terminal amino acid for the solvent system 1-propanol–water–(0.88) ammonia (60:10:30) on Whatman No. 4 chromatographic paper according to data given by MOORE AND BAKER[6]. $\bigcirc$ = Glycyl; $\triangle$ = alanyl.

*Plot 2.* $R_F \times$ 100 value of the dipeptide *versus* the $R_F \times$ 100 value of the C-terminal amino acid.

The straight line plots obtained were constructed by the method of least squares and are given in Figs. 1 and 2, respectively, for the solvent system 1. A study of these figures shows that the $R_F$ values of the dipeptides are linearly related to either the $R_F$ value of the C– or N–terminal amino acid, dependent upon the particular type of plot. Further, the gradients of the straight lines were approximately unity and the intercept on the abscissa varied with the structure of the terminal amino acid. For example, there is an increase in the $R_F$ of the dipeptide as the N-terminal amino acid goes from glycine, alanine, valine to leucine (Fig. 2). Reported PC data[6] gave similar straight line plots as shown in Fig. 3.

Since the dipeptides possessed has a greater variety of C-terminal amino acids and relatively few N-terminal amino acid combinations the remainder of this study consisted of an analysis of type 2 plots.

These linear empirical relationships may now be quantitated as follows:

$$R_F \ (X-Y) = mR_F \ (Y) + I$$

where $R_F \ (X-Y)$ is the $R_F$ value of the dipeptide $X–Y$; $R_F \ (Y)$ is the $R_F$ value of the free C-terminal amino acid; $m$ is the gradient and $I$ is the intercept of the plot on the abscissa.

Since the N-terminal amino acid is kept constant for a single plot, the intercept $I$ may be related to the $R_F$ of this amino acid either by the equation:

$$I = pR_F \ (X)$$

or

$$I = a + R_F \ (X)$$

where $p$ and $\ddot{a}$ are constants of the solvent system.

TABLE III

LINE CONSTANTS OBTAINED FROM A PLOT OF $R_F \times$ 100 VALUE OF DIPEPTIDE *versus* $R_F \times$ 100 VALUE OF THE C-TERMINAL AMINO ACID AFTER CHROMATOGRAPHY ON THIN LAYERS OF CELLULOSE USING THE SOLVENT SYSTEMS LISTED IN THE TEXT

| Solvent system | N-terminal amino acid residue | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Glycyl | | | Alanyl | | | Valyl | | | Leucyl | | |
| | m | p | a | m | p | a | m | p | a | m | p | a |
| 1 | 1.0 | −0.19 | −44 | 1.0 | 0.20 | −47 | 1.0 | 0.37 | −50 | 1.0 | 0.50 | −44 |
| 2 | 1.0 | −0.28 | −14 | 1.0 | 0.11 | −17 | 1.1 | 0.70 | −13 | 1.1 | 0.63 | −18 |
| 3 | 1.0 | −0.26 | −42 | 1.0 | 0.14 | −41 | 1.0 | 0.33 | −43 | 1.0 | 0.49 | −43 |
| 4 | 1.1 | −0.20 | −43 | 1.1 | 0.14 | −41 | 1.0 | 0.44 | −41 | 1.0 | 0.49 | −43 |
| 5 | 1.0 | −0.20 | −60 | 1.0 | 0.12 | −62 | 1.0 | 0.14 | −61 | 1.0 | 0.25 | −58 |

In order to understand the significance of the quantities $m$, $p$ and $a$ the plots were analysed for the four series of dipeptides in the solvent systems 1–5, and are tabulated above in Table III. From a study of the tabulated results it is apparent that in general the gradient $m$ is equal to unity. Exceptions may be due to several factors. First, the small number of peptides available containing valyl and leucyl residues leads to a small number of $R_F$ data. Secondly, in those cases where the values occur in a small range it is difficult to extrapolate with accuracy over the whole $R_F$ range. Finally, due to the non-equilibrium state of the solvent cellulose system and to demixing phenomena, $R_F$ values below 0.2 and above 0.8 are invalid for extrapolation purposes[11]. If these factors are taken into account then it is suggested that these non-linear plots are only apparent and do not represent real exceptions to the generally observed linear relationship between the $R_F$ value of the dipeptide and the $R_F$ values of the constituent amino acids.

On the contrary the quantity $p$ varies both with the solvent system and with the N-terminal amino acid, while the quantity $a$ varies only with the solvent system.

This analysis gives the exact relationship between the $R_F$ value of the dipeptide and its constituent amino acids on thin layers of cellulose for a given solvent system by the equation:

$$R_{F\ (X-Y)} = a + R_{F\ (X)} + R_{F\ (Y)}$$

TABLE IV

VALUES OF THE QUANTITY $a'$ OBTAINED FOR THE GIVEN TRIPEPTIDES AFTER CHROMATOGRAPHY ON THIN-LAYERS OF CELLULOSE USING THE SOLVENT SYSTEMS LISTED IN THE TEXT

| Solvent system | Tripeptides | | | |
| --- | --- | --- | --- | --- |
| | Gly–Gly–Gly | Ala–Gly–Gly | Val–Gly–Gly | Glu–Cys–Gly |
| 1 | − 79 | − 84 | − 88 | − 89 |
| 2 | − 34 | − 33 | − 35 | − 21 |
| 3 | − 82 | − 80 | − 84 | − 80 |
| 4 | −128 | −147 | −128 | − 74 |
| 5 | −120 | −112 | −110 | −115 |

In an attempt to extend the empirical relationship to tripeptides the values of the quantity $a'$ in the following equation are listed in Table IV

$$R_F \, (X-Y-Z) = a' + R_F \, (X) + R_F \, (Y) + R_F \, (Z)$$

A study of the values listed in this table shows that the quantity $a'$ varies only with the solvent system. Comparison of the values of the quantity $a'$ and the corresponding quantity $a$ given in Table III shows that these two constants are related by the equation:

$$a' = 2a$$

This suggests that the general equation relating the mobility of any di- or tripeptide to its constituent amino acids is:

$$R_F \, (peptide) = (n - 1) \, a + \Sigma R_F \, (amino \; acid) \tag{2}$$

where $a$ is calculated from known $R_F$ data.

In Table V the observed $R_F \times 100$ values in the solvent systems 1–5 are listed with the calculated values (eqn. 2) and the differences ($\Delta$) between the two for each of the di- and tripeptides studied. Using the same criterion adopted for the PARDEE[7] equation only 33 of the 115 dipeptides and 5 of the 20 tripeptides gave $\Delta$ values which differed by more than 4. This may be compared with the incorrect values obtained in 44 and 3 cases, respectively, by the PARDEE equation.

Since this empirical relationship (eqn. 2) is more convenient to use than the PARDEE equation and gave results which corresponded more closely to the observed values it was decided to see if it could be applied to higher peptides. Unfortunately, few of these higher peptides were available and so it was necessary to test the applicability of the equation with reported PC data[10]. For PC data PARDEE adopted the value of $R_F \times 100 \pm 4$ for the allowed deviation of duplicate determinations; thus a $\Delta$ value of greater than 8 in Table VI would imply that the equation was incorrect. A study of Table VI shows that the equation gave the incorrect value in 2 out of 9 cases for the tripeptides and 5 out of 6 cases for higher peptides. It was, therefore, found necessary to introduce further terms into eqn. 2 in order to fit the calculated data with the observed data[12].

It is suggested that the equation becomes:

$$R_F \, (peptide) = (n - 1) \, a + (n - 2) \, b + \Sigma R_F \, (amino \; acid) \tag{3}$$

for tripeptides, and

$$R_F \, (peptide) = (n - 1) \, a + (n - 2) \, b + (n - 3) \, c + \Sigma R_F \, (amino \; acid) \tag{4}$$

for tetrapeptides and so on for higher peptides.

Using eqn. 3 and eqn. 4 respectively, the $R_F$ data was recalculated for the tri- and tetrapeptides. The theoretical $R_F \times 100$ values are compared with the reported $R_F \times 100$ values in Table VII. From a study of this table it can be seen that the correct value is obtained in 8 out of 9 cases for the tripeptides and in all 4 cases for the tetrapeptides. The $R_F$ equations are therefore valid and can be used to determine the $R_F$ values of peptides in any solvent system provided the $R_F$ values of the individual amino acids and the constants are known. However, more data on the behaviour of peptides of greater chain length are needed to test the usefulness of the general

## TABLE V

COMPARISON OF OBSERVED $R_F \times 100$ VALUES AND $R_F \times 100$ DATA CALCULATED USING THE EMPIRICAL EQUATION 2 FOR SOME SYNTHETIC DI- AND TRIPEPTIDES AFTER CHROMATOGRAPHY ON THIN-LAYERS OF CELLULOSE USING THE SOLVENT SYSTEMS LISTED IN THE TEXT

| Peptide | Solvent systems | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | | 2 | | | 3 | | | 4 | | | 5 | | |
| | Obs. | Calc. | Δ | Obs. | Calc. | Δ | Obs. | Calc. | Δ | Obs. | Calc. | Δ | Obs. | Calc. | Δ |
| Ala-Ala | 65 | 69 | +4 | 26 | 19 | −7 | 55 | 49 | −6 | 68 | 65 | −3 | 58 | 50 | −8 |
| Ala-Asp | 56 | 60 | +4 | 1 | 3 | +2 | 45 | 36 | −9 | 44 | 48 | +4 | 19 | 22 | +3 |
| Ala-Glu | 64 | 68 | +4 | 5 | 3 | −2 | 58 | 50 | −8 | 56 | 58 | +2 | 29 | 28 | −1 |
| Ala-Gly | 50 | 49 | −1 | 17 | 15 | −2 | 36 | 37 | +1 | 46 | 47 | +1 | 46 | 45 | −1 |
| Ala-Phe | 94 | 94 | 0 | 52 | 56 | +4 | 86 | 78 | −8 | 84 | 86 | +2 | 65 | 71 | +6 |
| Ala-Ser | 52 | 51 | −1 | 17 | 24 | +7 | 36 | 37 | +1 | 41 | 45 | +4 | 49 | 45 | −4 |
| Gly-Ala | 52 | 49 | −3 | 18 | 15 | −3 | 37 | 37 | 0 | 43 | 47 | +4 | 46 | 45 | −1 |
| Gly-Asp | 43 | 40 | −3 | 0 | 0 | — | 30 | 24 | −6 | 29 | 30 | +1 | 13 | 17 | +4 |
| Gly-Gly | 34 | 29 | −5 | 13 | 11 | −2 | 22 | 25 | +3 | 29 | 29 | 0 | 34 | 40 | +6 |
| Gly-His | 7 | 3 | −4 | 16 | 18 | +2 | 5 | 1 | −4 | 16 | 17 | +1 | 32 | 36 | +4 |
| Gly-Ile | 81 | 82 | +1 | 49 | 48 | −1 | 65 | 67 | +2 | 80 | 85 | +5 | 75 | 66 | −9 |
| Gly-Leu | 82 | 82 | 0 | 51 | 52 | +1 | 67 | 74 | +7 | 87 | 79 | −8 | 80 | 68 | −12 |
| Gly-Lys | 14 | 8 | −6 | 11 | 10 | −1 | 2 | 1 | −1 | 21 | 19 | −2 | 27 | 26 | −1 |
| Gly-Phe | 75 | 74 | −1 | 50 | 52 | +2 | 65 | 66 | +1 | 76 | 68 | −8 | 67 | 66 | −1 |
| Gly-Pro | 47 | 50 | +3 | 17 | 21 | +4 | 39 | 45 | +6 | 45 | 50 | +5 | 44 | 51 | +7 |
| Gly-Ser | 32 | 31 | −1 | 13 | 20 | +7 | 22 | 25 | +3 | 28 | 27 | −1 | 36 | 40 | +4 |
| Gly-Tyr | 68 | 64 | −4 | 28 | 24 | −4 | 45 | 49 | +4 | 64 | 56 | −8 | 51 | 47 | −4 |
| Gly-Val | 72 | 71 | −1 | 36 | 32 | −4 | 56 | 59 | +3 | 73 | 69 | −4 | 62 | 61 | −1 |
| Leu-Ala | 97 | 100 | — | 56 | 56 | 0 | 88 | 86 | −2 | 90 | 97 | +7 | 89 | 73 | −16 |
| Leu-Gly | 82 | 82 | 0 | 52 | 52 | 0 | 67 | 74 | +7 | 84 | 79 | −5 | 83 | 68 | −15 |
| Leu-Val | 100 | 100 | — | 77 | 73 | −4 | 98 | 100 | — | 96 | 100 | — | 95 | 89 | −6 |
| Val-Gly | 67 | 71 | +4 | 38 | 32 | −6 | 56 | 59 | +3 | 70 | 69 | −1 | 69 | 61 | −8 |
| Val-Leu | 100 | 100 | — | 80 | 73 | −7 | 98 | 100 | — | 100 | 100 | — | 95 | 89 | −6 |
| Ala-Gly-Gly | 47 | 46 | −1 | 13 | 12 | −1 | 34 | 28 | −6 | 39 | 43 | +4 | 43 | 35 | −8 |
| Glu-Cys-Gly | 16 | 20 | +4 | 0 | 0 | — | 7 | 1 | −6 | 7 | 8 | +1 | 0 | 0 | — |
| Gly-Gly-Gly | 32 | 26 | −6 | 8 | 8 | 0 | 20 | 16 | −4 | 31 | 25 | −6 | 30 | 30 | 0 |
| Val-Gly-Gly | 65 | 68 | +3 | 28 | 29 | +1 | 52 | 50 | −2 | 65 | 65 | 0 | 61 | 61 | 0 |

TABLE VI

COMPARISON OF REPORTED PC $R_F$ × 100 DATA AND THEORETICAL $R_F$ × 100 VALUES CALCULATED USING THE EMPIRICAL EQUATION 2

Reported chromatography on Whatman No. 1 paper; solvent system: pyridine–isoamylalcohol.

| Peptide | $R_F$ × 100 values | | |
|---|---|---|---|
| | Obs. | Calc.[a] | $\Delta$ |
| Ala–Gly–Gly | 11 | 1 | − 10 |
| Gly–Asn–Leu | 19 | 20 | + 1 |
| Gly–Gly–Gly | 8 | − 5 | −13 |
| Gly–Tyr–Gly | 26 | 20 | − 6 |
| Leu–Ala–Ala | 34 | 33 | − 1 |
| Leu-Ala–Gly | 28 | 27 | − 1 |
| Leu–Gly–Ala | 29 | 27 | − 2 |
| Leu–Gly–Asp | 10 | 17 | + 7 |
| Leu–Gly–Leu | 51 | 47 | − 4 |
| Gly–Ala–Gly–Tyr | 26 | 18 | − 8 |
| Gly–Asn–Gly–Gly | 4 | −14 | − 18 |
| Gly–Gly–Gly–Gly | 7 | −13 | −20 |
| Leu–Gly–Gly–Gly | 24 | 7 | −17 |
| Ala–Gly–Gly–Ala–Gly–Gly | 9 | −17 | −26 |
| Leu–(Gly)₈–Gly | 17 | −24 | −41 |

[a] Constant $a = -19$.

equation summarized for tetrapeptides in eqn. 4. At present its use is therefore restricted to the identification of di- and tripeptides but further work is being carried out in order to extend the range of the equation. Because the two-dimensional solvent system of HEATHCOTE AND HAWORTH[3], utilising solvents sytems 1 and 2 on thin layers of cellulose, resolves some forty amino acids, it is used in this laboratory for the

TABLE VII

COMPARISON OF REPORTED PC $R_F$ × 100 DATA AND THEORETICAL $R_F$ × 100 VALUES CALCULATED USING THE EMPIRICAL EQUATIONS 3 AND 4

Reported chromatography on Whatman No. 1 paper; solvent system: pyridine–isoamylalcohol.

| Peptide | $R_F$ × 100 values | | |
|---|---|---|---|
| | Obs. | Calc.[a] | $\Delta$ |
| Ala–Gly–Gly | 11 | 6 | − 5 |
| Gly–Asn–Leu | 19 | 25 | + 6 |
| Gly–Gly–Gly | 8 | 0 | − 8 |
| Gly–Tyr–Gly | 26 | 25 | − 1 |
| Leu–Ala–Ala | 34 | 38 | + 4 |
| Leu–Ala–Gly | 28 | 32 | + 4 |
| Leu–Gly–Ala | 29 | 32 | + 3 |
| Leu–Gly–Asp | 10 | 22 | +12 |
| Leu–Gly–Leu | 51 | 52 | + 1 |
| Gly–Ala–Gly–Tyr | 26 | 32 | + 6 |
| Gly–Asn–Gly–Gly | 4 | 0 | − 4 |
| Gly–Gly–Gly–Gly | 7 | 7 | 0 |
| Leu–Gly–Gly–Gly | 24 | 27 | + 3 |

[a] Constants: $a = -19$; $b = +5$; $c = +4$.

chromatographic identification of all ninhydrin positive material. Accordingly the value of the constants $a$ and $b$ in eqn. 3 are given. For the solvent system 1 these constants have the values $-0.45$ and $+0.03$ and for the solvent system 2 the values are $-0.17$ and $0.00$ respectively.

ACKNOWLEDGEMENT

REFERENCES

1 J. G. HEATHCOTE, D. M. DAVIES, C. HAWORTH AND R. W. A. OLIVER, J. Chromatogr., 55 (1971) 377.
2 C. HAWORTH AND J. G. HEATHCOTE, J. Chromatogr., 41 (1969) 380.
3 J. G. HEATHCOTE AND C. HAWORTH, J. Chromatogr., 43 (1969) 84.
4 J. G. HEATHCOTE, R. J. WASHINGTON, C. HAWORTH AND S. BELL, J. Chromatogr., 51 (1970) 267.
5 B. SKARZYNSKI AND M. SARNECKA-KELLER, Adv. Clin. Chem., 5 (1962) 107.
6 T. B. MOORE AND C. G. BAKER, J. Chromatogr., 1 (1958) 513.
7 A. B. PARDEE, J. Biol. Chem., 190 (1951) 757.
8 A. J. P. MARTIN, Ann. N.Y. Acad. Sci., 49 (1948) 249.
9 P. J. SCHORN, Z. Anal. Chem., 205 (1964) 303.
10 C. A. KNIGHT, J. Biol. Chem., 190 (1951) 753.
11 L. S. BARK, in A. NIEDERWIESER AND G. PATAKI (Editors), Progress in Thin-Layer Chromatography and Related Methods, Vol. I, Ann. Arbor, London, 1970.
12 C. HAWORTH AND R. W. A. OLIVER, Biochem. J., in press.